MOLECULAR ASPECTS OF PAPOVAVIRUSES

Yechiel Becker, Series Editor
Julia Hadar, Managing Editor

DEVELOPMENTS IN MOLECULAR VIROLOGY

Becker, Y. (ed.) *Herpesvirus DNA* (1981)
Becker, Y. (ed.) *Replication of Viral and Cellular Genomes* (1983)
Becker, Y. (ed.) *Antiviral Drugs and Interferon: the Molecular Basis of Their Activity* (1983)
Kohn, A. and Fuchs, P. (eds.) *Mechanisms of Viral Pathogenesis from Gene to Pathogen* (1983)
Becker, Y. (ed.) *Recombinant DNA Research and Viruses. Cloning and Expression of Viral Genes* (1985)
Feitelson, M. (author) *Molecular Components of Hepatitis B Virus* (1985)
Becker, Y. (ed.) *Viral messenger RNA: Transcription, Processing, Splicing and Molecular Structure* (1985)
Doerfler, W. (ed.) *Adenovirus DNA: the Viral Genome and Its Expression* (1986)

DEVELOPMENTS IN VETERINARY VIROLOGY

Payne, L.N. (ed.) *Marek's Disease* (1985)
Burny, A. and Mammerickx, M. *Enzootic Bovine Leukosis and Bovine Leukemia Virus* (1987)
Becker, Y. (ed.) *African Swine Fever* (1987)
De Boer, G.F. (ed.) *Avian Leukosis* (1987)

DEVELOPMENTS IN MEDICAL VIROLOGY

Levine, P.H. (ed.) *Epstein-Barr Virus and Associated Diseases* (1985)
Becker, Y. (ed.) *Virus Infections and Diabetes Mellitus* (1987)

MOLECULAR ASPECTS OF PAPOVAVIRUSES

edited by

Yosef Aloni
The Weizmann Institute of Science
Rehovot, Israel

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

Distributors for the United States and Canada:
Kluwer Academic Publishers
101 Philip Drive
Assinippi Park
Norwell, Massachusetts 02061, USA

Distributors for the UK and Ireland:
Kluwer Academic Publishers
MTP Press Limited
Falcon House, Queen Square
Lancaster LA1 1RN, UNITED KINGDOM

Distributors for all other countries:
Kluwer Academic Publishers Group
Distribution Centre
Post Office Box 322
3300 AH Dordrecht, THE NETHERLANDS

Library of Congress Cataloging-in-Publication Data

Molecular aspects of papovaviruses.

(Developments in molecular virology ; 9)
Bibliography: p.
Includes index.
1. Papovaviruses. 2. Molecular biology.
I. Aloni, Yosef. II. Series.
QR406.M65 1987 576'.6484 87-19103

ISBN-13: 978-1-4612-9237-1 e-ISBN-13: 978-1-4613-2087-6
DOI: 10.1007/978-1-4613-2087-6

The figure on the cover is from DePamphilis M.L. "Replication of Simian Virus 40 and Polyoma Virus Chromosomes". The figure appears on page 3 of this book.

CONTENTS

CONTRIBUTORS

Raquel H. Adams
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

James C. Alwine
Department of Microbiology
School of Medicine
University of Pennsylvania
Philadelphia, PA 19104
U. S. A.

Christine Ambrose
Department of Chemistry
Purdue University
West Lafayette, IN 47907
U. S. A.

Claudio Basilico
Department of Pathology
New York University School of Medicine
550 First Avenue, Rm 551
New York, N.Y. 10016
U. S. A.

Avri Ben-Ze'ev
Department of Genetics
Weizmann Institue of Science
Rehovot 76100
Israel

Shelley Berger
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

Minou Bina
Department of Chemistry
Purdue University
West Lafayette, IN 47907
U. S. A.

Veronica Blasquez
Department of Chemistry
Purdue University
West Lafayette, IN 47907
U. S. A.

John Brady
Laboratory of Molecular Virology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
U. S. A.

Susan Carswell
Department of Microbiology
School of Medicine
University of Pennsylvania
Philadelphia, PA 19104
U. S. A.

Pierre Chambon
Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS
Unité 184 de Biologie Moléculaire et
de Génie Génétique de L'INSERM
Faculté de Médecine
11, rue Humann
67086 Strasbourg-Cédex
France

Murari Chaudhuri
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Lori Covey
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Pasquale Delli-Bovi
Department of Pathology
New York University School of Medicine
550 First Avenue, Rm 554
New York, N.Y. 10016
U. S. A.

Melvin L. DePamphilis
Department of Cell Biology
Roche Institute of Molecular Biology
Nutley, N.J. 07110
U. S. A.

Walter Dietrich
Institut für Klinische Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

Janet Duvall
Laboratory of Molecular Virology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
U. S. A.

William R. Folk
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

Pawel G. Fuchs
Institut für Klinische Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

Francis G. Kern
Breast Cancer Section/Medicine Branch
National Cancer Institute
National Institutes of Health
Building 10, Rm 12N 266
Bethesda, Maryland 20982
U.S.A.

George Khoury
Laboratory of Molecular Virology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
U. S. A.

Eva Kleiner
Institut für Klinische Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

Marie-Helene Kryszke
Department of Molecular Biology
Pasteur Institute
25, rue du Dr. Roux
75724 Paris, Cedex 15
France

Perry Kwok
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Gerlinde Lang
Institut für Klinische Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

John Lednicky
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

Mary Loeken
Laboratory of Molecular Virology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
U. S. A.

Henry Lowman
Department of Chemistry
Purdue University
West Lafayette, IN 47907
U. S. A.

James L. Manley
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Mark Martin
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

Sandra Pellegrini
I.C.R.F. Laboratories
P.O. Box 123
Lincoln's Fields
London, WC 2A 3PX
England

Jacques Piette
Department of Molecular Biology
Pasteur Institute
25, rue du Dr. Roux
75724 Paris, Cedex 15
France

Herbert Pfister
Institut für Klinische und Molekulare Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

Carol Prives
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Lisa C. Ryner
Department of Biological Sciences
Columbia University
New York, N.Y. 10027
U. S. A.

Gregor Sagner
Institut für Klinische Virologie
Universität Erlangen-Nürnberg
D-8520 Erlangen
Federal Republic of Germany

Walter A. Scott
Department of Biochemistry
The University of Miami School of Medicine
Miami, Florida 33101
U. S. A.

Wei-Jen Tang
Department of Microbiology
The University of Texas at Austin
Austin, Texas 78712
U. S. A.

Margaret Ann Thompson
Laboratory of Molecular Virology
National Cancer Institute
National Institutes of Health
Bethesda, Maryland 20892
U. S. A.

Alan Wildeman
Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS
Unité 184 de Biologie Moléculaire et
de Génie Génétique de L'INSERM
Faculté de Médecine
11, rue Humann
67086 Strasbourg-Cédex
France
Present address: Department of Molecular Biology and Genetics
University of Guelph
Guelph, Ontario NIG 2W1
Canada

Moshe Yaniv
Department of Molecular Biology
Pasteur Institute
25, rue du Dr. Roux
75724 Paris, Cedex 15
France

Martin Zenke
Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS
Unité 184 de Biologie Moléculaire et
de Génie Génétique de L'INSERM
Faculté de Médecine
11, rue Humann
67086 Strasbourg-Cédex
France
Present address: European Laboratory of Molecular Biology
Postfach 1022.40
6900 Heidelberg 1
West Germany

PREFACE

It is almost twenty years since the first DNA tumor virus meeting was held at Cold Spring Harbor. At this meeting studies on three tumor viruses were discussed: the papovaviruses, the adenoviruses and the herpesviruses. The present series *Developments in Molecular Virology* chose to reverse this sequence by first publishing books on the herpesviruses, followed by adenoviruses, and only now the papovaviruses.

All the DNA tumor viruses gained their original reputation by serving as model systems in animal cells for studying gene expression and gene regulation, but SV40 and polyoma have been the jewel in the crown in these studies, as λ phage was for the study of prokaryotes. SV40 was the first DNA tumor virus to be completely sequenced that enabled the definition of the *cis* controlling elements in DNA replication and transcription. I am continuously fascinated by the organization of the SV40 and polyoma genomes. Although they contain about 5000 bp that encode for only 6 to 7 proteins, the mechanisms which regulated their gene expression are varied and include almost any other type of gene regulation found today to regulate eukaryotic genes. Just to mention two: (i) the early promoter is a classical promoter that contains the TATA, CAAT and enhancer elements, while the late promoter is devoid of these elements, and (ii) the mRNA can be structurally and functionally monocistronic or dicistronic. This hints at the versatility in the control of gene expression at the transcriptional and translational levels.

Sometimes one may wonder why SV40 has been one of the best studied viruses? Is it a reflection of the scientists who have studied it or because of the special features of the viral genome? An answer to this question is exemplified in the discovery of the SV40 early promoter and enhancer. These regions contain repeated sequences that have triggered scientists to ask the obvious question: "Why are these sequences repeated"? The obvious experiments were to delete these sequences, and thus to determine their effect on gene expression and regulation. This approach has led to the discovery of the enhancer element and the *cis* controlling elements that constitute the promoter.

Although splicing was first discovered in adenoviruses, the SV40 system provided the first known use of splicing in generating two functional proteins from one primary transcript.

The oncogenes of SV40 and polyoma are among the best-studied proteins of this type and their multifunctional domains are now being unravelled. Last but not

least is the use of the SV40 minichromosome as a model system to study the nucleoprotein structure and topology of actively transcribed genes. The hypersensitive region upstream of a promoter was first demonstrated in the SV40 system.

The papovaviruses include, in addition to SV40 and polyoma, the papillomaviruses which have recently acquired their own reputation. All the above observations are dealt with in the articles in this volume.

The papovaviruses continue to be in the forefront of studies in molecular biology and oncogenesis, and it is only appropriate to include them in the present series.

I would like to thank all the contributors who for a long time have been intrigued with the papovaviruses, and agreed to submit their recent studies in this field. I am particularly indebted to Julia Hadar for critically reading the manuscripts.

I would like to dedicate this volume to the memory of Dr. George Khoury, a friend and scientist.

Yosef Aloni
Weizmann Institute of Science, Rehovot

MOLECULAR ASPECTS OF PAPOVAVIRUSES

1

REPLICATION OF SIMIAN VIRUS 40 AND POLYOMA VIRUS CHROMOSOMES

Melvin L. DePamphilis

Department of Cell Biology, Roche Institute of Molecular Biology, Nutley, N.J. 07110, USA

Abstract

Simian virus 40 and polyoma virus chromosomes have long been used as models for the replication and structure of chromatin in mammalian nuclei. This article highlights recent developments and provides simple working models for events common to both viral chromosomes. RNA-primed DNA synthesis by DNA primase-DNA polymerase-α or -δ is initiated repeatedly only on the retrograde arm of replication forks where one of many initiation sites is selected stochastically within a single-stranded DNA region ("initiation zone") that is defined by chromatin structure rather than DNA sequence. Old histone octamers are distributed to both arms of a fork and newly replicated DNA is rapidly assembled into nucleosomes with little regard for sequence specificity or strand preference. RNA primed-DNA synthesis is first initiated on the early mRNA template strand of the origin of replication(ori) by the same mechanism used to initiate Okazaki fragments at replication forks. Bidirectional DNA replication begins at a unique site at one end of the required ori sequence (ori-core) following binding of a T-antigen/permissive cell-factor initiation complex; binding to ori-core is facilitated by either promoter or enhancer elements adjacent to the late-gene side of ori-core. These transcriptional elements can determine cell-type, but not species, specificity for ori activation. Replication terminates at whatever sequence is 180° from ori, but the termination site strongly affects the way in which sibling chromosomes are separated. Topoisomerase II appears to be required specifically for termination of replication although formation of catenated intertwines is not an obligatory pathway in the separation of sibling chromosomes, suggesting that topoisomerase II acts behind replication forks rather than in front of them.

Introduction

The replication and structure of simian virus 40 (SV40) and polyoma virus (PyV) chromosomes have been reviewed in detail (1,6-12), and the reader should refer to these articles for a comprehensive list of literature citations, and to Kornberg (13) for a general review of DNA replication. In general, only recent references are cited that will direct the reader to earlier

Aloni, Y (ed), Molecular Aspects of Papovaviruses. *ISBN 0-89838-971-2.*

papers. SV40 and PyV are small (5.2 kb) circular, double-stranded DNA genomes that replicate in the nucleus of mammalian cells as chromosomes with a histone composition and nucleosome structure that are essentially the same as those of their host. With the exception of large tumor-antigen (T-Ag), which is required for initiation of viral DNA replication, all steps in the replication and assembly of these viral chromosomes are carried out by cellular components, exclusively. In fact, events at SV40 and PyV replication forks are remarkably similar to those in eukaryotic cell chromosomes (1). Even the separation of sibling molecules and termination of DNA synthesis may be the same since the topological problems involved in unwinding DNA in front of two converging replication forks in a small, covalently-closed, circular DNA molecule are the same as in a large linear DNA molecule containing multiple replication bubbles. Neither situation allows free rotation of one arm of a replication fork about the other.

The most significant differences between viral and cellular DNA replication likely occur in the mechanism for initiation of new rounds of DNA replication. Cellular chromosomes must replicate all of their DNA once and only once during cell division, whereas viral chromosomes replicate many times during a single S-phase. Therefore, T-Ag may be the viral equivalent of cellular proteins that promote amplification of specific genes under certain conditions (e.g. early stages in development or under selective pressures). The stringent specificity of interactions between T-Ag, viral DNA and cellular replication factors is apparent from the fact that SV40 DNA replicates efficiently only in certain monkey and human cells ("permissive cells", refs. 2,3), while PyV DNA replicates efficiently in most differentiated mouse cells (4), but not in embryonic or transformed mouse cell lines or non-mouse cells ("non-permissive cells", refs. 5,6). Interestingly, PyV DNA does replicate in mouse preimplantation embryos (14).

DNA Synthesis at Replication Forks

Okazaki fragments (Fig. 1)

During semiconservative DNA replication in SV40, PyV and mammalian chromosomes, newly synthesized DNA appears coincidentally on both arms of replication forks, yet all known DNA polymerases extend polynucleotide chains only at their 3-OH termini. This presents a paradox since the two parental DNA strands are anti-parallel. The solution is the repeated initiation of short nascent DNA chains ("Okazaki fragments") on the retrograde arms of replication forks (i.e. that side of the fork where the direction of DNA synthesis must be opposite to the direction of fork movement). However, this raises a second paradox since purified mammalian DNA polymerases cannot initiate synthesis *de novo*, but require the 3'-OH end of an oligonucleotide primer that is annealed to a DNA template. The solution is a second enzyme, DNA primase, that synthesizes a short oligoribonucleotide primer (referred to either as an "RNA primer" or "initiator RNA (iRNA)") at replication forks that is eventually excised to allow ligation of Okazaki fragments into a continuous DNA chain. Once a steady-state is achieved in SV40 and PyV replicating DNA, each fork contains an average of 0.25 to 1 Okazaki fragment and 50% of

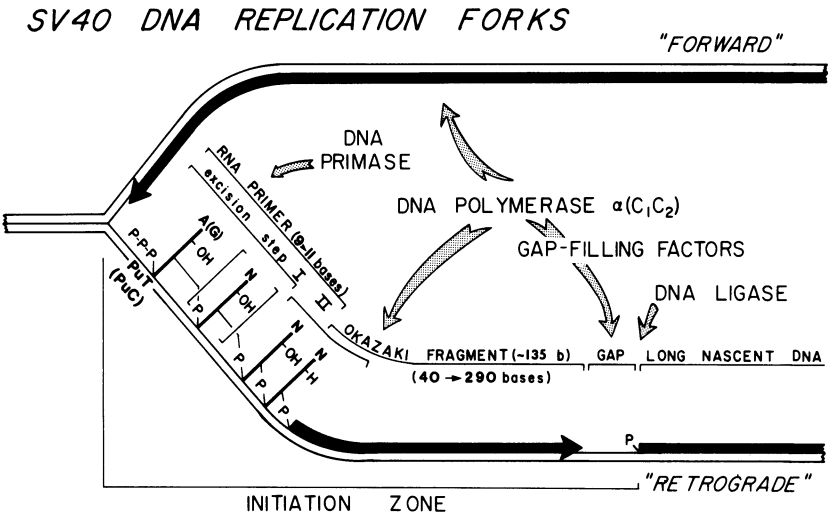


Figure 1. SV40 replication fork - DNA, RNA and enzyme components.

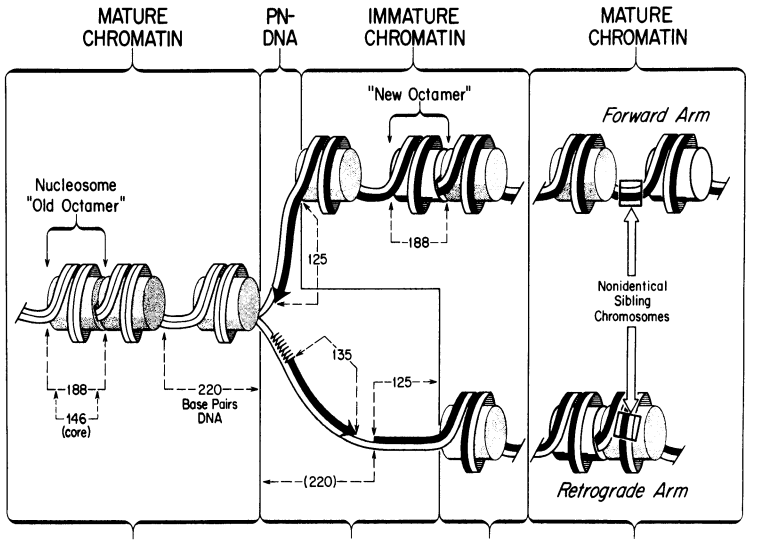


Figure 2. SV40 replication fork - nucleosome arrangement.

the Okazaki fragments contain iRNA covalently attached to their 5'-end (reviewed in 1,8,12). Approximately 3/4 of these RNA primers are intact (i.e. begin with 5'-(p)pprN), while the remainder are partially degraded. The structure of Okazaki fragments, their distribution at replication forks, and their validity as transient intermediates in viral DNA replication are well documented; Okazaki fragments do not result from damage and repair of nascent DNA. SV40 Okazaki fragments originate predominantly, if not exclusively, from retrograde templates. Therefore, DNA synthesis on the forward arm is a relatively continuous process. Recent experiments in which PyV and SV40 replicating DNA was analyzed in parallel by the same techniques revealed that at least 90% of PyV Okazaki fragments also originate from retrograde templates (15; unpublished data).

Theoretically, Okazaki fragments must originate from retrograde templates, whereas reinitiation of DNA synthesis on forward templates is not required and, in fact, may be prohibited. Once DNA synthesis begins at the origin of replication, synthesis in the same direction as fork movement could continue until termination of replication occurs. If replication forks advance faster than DNA synthesis can keep pace, then Okazaki fragments may be initiated on forward as well as retrograde templates. However, if the enzyme complex responsible for initiation of Okazaki fragments in eukaryotes behaves like the *E. coli* primasome and slides along the DNA template in a 5' to 3' direction (16), it could only operate on the retrograde template where the direction of sliding is toward the fork. On the forward template, the initiation complex would collide with DNA polymerase coming in the opposite direction. In fact, DNA primase-DNA polymerase-α strongly favors initiating DNA synthesis on the 3'-end of preformed RNA or DNA primers over *de novo* initiation of a new primer (17). Therefore, once DNA synthesis is initiated on the forward arm, further RNA-primed initiation events on that side of the fork are not favored. Even if the DNA polymerase falls off, synthesis is more likely to reinitiate on the 3'-end of the nascent DNA chain.

DNA Primase-DNA Polymerases α and δ

Historically, the major DNA polymerase activity in mammalian cells has been identified as DNA polymerase-α, a multi-protein enzyme with subunits ranging from about 15,000 to 250,000 daltons, depending on the method of purification (21,51-55). However, DNA polymerase-δ, generally found in relatively small amounts, is similar to DNA polymerase-α in its size and subunit composition, its sensitivity to aphidicolin, N-ethylmalimide, ara-CTP and ara-ATP, its resistance to ddTTP, its stimulation by ATP on homopolymer primer-templates, and its processivity. DNA polymerase-δ is distinguished from DNA polymerase-α by its tight association with a 3'-5' exonuclease, its preference for homopolymer primer-templates over DNase I activated DNA, its chromatographic properties, and its resistance to inhibition by p-n-butylphenyl-dGTP (BP-dGTP) and by some monoclonal antibodies directed against α-polymerase (39,40,57-59).

Previously published data supporting DNA polymerase–α as the enzyme solely responsible for DNA synthesis during SV40, polyoma virus and mammalian chromosome replication (1,11) are also consistent with DNA polymerase–δ. First, with the exception of BP-dGTP, DNA polymerases–α and –δ are remarkably similar in their sensitivities to inhibitors. Second, in virtually all of the studies attempting to correlate α–polymerase activity with DNA replication activity, δ–polymerase activity was not considered. Third, the most direct evidence that DNA polymerase–α is involved in SV40 DNA replication comes from elimination of α–polymerase activity in cellular extracts either by treatment with N-ethylmalimide (41) or by passing them over immobilized monoclonal antibody (SJK-287) directed against α–polymerase (12) and then demonstrating that only purified α–polymerase will restore activity. However, SJK-287 partially crossreacts with δ–polymerase (57), although a related antibody, SJK 132, does not (40), and the absence of δ–polymerase in the purified preparations of α–polymerase was not examined.

Direct support for the involvement of DNA polymerase–δ in DNA replication comes from the observation that initiation and continuation of DNA replication in SV40 chromosomes, like DNA synthesis by DNA polymerase–δ, is 1000 times more resistant to BU-dGTP than DNA polymerase–α (25). An alternative explanation is that a complex form of DNA polymerase–α is involved in replication that may be resistant to BP-dGTP. However, the largest form of α–polymerase from HeLa cells is still sensitive to BP-dGTP (E. Baril, personal commun.). Perhaps the enzymes' sensitivity to BP-dGTP is a function of structure of the template as well as the catalytic subunit. For example, utilization of dNTP substrates by DNA polymerase–α varies 200-fold as a function of the template utilized (62), and SV40 *ori*-dependent plasmid DNA replication is 10 times less sensitive to BP-dGTP than SV40 chromatin replication (25). Unfortunately, the more complex DNA primase-DNA polymerase–α, like simpler forms of this enzyme, fails to select the same DNA primase initiation sites used *in vivo*, and synthesizes RNA primers about half the size of those synthesized in vivo (63). The properties of DNA primase-DNA polymerase–δ remain to be examined in similar detail.

DNA polymerase-α can be isolated as a multi-protein complex containing DNA primase and DNase activity, an Ap_4A binding protein, and primer recognition proteins C_1C_2 (17,27-30). DNA-dependent ATPase activity, RNase H, and topoisomerase II activities are associated with α-polymerase at earlier steps in its purification (29-31). Ap_4A has been implicated in the initiation of mammalian nuclear DNA replication (32), but it does not stimulate initiation of DNA replication in SV40 chromosomes in vitro (25). RNase H is an obvious candidate for excision of the RNA portion of RNA primers, but it does not excise rN-p-dN covalent linkages (34).

Topoisomerase II may be involved specifically in the termination of DNA replication (see "Termination"). The DNA-dependent ATPase activity suggests the presence of a helicase activity.

The most detailed studies so far have focused on DNA primase-DNA polymerase-α and its cofactors C_1 and C_2 (12). The C_1C_2 complex specifically stimulates (180 to 1800-fold) α-polymerase to incorporate the first dNTP onto an RNA or DNA primer associated with a DNA template of at least 50 nucleotides or more, but C_1C_2 has no effect on α-polymerase activity with shorter DNA templates. C_1C_2 also has no effect on dNTP Km values, recognition of DNA template signals that arrest α-polymerase, and α-polymerase processivity (9-11 bases). C_1C_2 specifically reduce the Km of the primer itself. Therefore, C_1C_2 increases the ability of α-polymerase to find a primer and insert the first nucleotide. The fact that extensive single-stranded (ss) DNA primer-templates are better substrates for DNA polymerase-α[C_1C_2] than DNase I activated DNA is for DNA polymerase-α alone, demonstrates that ssDNA participates in the reaction, presumably by allowing the enzyme to slide along the template until it finds a primer. In contrast, ssDNA inhibites α-polymerase, demonstrating that binding is nonproductive, and that the enzyme must continually dissociate and rebind in order to locate a primer. C_1C_2 may also be involved in initiation of iRNA synthesis.

Purified DNA primase-DNA polymerase-α initiates DNA synthesis *de novo* by synthesizing a short oligoribonucleotide on a bare ssDNA template that serves as a primer for DNA synthesis. Table 1 compares the properties of this reaction with those of the endogenous "replicase" associated with SV40 or PyV replicating chromosomes *in vivo* and in subcellular replication systems. iRNA synthesis by either system is resistant to α–amanitin and aphidicolin, initiates primarily with ATP and secondarily with GTP depending on the ratio of ATP/GTP in the reaction mixture, changes to DNA synthesis at an unspecified sequence, is template-dependent, and results in oligoribonucleotides of similar, but not identical, size.

One major difference between iRNA synthesis on purified DNA templates and natural replicating DNA intermediates is that purified DNA primase-DNA polymerase-α initiates synthesis at pyrimidine-rich sequences that occur about once every 25 bases, while the endogenous replicase activity initiates synthesis at purine-T and purine-C sites that occur about once every 7 nucleotides. Otherwise, there is no identifiable sequence preference. In fact, all 16 possible rN-p-dN covalent linkages are represented *in vivo* at frequencies that suggest a near random distribution of initiation sites on the DNA template. Replicase and primase-polymerase initiation sites were shown to be distinctly different by mapping them on the same templates (28,46). None of the sites selected in the SV40 ori region during viral DNA replication in CV-1

cells corresponded to any of the sites selected by CV-1 cell DNA primase-DNA polymerase-α *in vitro*. Changing the ratio of ATP/GTP in the reaction mixture favored initiation at sites that used the ribonucleotide in highest concentration but the enzyme still selected sites from the same family of template sequences (17,28). Similarly, a more complex form of the enzyme that included C_1C_2 inherently preferred to initiate with ATP, but still selected pyrimidine-rich sites distinctly

Table 1. Synthesis of RNA-primed DNA chains on natural replicating chromosomes (RC) compared with synthesis by purifed DNA primase-DNA polymerase-a on ssDNA templates.

RNA Primer Characteristics	PyV(RC)[1]	SV40(RC)[2]	DNA pol-a DNA primase[3]	DNA pol-a DNA primase C_1C_2[4]
Primase Inhibited by:				
α-amanitin	no	no	no	no
aphidicolin		no	no	no
5'-ppprN (%)				
A	80	70	70	70
G	20	30	30	30
C	0	0	0	0
U	0	0	0	0
A/G Starts = ATP/GTP Ratio	yes		yes	
rN-p-dN	'random'	'random'	'random'	
Complementary to DNA Template		yes	yes	
DNA Template Initiation Sites[5]				
retrograde arm	90%	95%	---------- no preference ----------	
sequence				
primary (80%)		3'-PuT	-------- 3'-(Py)nCTTT(Py)n --------	
secondary (20%)		3'-PuC	-------- 3'-(Py)nCCC(Py)n --------	
sites/100 bases		7 - 20	3 - 8	
Size (bases)[6]				
range	5-12	2-12	1-9	1-5
peak	10	10	7	3
Length Heterogeneity at Unique Site		yes	1-5 bases	1-3 bases

Data were taken from: [1] refs 35-39; [2] refs 15,25,40-49; [3]refs 17,28, 50; [4] ref 50. [5]**Bold** deoxyribonucleotide in template encodes first ribonucleotide of primer. [6]Intact primers containing 5'-(p)ppN.

different form those used *in vivo* (50). Selection of initiation sites *in vitro* depends on the

template DNA sequence, ATP/GTP ratio, template secondary structure (51), and enzyme complexity.

A second difference is that the complex form of α-polymerase made RNA primers that were about 70% shorter than those in replicating viral DNA and about 50% shorter than those made by DNA primase-DNA polymerase-α alone. Although the sizes of RNA primers produced by the same enzyme varied considerably among different initiation sites, the more complex form of DNA primase-DNA polymerase-α produced shorter primers than DNA primase-DNA polymerase-α alone even at the same nucleotide initiation site. This difference between the two enzymes appears to be due to the C_1C_2 complex. Since C_1C_2 allows α-polymerase to use preformed primers as short as 2 nucleotides while α-polymerase alone requires primers of 8 to 10 nucleotides for maximal activity (27), C_1C_2 may increase the rate at which RNA synthesis switches to DNA synthesis.

Since DNA polymerase-α is the most abundant polymerase in the cell, it seems reasonable that DNA primase-DNA polymerase-α is the enyzme responsible for initiation of DNA synthesis at replication forks, However, either the enzyme isolated so far is missing an important factor, or purified ssDNA is not appropriate as a template for initiation studies unless associated with the appropriate binding proteins. Alternatively, DNA primase at replication forks may be associated with a different DNA polymerase such as DNA polymerase-δ.

Chromatin Structure and Assembly at Replication Forks (Fig. 2)

The structure of chromatin at SV40 DNA replication forks (and apparently cellular and PyV forks as well) can be divided into at least four distinct domains based on both nuclease digestion and electron microscopic analyses (reviewed in 1,8,12). Pre-replicative, mature chromatin in front of replication forks is equivalent to mature unreplicated chromatin whose histone composition and nucleosome structure is indistinguishable from the host cell, and nucleosomes are separated by highly variable regions of internucleosomal DNA. Recent electron microscopic analysis of DNA from psoralen crosslinked SV40 chromosomes observed an average of 27±2 nucleosomes per genome (53), whereas previous EM observations of viral chromosomes observed 24±2 nucleosomes per genome (173). The actual sites of DNA synthesis is free of nucleosomes and therefore referred to as pre-nucleosomal (PN) DNA. Thus, the enzymes responsible for DNA synthesis utilize non-nucleosomal DNA templates, although their relative resistance to single-strand specific endonucleases indicates the presence of bound proteins (151). Based on nuclease digestion of nascent DNA, PN-DNA consists of an average of 125 bp of newly replicated DNA on the forward arm and one Okazaki fragment plus an average of 125 bp of the newly replicated daughter duplex DNA on the retrograde arm. Crosslinking DNA in SV40 replicating chromosomes with trimethyl-psoralen, either *in vivo* or *in vitro*, revealed that the

distance from the branch point in replication forks to the first nucleosome is 225±145 nucleotides on the forward arm and 285±120 nucleotides on the retrograde arm (53). Thus, newly replicated DNA is organized into nucleosomes as rapidly as sufficient double-stranded DNA becomes available, but the initial structure is immature chromatin because it is hypersensitive to nonspecific endonucleases. Post-replicative, mature chromatin consists of newly assembled nucleosomes that cannot be distinguished from mature chromatin. The structure, spacing and phasing with respect to DNA sequence of nucleosomes on replicating viral chromosomes is indistinguishable from mature viral chromosomes.

"Old" pre-fork histone octamers are distributed in an apparently random fashion to both arms of the fork, new histone octamers are assembled on both arms, and chromatin maturation occurs on both arms. Although the fate of "old" histones in SV40 was studied only in the presence of cycloheximide to block synthesis of new histones, the results are in agreement with studies on cellular chromatin, some of which were done in the absence of protein synthesis inhibitors (reviewed in 52,53). Nucleosomes do not occupy the same DNA sites on all chromosomes, but many different preferred sites that approximate a 'random' distribution of nucleosomes with respect to DNA sequence (nucleosome phasing). Nucleosome phasing on one arm of a single replication fork is not identical to phasing on the other arm, suggesting that, at least in virus chromosomes, nucleosome phasing in front of replication forks is not maintained during replication. Nucleosome assembly during *in vitro* DNA replication occurs preferentially at active replication forks, suggesting that single-stranded regions at replication forks effectively relieve torsional strain that forms during nucleosome assembly (54).

The nuclease hypersensitive region in the SV40 ori-enhancer segment, which appears in only 20% to 25% of viral chromosomes (reviewed in 53), is the only region where some nucleosome phasing occurs. This phasing does not during replication because it is absent from newly replicated chromatin (55,56), and chromatin-specific hypersensitive sites are assembled on histone genes injected into *Xenopus* oocytes where DNA replication does not occur (57). However, the extent of chromatin hypersensitivity in the SV40 ori-enhancer region was greater in chromosomes that underwent replication than in those that did not (58). Therefore, the presence of immature chromatin in newly replicated chromosomes may facilitate subsequent alteration of chromatin structure by sequence-specific proteins.

"Initiation Zone" Model for Replication Forks

The data available for SV40 and PyV DNA replication forks can be interpreted on the basis of a simple model that has been described previously (1,8,12,41,45,46). As replication forks advance, continuous DNA synthesis maintains the forward arm as double-stranded (ds) DNA while a ssDNA region is exposed on the retrograde arm that acts as an "initiation zone" for the synthesis of Okazaki fragments. Assuming that nucleosome disassembly in front of the fork is the rate limiting step in fork movement, the size of this initiation zone should be 220±73

nucleotides, the average distance from one nucleosomal core to the next. Both the size and sequence of initiation zones vary extensively among individual replicating chromosomes because nucleosomes are arranged in a near-random fashion with respect to DNA sequence.

Since DNA synthesis at replication forks, including synthesis and completion ("gap-filling") of Okazaki fragments (33), can be inhibited by aphidicolin but not by ddTTP, DNA polymerase-α (or conceivably DNA polymerase-δ) is solely responsible for DNA synthesis at SV40 replication forks. Therefore, by some stochastic process, DNA primase-DNA polymerase-α[C_1C_2] selects one of many possible sites on the template within the initiation zone to initiate synthesis of a short iRNA on whose 3'-OH end DNA polymerase-α rapidly initiates DNA synthesis. Initiation events occur 80% of the time at 3'-dPuT and 20% of the time at 3'-dPuC in the template, with the first ribonucleotide complementary to either dT or dC. Although these sites occur, on average, once every 7 nucleotides, initiation events occur only once every 135 bases (the average size of mature Okazaki fragments (41)). Therefore, once synthesis of an iRNA is initiated, elongation of the resulting nascent chain is rapid enough to prevent additional initiation events downstream. Neither the template sequence encoding the iRNA, nor the transition point in the template where RNA synthesis changes to DNA synthesis reveals sequence preference. The transition from RNA to DNA synthesis generally occurs 9-11 bases downstream, but it can vary from 2-12 bases, depending on the template initiation site. DNA synthesis continues until 10-15 nucleotides remain to be incorporated whereupon one or more protein "gap-filling factors" are required to allow α-polymerase to complete DNA synthesis. DNA ligase then seals the 3'-end of the Okazaki fragment to the 5'-end of the long, nascent DNA strand. Okazaki fragments that have completed their synthesis range in size from 40 bases (those that initiated closest to the growing daughter strand) to 290 bases long (those that initiated closest to the replication fork). Thus, the structure of a typical RNA-primed Okazaki fragment is (p)ppA/G$(pN)_{7-9}$pN-p-dN$(pN)_{40-290}$.

Excision of iRNA occurs concurrently with DNA synthesis and ligation. RNA primers are excised at the same rate at which Okazaki fragments are ligated to the 5'-ends of long nascent DNA chains. Removal of the bulk of the primer does not require concomitant DNA synthesis whereas removal of the RNA-p-DNA junction is facilitated by DNA synthesis. Excision of RNA primers does not stop at the RNA-DNA junction, but removes a variable number of residues from the 5'-end of the DNA chain as well (46). Thus, an RNase H activity could remove the bulk of the primer, but the second step must involve an exonuclease that degrades both RNA and DNA. The 5'-end of the degraded Okazaki fragment becomes the 5'-end of the long nascent DNA strand, the fork advances to another nucleosome, a new Okazaki fragment initiation zone is exposed and the process begins again.

The essence of this model is that the periodic structure of chromatin in front of replication

forks, rather than unique DNA sequence signals, dictates the repeated initiation of Okazaki fragments an average of once every 135 bases and limits their size to about 300 bases. The only requirement for unique cis-acting sequences in DNA replication is ori.

Initiation of Replication - "Requirements"

Initiation of viral DNA replication requires the interaction of virally encoded large tumor antigen (T-Ag) with the cis-acting sequence from the same virus that functions as the *genetic origin of replication* (ori), and one or more factors from appropriate permissive cells. The result is bidirectional replication from the *origin of bidirectional replication* (OBR). Based on the initial observations of Li and Kelly (47), soluble systems have been developed that can initiate replication in circular double-stranded DNA molecules containing either the SV40 (24,33,49,54,64,65,121,138) or PyV (59) ori. DNA replicates only if it contains a functional ori in the presence of high concentrations of the corresponding T-Ag and permissive cell extract. DNA replication is bidirectional from ori, and multiple rounds of replication occur. One of these systems (33,49,138) initiates bidirectional replication at ori in SV40 chromosomes as well as purified SV40 ori-containing plasmid DNA. Linear duplex DNA is at least at least 10X less efficient than circular DNA as substrates in these systems, suggesting that activation of ori, requires negative superhelical turns. This could explain why initiation occurs preferentially in circular viral DNA molecules with the highest negative superhelical density (139). Interestingly, mature SV40 chromosomes are not completely relaxed despite the presence of nucleosomes and associated topoisomerase I and II activities (172), but contain about 2 additional negative superhelical turns (126,140).

Origin of Replication (Fig. 3)

Analysis of deletions and substitutions in and around the SV40 ori region reveal three elements: a 64 bp sequence that is absolutely required for replication (ori-core, nucleotides 5209 to 29), and two 43-45 bp auxiliary sequences (aux-1, 5164 to 5208; aux-2, 30 to 72), one on each side of ori-core that facilitate replication 2-5 fold (reviewed in 12). Single bp deletions have now defined ori-core as the 64 bp from positions 5211 to 31 (63). Mapping the nucleotide locations of 5'-ends of RNA-primed nascent DNA chains in SV40 replicating DNA has identified the OBR (the transition from discontinuous to continuous DNA synthesis on each strand of ori) at the junction between T-Ag site 1 and ori-core (5210 to 5211; arrows on replication origin in Fig. 4 mark beginning of continuous DNA synthesis). Ori function is markedly affected by the number of ori copies, their proximity to one another, their relative orientations, and the presence of intervening sequences (62). When both aux elements are absent, replication is reduced about 20-fold (60). Aux-1 contains the strongest T-Ag binding site, and aux-2 contains the G/C-rich repeats and part of the early gene promoter. Aux-1 cannot substitute for aux-2 (60). However,

in the absence of the 21 bp repeats, the 72-bp repeats, which contain the enhancer elements, can substitute for aux-2 with about 50% effectiveness (60,61,64,75); replication is reduced about 100-fold in the absence of both 72-bp repeats and aux-2 (75). Aux-2 function depends on its distance and orientation with respect to ori-core, and no replication is detected if aux-2 is placed on the early-gene side of ori-core (61). Therefore, activation of ori-core is greatly facilitated by an adjacent transcriptional element on its late gene side. In vitro SV40 DNA replication also requires ori-core, and is facilitated by aux-1 (about 5192 to 5216) but not by aux-2 (64,65).

The PyV ori consists of a 66 bp ori-core (nucleotides 44 to 5274 in strain A3; numbering system is in the opposite direction from SV40), and two additional cis-acting sequences [α (5073 to 5126) and β (5147 to 5218)] that map within the borders of gene enhancer elements 2 (5073 to 5130) and 3 (5131 to 5229) (reviewed in 12,66-69). The minimal ('core') α and β sequences (5108 to 5126 and 5172 to 5202, respectively) are 5-10X less active than the complete elements (J. Hassell, unpublished data). As discussed later, at least one enhancer element is required to activate PyV ori-core. The PyV ori-core is structurally homologous to the SV40 ori-core: each contains a 17bp (SV40) or 15 bp A/T rich sequence in which 77% and 79%, respectively, of the T residues reside in one strand, each contains a palindromic sequence (13 bp, 77% G/C rich, in SV40; 15 bp, 73% G/C rich in PyV), and each contains a block of 16 (SV40) to 18 (PyV) residues in which 88% to 100%, respectively, of the purines are in one strand (Fig. 5). The locations of 'GAGGC' boxes (see "T-antigen") are similar in the two ori-core regions, but not identical. The PyV OBR has been located in ori-core by mapping rN-p-dN covalent linkages in PyV replicating DNA from PyV-infected 3T3 cells (15). The transition point between discontinuous and continuous DNA synthesis on the template encoding late mRNA lies at the junction between ori-core and a strong DNA binding site for T-Ag; virtually the same nucleotide location as seen on the template encoding late mRNA in SV40 (Fig. 5). However, the transition point on the early mRNA template of PyV lies 18-20 bp further into ori-core than it does in SV40. The striking similarity between the SV40 and PyV OBR suggests that the OBR is determined by the sequences defining ori-core rather than strong T-Ag DNA binding sites, transcriptional elements, or mRNA initiation sites.

T-antigen

Large T-Ag is a multifunctional regulatory protein synthesized by papovaviruses early after infection of the host cell and found predominantly in the nucleus (reviewed in 12). There are 708 amino acids in SV40 T-Ag and 785 amino acids in PyV T-Ag, consistent with molecular weights of 81.5 Kd and 89.0 Kd, respectively. However, purified SV40 T-Ag is 88-94 Kd and PyV T-Ag is 94-100 Kd; the larger numbers apparently result from post-translational modifications and proteolytic alterations. T-Ag is absolutely required for initiation of viral DNA replication.

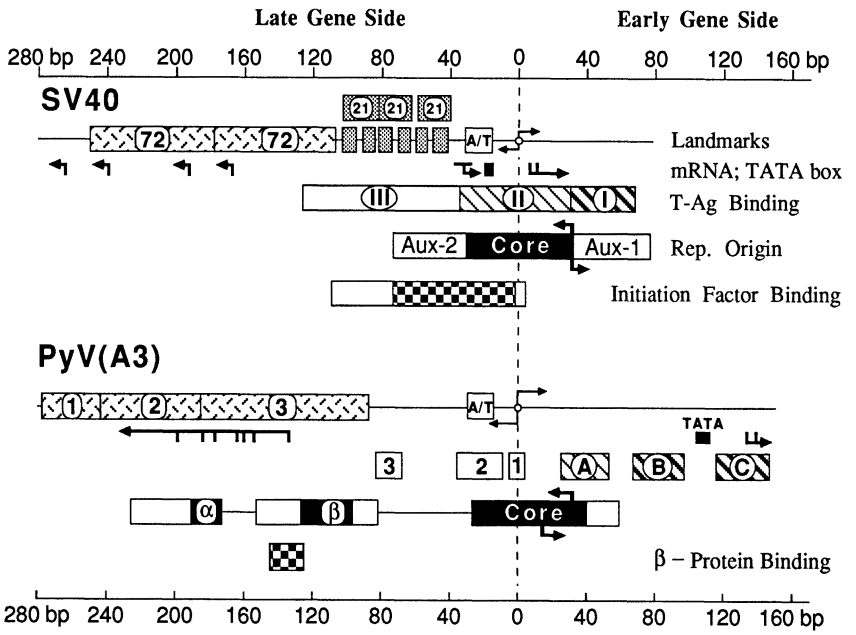


Figure 3. Comparison of SV40 and PyV ori-regions.

In addition to its role in cell transformation, T-Ag also induces replication of cellular DNA, synthesis of cellular mRNA and ribosomal RNA, and regulates viral mRNA synthesis. Purified T-Ag binds DNA (see below) and nucleotides (70), hydrolyzes ATP (71), interacts with a 53 Kd cellular protein (72,73), exhibits DNA helicase activity (74), and binds DNA polymerase–α (101). Temperature-sensitive mutations in the gene encoding T-Ag (tsA) affecting the initiation of viral DNA replication also produce defects in most of these *in vitro* activities.

T-Ag is required to initiate viral DNA replication, but its precise function(s) is not clear. T-Ag association with p53 does not appear to be required because monoclonal antibodies directed against p53 do not prevent initiation of SV40 DNA replication (R. Possenti, unpublished data), although they do prevent G_0 cells from entering S-phase and SV40 stimulation of cellular DNA replication (76). T-Ag does not appear to be required at ongoing viral replication forks because replicating intermediates in cells that were infected with tsA mutants (i.e. thermolabile T-Ag) complete replication at the restrictive temperature although new rounds of viral DNA replication are not initiated (77-80). Monoclonal antibodies directed against different SV40 T-Ag epitopes inhibit initiation of SV40 DNA replication *in vitro* with no apparent effect on replication forks at

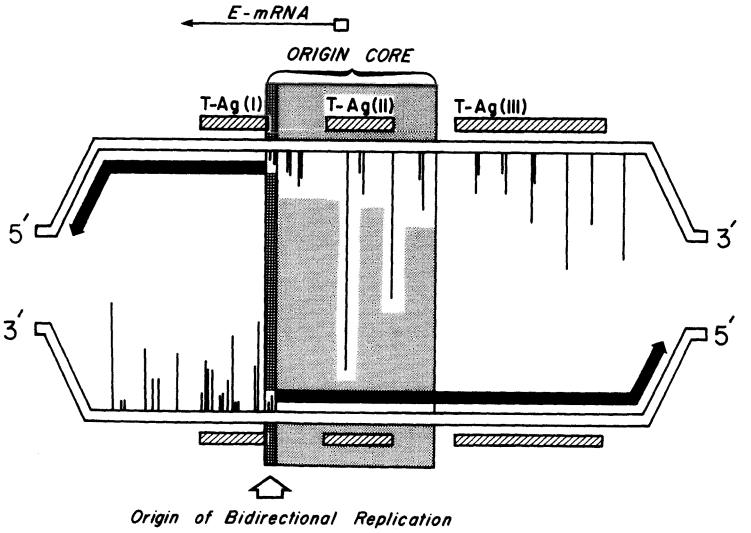


Figure 4. RNA-DNA covalent linkages in the ori region of SV40 replicating DNA intermediates (45,46). The transition from initiation events (vertical bars; relative lengths are proportional to their abundance) to the absence of initiation events (solid arrows) on each side of ori defines the origin of bidirectional DNA replication.

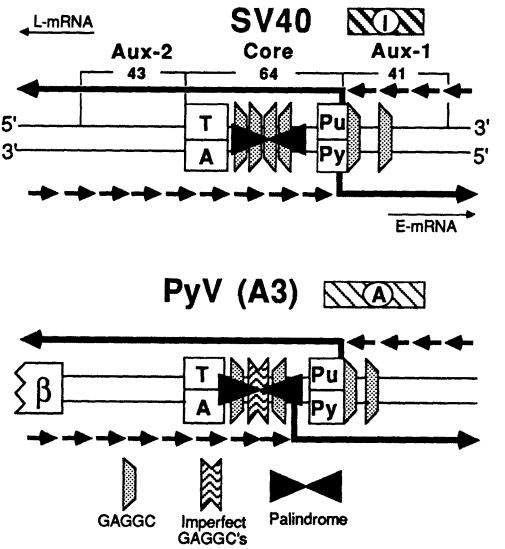


Figure 5. Comparison of SV40 and PyV ori-core regions. The origin of bidirectional replication is indicated by the broken and continuous arrows on each side of ori. Broken portion of arrow is region where RNA-primed DNA synthesis initiation events map.

non-ori regions (R. Possenti, unpublished data), although one monoclonal anti-T-ag antibody does inhibit DNA synthesis SV40 replicative intermediates (81). It was suggested that T-Ag helicase activity may be required at replication forks (74). However, since the association of T-Ag with replicating SV40 chromosomes is lost during the final stages of replication (82), it seems unlikely that a functional T-Ag is required for replication once initiation has occurred.
T-Ag has both nonspecific and specific binding affinities for DNA, as reflected in its affinity for

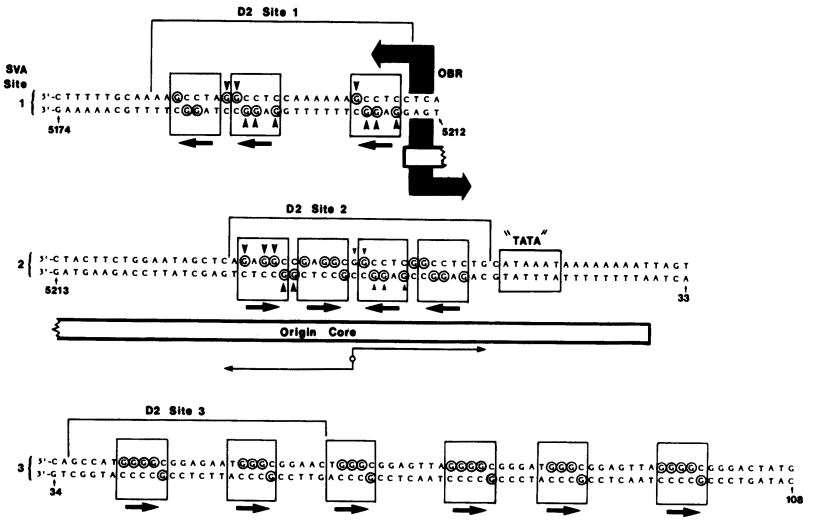


Figure 6. SV40 T-antigen DNA binding sites. SVA sites 1, 2 and 3 were determined with wild-type T-Ag (86,87,89,100,134-136) and D2 sites 1, 2 and 3 were determined with an analog of T-Ag containing about 10K of an Ad2 protein and about 90K of T-Ag (85,86,96,97). Guanines protected by bound T-Ag against methylation are circled. Guanines whose methylation interferes with T-Ag binding are indicated by arrowheads; large arrowheads indicate strong sites. The pentanucleotide sequence 5'-(G>T)(A>G)GGC-3' is boxed with an arrow to indicate 5'-3' polarity. The ori-core, origin of bidirectional replication (OBR), TATA box, and 27 bp palindrome are also indicated.

dsDNA and ssDNA under various conditions of pH and ionic strength. Only about 10% of the total T-Ag pool is able to bind viral ori-region DNA specifically, and only about 1% of the T-Ag pool is associated with SV40 chromosomes in infected cells (83). Alterations of T-Ag that

reduce its affinity for the ori-region eliminate its ability to replicate DNA (12). Low concentrations of ATP and other nucleotides that bind to T-Ag inhibit binding of T-Ag to SV40 ori (84). Surprisingly, high ATP concentrations are optimal for initiation of RNA-primed DNA synthesis by DNA primase-DNA polymerase-α, indicating a conflict of interests that could control the rate of viral DNA replication. Both SV40 and PyV T-Ag binds to the consensus pentanucleotide 5'-(G>T)(A>G)GGC-3' located within specific T-Ag DNA binding sites in and around their respective ori regions (85-88). The degree of T-Ag affinity for these "GAGGC" sequences depends on their number, proximity, relative orientation and spacing (89-91). PyV DNA high affinity binding sites A, B and C all lie to the early gene sided of ori; a 10-fold higher concentration of PyV T-Ag is required to detect viral DNA binding sites within ori-core (88,92,93,134). The poor binding of purified PyV T-Ag to PyV ori-core, which may reflect the absence of perfect GAGGC repeats (Fig. 5), represents a striking difference with SV40. Purified SV40 T-Ag binds most strongly to the early gene side of ori-core (site 1), 5-7 times less strongly to ori-core (site 2, ref. 94), and marginally to aux-2 and beyond (site 3; Figs. 3,6). Mutations within sites 1 and 2 alter the affinity of T-Ag for DNA (89,95-97). These sites are biologically significant because mutations within ori that reduce or prevent replication can be suppressed by mutations in T-Ag (98), and mutations in site 1 that reduce the ability of T-Ag to regulate its own synthesis are overcome by mutations in T-Ag (95). Since most of sites 1 and 3 can be deleted with only a marginal effect on replication, initiation of DNA replication must involve binding of T-Ag to site 2 (ori-core), perhaps facilitated by the strong affinity of T-Ag to site 1. Thus, one might expect ori-core to be the primary target for the SV40 initiation complex. However, at least one factor required for initiation of SV40 DNA interacts with sequences outside T-Ag binding sites 1 and 2, and outside ori-core (see below).

T-Ag has 8-10 phosphorylated sites at seryl and threonyl residues localized in two clusters on opposite ends of the polypeptide chain (99,100). Enzymatic dephosphorylation of purified T-Ag markedly stimulates its ability to replicate SV40 DNA *in vitro*, and increases its affinity for ori-core (Y. Gluzman, personal communication). This is consistent with previous observations that the DNA-binding properties of the protein were highest when the peptide was relately new and the phosphate content was low (102), particularly at N-terminal sites (103), and that phosphorylation increased with time (104,105). T-Ag has also been reported to be glycosylated (106), adenylated (107), ADP ribosylated (108), and acetylated (109). These modifications, which appear to affect only a small fraction of the available protein and may differ from one T-Ag molecule to the next, may serve to modulate T-Ag function and/or direct the protein to various cellular locations.

Purified T-Ag exists in several oligomeric forms that differ in their enzymatic activities and DNA binding properties (110-113). However, only the purified monomer is active in replicating SV40 DNA *in vitro* (B. Weiner, unpublished data). This form of T-Ag is the most efficient at binding to SV40 ori, is underphosphorylated, and is devoid of ATPase activity in either the

presence or absence of SV40 DNA. Therefore, assuming that T-Ag ATPase activity is required for initiation of viral DNA replication [e.g. helicase activity (74)], T-Ag monomer must form a complex with ori in the presence of permissive cell factors that reconstitutes the ATPase activity normally associated only with multimeric forms of T-Ag. In fact, the conformation of T-Ag bound to replicating SV40 chromosomes differs significantly from unbound T-Ag (83), and a stable pre-elongation complex between SV40 T-Ag, ori and cellular factors can be isolated and subsequently elongated in the presence of added nucleotides (24). Whether or not this complex contains T-Ag specific ATPase activity remains to be determined.

Permissive Cell Factors and Ori-Core

Even with sensitive methods of detection, SV40 (115) and polyoma (116,117) ori-dependent, T-Ag dependent DNA replication has not been observed in nonpermissive fibroblast cells. Both require soluble factors that are present only in permissive cells during S-phase, and that appear to act in a positive manner, rather than by relieving inhibition by a nonpermissive cell repressor. There are three basic mechanisms by which permissive cell factors can control initiation of viral DNA replication: (i) regulate the concentration of T-Ag at the level of transcription or translation, (ii) modify T-Ag post-translationally to an active or stable form, and (iii) participate directly by forming a T-ag-initiation factor(s)-ori complex (reviewed in 12). Although little direct information is yet available, the third hypothesis seems the most likely.

Permissivity for papovavirus DNA replication lies in the stringent specificity of interactions between T-Ag, ori-core and permissive cell factors. Bennett et al. (118) showed that a plasmid carrying the SV40 ori-core sequence replicates only in monkey cells expressing SV40 T-Ag from an integrated SV40 ori$^-$ genome (COS cells), and not in mouse cells expressing SV40 T-Ag (MOS cells) or in monkey cells expressing PyV T-Ag (COP cells). This host preference was not affected by the presence or absence of SV40 or PyV enhancer sequences. Similarly, a plasmid carrying the PyV ori replicated only in mouse cells expressing PyV T-Ag (MOP cells), regardless of the presence of either the normal PyV enhancer region or the SV40 enhancer-promoter sequence (72 bp repeats + 21 bp repeats). The SV40 enhancer-promoter region allows PyV T-ag-dependent, ori-core dependent DNA replication in mouse differentiated (116,117,119,120) and embryonic (D. Wirak, unpublished data) cells. Therefore, although enhancer elements are required for activation of PyV ori-core in mouse cells, *enhancer elements do not alter species-specificity although they can alter cell-specificity* (see below). Murakami et al. (59,121) have reported that extracts of either permissive or nonpermissive mammalian cells will support SV40 ori-dependent DNA replication when SV40 T-Ag and either HeLa or monkey DNA primase-DNA polymerase-α are present, but not when DNA primase-DNA polymerase-α is taken from nonpermissive cells. The same is true for PyV ori-dependent replication which requires PyV T-Ag and DNA primase-DNA polymerase-α from mouse cells. These data, together with the affinity of T-Ag for DNA polymerase-α (101), suggest that DNA primase-DNA

polymerase-α, or a factor that binds to this enzyme, is the permissive cell factor with which T-Ag interacts.

One problem in concluding that enhancer elements cannot change species-specificity is that the level of T-Ag produced in MOS or COP cells may be insufficient to support SV40 and PyV ori-dependent replication in a nonpermissive environment. deVilliers et al. (119) observed that plasmids carrying a PyV genome replicated in monkey or human cells if the SV40 72 bp repeats were substituted for the normal PyV enhancer region. However, this observation was not confirmed using either cloned PyV genomes or reconstructed PyV virions (116,117). Furthermore, substitution of the PyV enhancer for the SV40 enhancer allowed efficient SV40 T-Ag production in mouse cells, but did not result in SV40 DNA replication, suggesting that an inadequate concentration of viral T-Ag was not responsible for failure of SV40 DNA replication in nonpermissive cells (122). Using an analogous approach, Y. Gluzmann and coworkers (unpublished data) constructed a mouse cell line that produces 5-times more SV40 T-Ag than normally occurs during SV40-infection of monkey cells, but they could not detect replication of SV40 DNA. It is possible that mouse cells cannot modify SV40 T-Ag to its "active" form, but this seems unlikely because of the similarities between SV40 and PyV T-Ag (12), and the fact that low levels of SV40 ori-dependent, T-Ag dependent DNA replication has been observed in mouse preimplantation embryos (123). This question may be answered when SV40 T-Ag is raised in nonpermissive cells (e.g. mouse or *E. coli)* and tested for its activity in cell free systems that initiate SV40 DNA replication.

Permissive Cell Factors and Enhancer Elements

Cell-specific activation and suppression of PyV ori-core by cis-acting enhancer elements provides a second line of evidence demonstrating a direct interaction between permissive cell factors and ori (Table 2; Fig. 3). At least three different configurations of the PyV ori (α–β-core, α-core, and β–core) as well as two alien configurations (α–MuLV-core, and 72s-core) can initiate PyV T-Ag dependent DNA replication in mouse fibroblasts, but not necessarily in other mouse cells. In these constructions, α is enhancer regions 1+2, and β is enhancer region 3 (Fig. 3). α–β-Core does not replicate in mouse lymphocytes, mast cells, or embryonal carcinoma (EC) cells unless it is modified. Removing the β-element allows replication in T-lymphocytes and mast cells but not in EC cells, revealing the ability of the β-element to suppress α-core replication in some cell types. Replacing the β-element with the Moloney murine leukemia virus enhancer extends the host range for DNA replication to lymphocytes and mast cells, whereas replacing β with the mouse immunoglobulin enhancer allows replication in lymphocytes, but suppresses the ability of α-core to replicate in fibroblast or mast cells. Similarly, the SV40 enhancer can also activate ori-core in fibroblasts, but only in the absence of the PyV α-element. Thus, cell-type

specific activation of ori-core in fibroblasts, as well as in embryonic cells (see below), is a function of associated enhancer element. However, some combinations of enhancer elements can interfere with each other in some cell types, resulting in the absence of ori activity.

In some cases, the PyV enhancer elements can be genetically altered to allow DNA replication in nonpermissive mouse cells. For example, mutations in PyV selected for growth on normally nonpermissive EC cells, Friend erythroleukemic cells and neuroblastoma cells are found in the α or β–elements, depending on the cell line used (5,124,125). All of these mutations still replicate in mouse 3T3 fibroblasts, but not necessarily in all other nonpermissive mouse cell lines. Therefore, assuming that one or more PyV permissive cell factors is an enhancer-binding protein, mouse fibroblasts must express all of them (or all their determinants on a single protein) while other mouse cells must contain only a subset. A protein has been identified in mouse cells that binds specifically to the β-element and may distinguish between wild-type and EC-mutant DNA (127,128; Fig. 3).

Table 2. The ability of various enhancer elements to activate PyV ori-core

Mouse Cells	Enhancer Elements								
	1+2+3	1+2	3	1+2+IgG	IgG	1+2+MuLV	1+2+F101	1+2+72s	72s
fibroblasts[1]	+	+	+	-	-	+	+	-	+
T-lymphocytes[2]	-	+		+		+			
B-lymphocytes[2]	-	-		+	+	+		-	
Mast cells[2]	-	+		-		-		-	
EC cells[3]	-	-		-		-	+		
1-cell embryos[4]	+	+	+						
2-cell embryos[4]	+	-	+				+		+

Symbols: +, replicates; -, little or no replication; 1, 2 and 3, PyV enhancer elements (Fig. 3); IgG, mouse immunoglobulin enhancer; MuLV, Moloney murine leukemia virus long terminal repeat; F101, PyV 3-element mutation selected for PyV growth in EC cells (Fig. 3); 72s, SV40 72 bp repeat enhancer elements (Fig. 3). Enhancer effects on DNA replication are cis-acting, and therefore are not do to a requirement for activation of T-Ag synthesis. Results were the same with enhancers in either orientation.

[1]Mouse 3T3 and MOP cell line data are from refs. 67,117,120, and129.
[2]Lymphocyte data are from ref. 117.
[3]Embryonal carcinoma PCC4-Aza and F9 cell line data are from refs. 117 and 125.
[4] Mouse 1-cell and 2-cell preimplantation embryos were injected with DNA intranuclearly and then allowed to continue development *in vitro* (14; D. Wirak, D. Cupo, E. Martinez-Salas, J. Hassell, and M. DePamphilis, unpublished data).

In the examples cited above, the ability of a cis-acting DNA sequence to activate PyV ori-core correlated with the ability of the same sequence to enhance gene expression in the same cell type, strongly suggesting that it is the enhancer activity in these sequences that is required to activate ori-core. This hypothesis is further strengthened by the fact that mouse 1-cell embryos can both

replicate DNA plasmids containing only the PyV ori-core and T-Ag (D. Cupo and E. Martinez-Salas, unpublished data), and fully activate the HSV tk promoter in the absence of an enhancer element (D. Wirak, unpublished data). Ori-core has never been observed to replicate alone in any other cell, and the HSV tk promoter requires an enhancer element for activity in other mouse cells, including 2-cell embryos. Thus, when a cell requires an enhancer to activate promoters, it requires the same activity to activate the PyV ori-core; those sequences that are recognized by a particular cell as enhancer elements will activate ori-core, and those sequences that are not recognized will not activate ori-core.

Permissive Cell Factors and Embryonic Cells

A further complexity in understanding regulation of DNA replication is the possibility that the rules for replicating DNA in embryonic cells may be different that those in differentiated cells. For example, embryonic cells may not require cis-acting sequences to initiate DNA replication. All DNAs (including SV40 and PyV) injected into *Xenopus* eggs replicate under apparent control of the cell division cycle, but with no apparent requirement for specific DNA sequences (including T-Ag; ref. 130). In *Drosophila*, replication bubbles are at least 5X more frequent in embryos than in differentiated cells, suggesting that sequences can function as an ori in embryos but not recognized as ori in differentiated cells (131). Other results suggest that embryonic cells may not respond to control signals recognized by differentiated cells. PyV replicates in most mouse differentiated cell types, but PyV neither replicates its DNA nor transcribes its genes in mouse embryonal carcinoma (EC) cell lines which are thought to represent pluripotent cells from 5-6 day old mouse embryos (5,6,12). Thus, some experiments indicate that DNA replication in early embryonic cells has no sequence requirements while other experiments indicate that they have stringent requirements that differ from those in more developed cells.

In an effort to resolve this paradox, Wirak et al. (14) injected viral and plasmid DNA into the nucleus of mouse oocytes, 1-cell and 2-cell embryos, and then allowed development to continue *in vitro*. They demonstrated that wild-type PyV DNA replicates in mouse embryos, but not in mouse oocytes, consistent with the biological properties of these cells. Furthermore, injection of plasmid DNA carrying PyV sequences revealed that DNA replication required a functional PyV ori and PyV T-Ag. Although SV40 DNA replication has not been detected in mouse fibroblasts, it replicates about 3% as well as PyV DNA when injected into mouse embryos and requires a functional SV40 ori and SV40 T-Ag (123). Therefore, mouse embryos, in contrast to amphibian embryos, require unique cis-acting DNA sequences for replication; all DNAs tested so far that lacked a functional viral origin failed to replicate in mouse embryos, with or without T-Ag present (D. Cupo, unpublished data). Thus, it may be possible to identify cellular sequences that will allow DNA replication in preimplantation mouse embryos. Ori sequences in mammalian embryos may also be cell-type specific (14). Two configurations of the PyV ori, α–β-core and β-core, function either in mouse differentiated cells or in mouse embryos, but another

configuration, α-core, functions only in mouse differentiated cells. The α–β sequence alone is inactive. Therefore, one configuration (α-core) of the PyV ori is active only in differentiated mouse cells while two other configurations (α–β-core and β-core) are active in both embryonic and differentiated cells.

Since the α and β-elements encompass gene enhancer sequences, the ability of these sequences to enhance transcription under the same conditions was analyzed using a plasmid vector containing the *E. coli* chloramphenicol acetyltransferase gene (CAT) coupled to the HSV thymidine kinase promoter (137). PyV enhancer sequences 1, 2 and 3 (Fig. 3) were inserted individually and in combination at a site 600 bp upstream of the promoter. The same results were obtained regardless of which orientation was examined. Transfection of differentiated mouse cells (3T3) revealed that maximum enhancer activity required elements 1, 2 and 3 (Table 3). No significant difference in DNA replication was detected when 3-core, 1+2-core, and 1+2+3-core plasmids were used to transfect MOP cells (3T3 cells expressing PyV T-Ag from an integrated

Table 3. PyV Sequences that Control Gene Expression and DNA Replication Respond Differently in Mouse Embyros and Differentiated Cells

Sequences*	Differentiated Cells		2-Cell Embryos	
	Enhancer Activity#	DNA Rep&	Enhancer Activity#	DNA Rep&
pTKcat or pML-1	1	-	1	-
1	1		1	
2	1		2	
3	3	+	15	+
1+2	13	+	1	-
1+3	40		1	
2+3	60		8	
1+2+3	80	+	1	+

*Sequences are in positive orientation relative to promoter (negative orientation gave similar results).
#Ratio of CAT activity from pTK+PyV sequence to CAT activity from pTK alone.
&Replication of pML-1 DNA containing PyV ori-core adjacent to the indicated PyV sequence.

ori⁻ PyV genome), even though a 27-fold variation was observed in their ability to enhance CAT expression. When the same plasmids were injected into one of the nuclei in 2-cell embryos, element-3 was the only component with significant enhancer activity; element-1+2 was inactive. Since element-3 was required for DNA replication in embryos and element-1+2 was inactive,

these data supported the hypothesis that cell-specific activation of PyV ori-core depended on cell-specific recognition of a gene enhancer.

One result is not consistent with the hypothesis that enhancer activity activated ori-core. Element-1+2+3 fails to stimulate CAT expression in embryos, although the same plasmid stimulates CAT expression 80-fold in 3T3 cells (Table 3). Since 1+3 is also inactive as an enhancer while 2+3 is at least 50% as active as 3 alone, element-1+2 (primarily the element-1 component) appears to act as the target for a repressor protein present in embryos, but not in differentiated mouse cells. Since either orientation of this sequence inactivates transcription at a distance (at least 600 bp) from the promoter, it acts as an *embryo-specific negative enhancer element* (i.e. a "silencer"). However, the repressor does not interfere with the ability of element-1+2+3 to activate ori-core. Therefore, either the silencer does not interfere with enhancer activity (e.g. the silencer acts directly on promoters), or element-3 enhancer activity is different than its ori-core activation activity. Presumably, this embryonic repressor either does not act on the PyV promoter used to produce T-Ag, or, more likely, that low levels of T-Ag are sufficient for PyV replication.

Two other pieces of data also indicate that enhancer elements are not acting on ori-core in the same way they act on promoters. First, the minimal α-element (Fig. 3) is 3X more effective in activating ori-core than activating the PyV promoter (67). Second, activation of ori-core by the β-element exhibits an optimal distance effect of not closer than 20 bp and not further away than 200 bp (W. Muller and J. Hassell, unpublished data). Thus, it is possible that enhancer elements can activate origins of replication by some variation of their interaction with promoters.

The difference in requirements for DNA replication in mammalian and amphibian embryos probably reflects the difference in the amount of maternally inherited proteins and mRNA. Fertilization of *Xenopus* eggs results in a large increase in the rate of protein synthesis, with no requirement RNA synthesis due to the availability of previously nontranslatable oocyte mRNAs. This allows *Xenopus* to undergo its first 12 cleavage events in the absence of zygotic gene expression (132), whereas fertilized mouse eggs can undergo only one or two cleavage events (133). Therefore, amphibian eggs may contain such a high concentration replication factors that the difference in Km between ori and non-ori DNA substrates is not observed with injected plasmid DNA. Alternatively, the rapid initial rate of cell division in amphibian (35 min/cleavage) and *Drosophila* (10 min/cleavage) embryos and may necessitate a different approach to regulation of DNA replication than taken by mammalian embryos (20-24 hrs/cleavage).

The fact that mouse embryos are permissive for PyV DNA replication and marginally permissive for SV40 DNA replication while EC cell lines are nonpermissive for both reveals that the relationship between EC cell lines and mouse embryos needs further clarification. Preliminary experiments suggest that replication and transcription of DNA from PyV EC host range mutants in mouse embryos are significantly better than with wild-type PyV (D. Wirak, unpublished data). Therefore, PyV permissive cell factors may be of two kinds, embryo-specific

and differentiated cell-specific. Since the maximum level of enhancement by PyV elements 1, 2 and 3 was 5X greater in differentiated mouse cells than in mouse embryos (Table 3), mouse embryos appear to express differentiated cell factors in low abundance and embryo-specific factors in high abundance. EC and other embryonic cell lines may originate from pluripotent cells derived at later stages in development (6), and may be missing differentiated cell factors entirely. Direct comparisons of the requirements for DNA replication and gene expression in real mouse embryos with various EC and embryo-derived cell lines should clarify the lineages of stable cell lines and their usefulness in investigating early embryonic development.

Initiation of Replication - "Mechanism"

DNA Binding Site for an Initiation Factor(s) in SV40 Includes the Promoter

The DNA binding site for factors required to initiate SV40 DNA replication has been identified by its ability to interfere with replication of pML-1 plasmids containing the SV40 ori region (pSV40) when the competitor DNA, in the form of a double-stranded circular plasmid, was added to the in vitro system (138). The competitor DNA presumably inhibits replication by adsorbing one or more required factors. Thus, the DNA binding site for the factor(s) present in lowest concentration were identified, since the availability of this factor would be the rate limiting step in replication. The data summarized in figure 7 identifies nucleotide 108 on the late gene side of ori to nucleotide 5236 on the early gene side as the binding site for one or more factors required to replicate SV40 DNA. The region most stringently required was nucleotide 72 to 0/5243; some of the flanking sequences (open boxes) facilitated binding.

This DNA sequence is assumed to bind an initiation factor because it is unique to the SV40 ori sequence. It was not present in any plasmid DNA, the remainder of SV40 DNA, or in PyV DNA. Interference did not result from simply binding T-Ag because plasmids with binding sites 1 and 2 alone (XS15, XS16), and plasmids that bound SV40 T-Ag nonfunctionally (pXho1, pPyV) had little ability to interfere with DNA replication, and addition of T-Ag did not overcome the competition. Thus, T-Ag appears to be in excess. Since plasmids containing the PyV ori instead of the SV40 ori as well as plasmids that lacked either viral ori neither replicated nor interfered with SV40 ori replication, this initiation factor binding site is specific for the SV40 ori. Therefore, this initiation factor binding site is the target for either a unique subfraction of T-Ag, or a permissive cell factor required to initiate SV40 (but not PyV) DNA replication. Such a factor may form a specific complex with SV40 T-Ag. The initiation factor binding site includes only part of ori-core and all of the aux-2 sequence, suggesting that the function of aux-2 is to facilitate binding of the initiation complex to ori-core. Interestingly, aux-2 facilitates replication *in vivo* but not in some *in vitro* systems (64,65)

RNA-Primed DNA Synthesis at Ori.

Initiation sites for RNA-primed DNA synthesis been identified on one strand of the SV40 (45,46) and PyV (15) ori-sequences. These sites are indistinguishable from those found outside ori in terms of their frequency, sequence composition, confinement to retrograde sides of forks, and average lengths of RNA primers, suggesting that the first DNA chain at ori is, in essence, synthesis of the first Okazaki fragment (46). Therefore, the initiation sites in ori appear to be utilized by DNA primase in conjunction with DNA polymerase–α, although the failure of BP-dGTP to inhibit initiation of SV40 chromatin replication (25) and the possibility that some antibodies raised against α–polymerase cross-react with δ–polymerase means that DNA polymerase-δ may be the replicative enzyme (see "DNA Primase-DNA Polymerases–α and –δ").

Murakami et al. (121) removed DNA primase-DNA polymerase-α from HeLa cell lysates by treating the extracts with a monoclonal antibody directed against human DNA polymerase-α, and then discovered that SV40 ori-dependent DNA replication was not observed unless HeLa or CV-1 DNA primase-DNA polymerase-α was added back to the reaction. Addition of other DNA polymerases, including the same enzyme from nonpermissive cells, failed to restore activity. These results demonstrate that DNA primase-DNA polymerase-α is required to replicate the DNA plasmids added to this system, and the specificity for a permissive cell enzyme strongly suggests that DNA primase-DNA polymerase-α is required to initiate DNA synthesis at ori as well as to synthesize DNA at replication forks.

Aphidicolin, a specific inhibitor of DNA polymerases α and δ, provided a novel method for distinguishing between initiation of replication at ori and continuation of DNA replication beyond ori (49). In the presence of sufficient aphidicolin to inhibit total DNA synthesis by 50%, initiation of DNA replication in SV40 chromosomes (or plasmid DNA containing SV40 ori) continued *in vitro* while DNA synthesis in the bulk of SV40 replicative intermediate (RI) DNA that had initiated replication *in vivo* was rapidly inhibited. This resulted in accumulation of early RI in which most nascent DNA was localized within a 600 to 700 bp region centered at ori. Accumulation of early RI was observed only under conditions that permitted initiation of SV40 ori-dependent, T-ag-dependent DNA replication, and only when aphidicolin was added to the *in vitro* system. Increasing aphidicolin concentration revealed that DNA synthesis in the ori region was not completely resistant to aphidicolin but simply less sensitive than DNA synthesis at forks further removed. Since DNA synthesized in the presence of aphidicolin was concentrated in the 300 bp on the early gene side of ori, the initial direction of DNA synthesis is the same as that of early mRNA synthesis (Fig.8). These data are consistent with the observation that aphidicolin does not prevent reinitiation of SV40 DNA replication when tsA mutant-infected CV-1 cells are shifted from the restrictive to the permissive temperature, resulting in RI with newly synthesized DNA localized in the ori region (141). Therefore, *in vitro* or *in vivo*, DNA synthesis in the SV40 ori region is less sensitive to aphidicolin than DNA synthesis throughout the remainder of the

genome.

The effects of aphidicolin on SV40 DNA replication are similar to its effects on DNA primase-DNA polymerase-α, suggesting that this enzyme initiates DNA synthesis at ori. Insensitivity to aphidicolin appears to be a general property of eukaryotic DNA primases

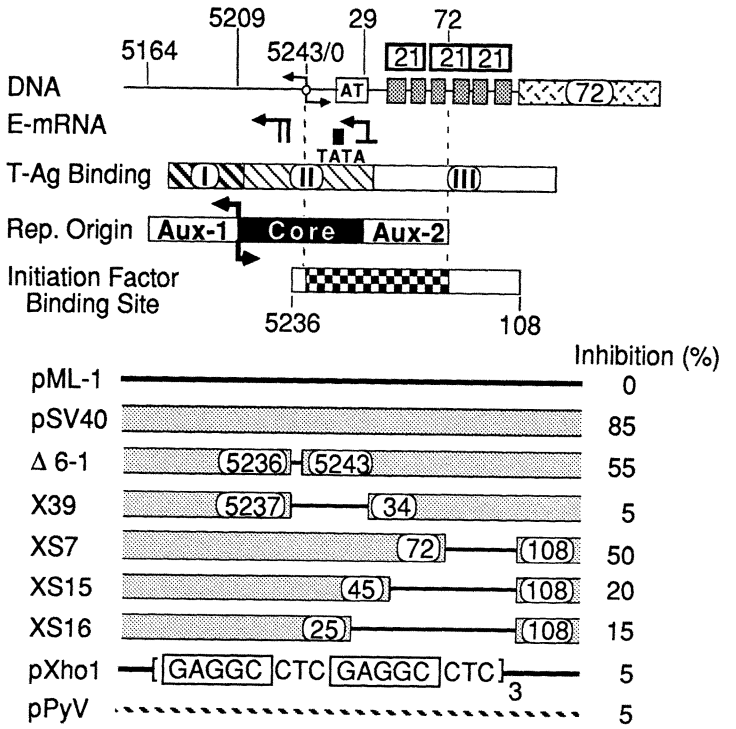


Figure 7. Identification of a DNA binding site for a factor(s) specifically required for initiation of SV40 DNA replication (138). pML-1 (pBR322 minus the poison sequence) is the parent plasmid. pSV40 represents several different plasmids containing the complete SV40 ori. Deletions in pSV40 plasmids are indicated with the nucleotide locations of their end points. All deletions of ori-core completely inactivate the plasmids ability to replicate. pXho1 contains tandem Xho1 linkers that generate a strong binding site for T-Ag. pPyV represents several different plasmids containing the PyV ori. Locations are indicated for one 72 bp repeat (enhancer), six $G_{3-4}CG_2Pu_2$ repeats overlapping three 21 bp repeats (promoter), TATA box, and early mRNA initiation sites. The origin of bidirectional replication is indicated by arrows on each side of ori pointing in the direction of synthesis.

(142-144). Aphidicolin reduces the frequency of initiation by purified CV-1 DNA primase-DNA polymerase-α on a single-stranded DNA template by 30%, but once initiation occurs, incorporation of the first 30 residues is completely resistant to aphidicolin (49). At this point, the probability of extending nascent DNA chains decreases rapidly. Since the 20-25 residues incorporated by CV-1 DNA primase-DNA polymerase-α per initiation event (17) corresponds closely to the number of nucleotides whose synthesis is completely resistant to aphidicolin, DNA primase alone may synthesize a $pppN(N)_{5-7}$ $(dN)_{20-30}$ moeity before DNA polymerase-α takes over.

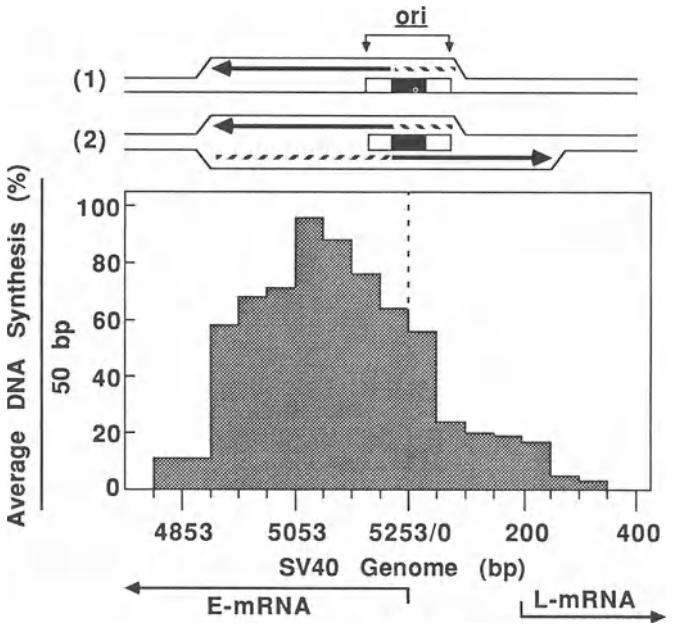


Figure 8. Direction of DNA synthesis in the SV40 ori region. DNA replication in SV40 wt800 chromosomes was initiated *in vitro* in the presence of radiolabeled [α-^{32}P]dNTPs and aphidicolin to allow accumulation of newly replicated SV40 RI DNA (49). These molecules were purified and then digested with various restriction endonucleases to determine the genomic distribution of nascent DNA. The data were averaged together over 50 bp segments.

Model for Initiation of SV40 DNA Replication

All of the available data are consistent with the model proposed by Hay and DePamphilis (45) for initiation of SV40 DNA replication (Fig. 9A). A preinitiation complex binds to the initiation factor binding site, opens the ori region to create a replication bubble, and thus allows DNA primase-DNA polymerase-α to select, by some stochastic process, one of several possible template initiation sites within this first initiation zone. Since the initiation sites identified within ori are all located on the strand encoding early mRNA (Fig. 9B), the first DNA chain synthesized should extend from ori towards the early genes. In fact, we found that the first fully synthesized DNA sequences included ori and the region about 300 bases on its early gene side (Fig. 8). Since bidirectional DNA replication originates at the junction of T-Ag binding site 1 and ori-core (Figs. 4,5), initiation of the first DNA chain must occur somewhere upstream of this point. It is interesting to note that each of the GC-rich repeats promotes initiation of an RNA-primed DNA chain (Fig. 9B). It is this nascent RNA-primed DNA chain that becomes the continuously synthesized forward arm of replication forks advancing in the same direction as early mRNA synthesis. DNA synthesis proceeds an average of 250 ± 150 bp (Fig. 8, diagram 1) before a second initiation zone is created on the retrograde arm of the fork to allow DNA synthesis to begin in the direction of late mRNA synthesis (Fig. 8, diagram 2). This is consistent with the hypothesis that the average center-to-center distance between nucleosomal cores in front of replication forks (220 ± 72 bp) determines the average size of the initiation zone which, in turn, determines the sizes of Okazaki fragments (1,8,12,41).

Because the initiation factor binding site includes the early gene promoter, TATA box and start sites for early mRNA synthesis (Fig. 3), it is tempting to suggest that initiation may be activated by transcription through ori. However, DNA replication in chromosomes or plasmid DNA was completely resistant to α-amanitin, a specific inhibitor of RNA polymerases II and III. Since SV40 genes are normally transcribed by RNA polymerase II, activation of ori by RNA synthesis is unlikely. Never the less, binding of transcriptional initiation factors that specifically recognize the GC-rich repeats (e.g. Sp1), may activate ori by facilitating binding of DNA replication initiation factors to ori-core. For example, the initiation factor binding site overlaps T-Ag binding site 2 (Fig. 9B), suggesting that the permissive cell factor that binds to the initiation factor site may facilitate binding of T-Ag to site 2 which in turn uses its helicase activity to separate the two ori strands, but only if negative superhelical strain is present to promote DNA melting. If DNA primase-DNA polymerase-α is associated with the T-ag-initiation complex, this could position the enzyme to initiate RNA primed-DNA synthesis somewhere upstream on the E-mRNA template.

Model for Initiation of PyV DNA Replication

The characteristics of RNA-primed DNA synthesis initiation sites (15), the location of the

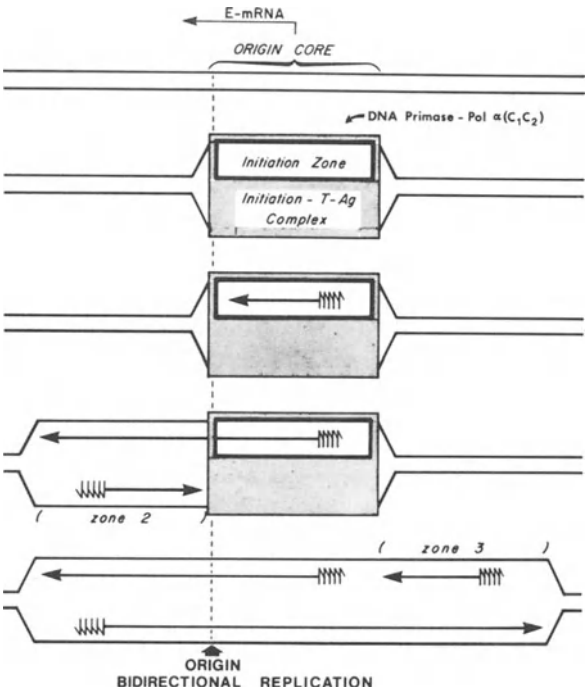


Figure 9A. Model for initiation of SV40 DNA replication - Sequence of events.

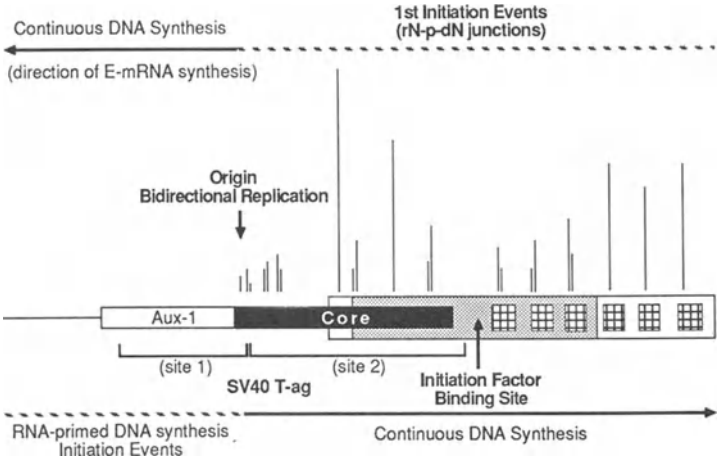


Figure 9B. Model for initiation of SV40 DNA replication - Close-up of ori region.

OBR with respect to ori-core (15), and the initial direction of DNA replication in the early gene direction (146) are all essentially the same as in SV40. Therefore, assuming that the mechanism for initiation of PyV DNA replication is analogous to that for SV40, enhancer elements play the same role in PyV that the early gene promoter (and to a lesser extent the SV40 enhancer) does in SV40. Specifically, these transcriptional elements may facilitate binding of T-Ag to ori-core, either directly through interaction with the T-ag-initiation complex, or indirectly by making the ori sequence more accessible to in initiation factors. Note that the three major DNA binding sites for PyV T-Ag (A, B and C) lie to the early gene side of ori-core (Fig. 3), revealing that binding of PyV T-Ag to PyV ori-core is significantly weaker than binding of SV40 T-Ag to its ori-core. This could account for the strong requirement for transcriptional elements as part of the PyV ori compared with their relatively weak requirement as part of the SV40 ori. Furthermore, since both early and late gene promoters and mRNA start sites lie outside PyV ori, each directing mRNA synthesis away from ori-core, it is unlikely that PyV ori-core is activated by RNA synthesis. Thus, it is possible that the protein binding specifically to the β–element (127,128; Fig. 3) facilitates binding of the T-Ag-initiation complex to ori-core. In differentiated mouse cells, ori-core is activated by proteins that bind to either or both α- and β-elements, whereas, in embryonic mouse cells, only the β-element is recognized. However, the protein recognizing β in embryonic cells may not be identical to the one used in differentiated cells. β-Enhancer activity in embryos was 5-fold less than α–β-enhancer activity in 3T3 (Table 3), and PyV host-range mutants that replicate on EC cell lines have altered β-elements (5,6). The role enhancer elements play in activating ori-core can be distinguished from their role in activating promoters by the fact that the putative embryo-specific silencer sequence inactivates transcription by not DNA replication.

Termination of DNA Replication

Replicating Intermediates

Of the 11 forms of viral DNA identified in a lytic infection (8), only 5 appear as major intermediates or products of replication during the peak period of viral DNA synthesis (Fig. 10A). Other forms of viral DNA such as circles with a tail ("rolling-circle intermediates"), linear monomers, and large circular and linear concatemers containing 2 to 10 tandemly arranged, head-to-tail molecules accumulate only under suboptimal conditions for replication or late in the infection cycle (8,12). The bulk of nascent DNA is synthesized semiconservatively and rapidly appears in circular replicating intermediates (RI) consisting of covalently-closed parental DNA strands and two replication forks traveling in opposite directions (8,12). Although replication forks in the total RI population terminate replication at a position about 180^{o} from ori (155,156,160-162), the two forks in a single RI molecule do not migrate synchronously

(145,146). However, only 1/3 of the RI molecules contain two forks the same distance from ori; the average pair of replication forks are out of synchrony by 12% of the distance traveled. This degree of asymmetry is similar at all stages of replication, suggesting that forks move at different rates instead of, or in addition to, leaving ori at different times. Therefore, forks close to ori appear to initiate concurrently and move out at similar rates (147), but as a result of nonuniform rates of travel, when the average replication fork traverses 50% of the genome, its sibling has traversed only 44% of the genome, leaving the two forks separated by an average of 315 bp (145). The extent of asymmetry in PyV fork movement appears to be twice as great as in SV40 (about 400 bp in PyV(RI) compared with 190 bp in SV40(RI) (146), but the degree of fork asymmetry was still the same in early RI as in late RI (148). Therefore, the two replication forks in a single RI frequently do not arrive in the termination region at the same time.

Elongation of nascent DNA chains in SV40(RI) proceeds bidirectionally until replication is 85-95% completed. At this point, late replicating DNA intermediates (RI*, Fig. 10A) accumulate to a level 2-3 times greater than observed for an equivalent sample of RI at earlier stages in replication, indicating that separation of sibling molecules is a slow step in replication (48,82,148-150,152). Earlier reports that SV40(RI*) accumulate up to 8-fold over RI appear to have overestimated the extent of accumulation because of technical complications (8,12). In addition, if parental DNA is topologically relaxed (e.g. a single "nick") or one of the replication forks broken, the distribution of RI during various fractionation procedures can be dramatically altered. In fact, a previous report that some RI accumulated at about 80% completion (149) most likely resulted from RI in which ssDNA at one replication fork was broken (48,154). Finally, the fraction of RI* depends on the physiological state of the infected cells, the multiplicity of infection, and the DNA sequence in the termination region (137,156). Electrophoretic fractionation of PyV(RI) reveals a distribution of molecules remarkably similar to that seen with SV40(RI) (148,157-159), although the authors did not interpret the data as an accumulation of PyV(RI*).

Replication forks in SV40(RI*) have been arrested at specific sites in the termination region (145). The two major arrest sites are separated by about 470 bp of unreplicated DNA centered at nucleotide 2743. This is about 52% of the genome from the Bgl 1 site in ori-core and encompasses the normal termination region for SV40 DNA replication (155,156,160-162). Therefore, most replication forks pause when replication is about 91% complete, consistent with the accumulation of RI*. Replication forks have also been shown to accumulate at specific sites flanking the termination region of PyV DNA (163). These and other studies (164,165) also reveal that replication forks pause at many sites throughout the SV40 and PyV genomes, although accumulation of forks is most evident in the termination region. The locations of about 80% of the most prominent in vivo SV40 replication arrest sites correspond to sites that also arrest DNA polymerase-α on the same DNA template in vitro (165). However, this correlation is more likely a consequence of the termination process rather than a cause of it (see below).

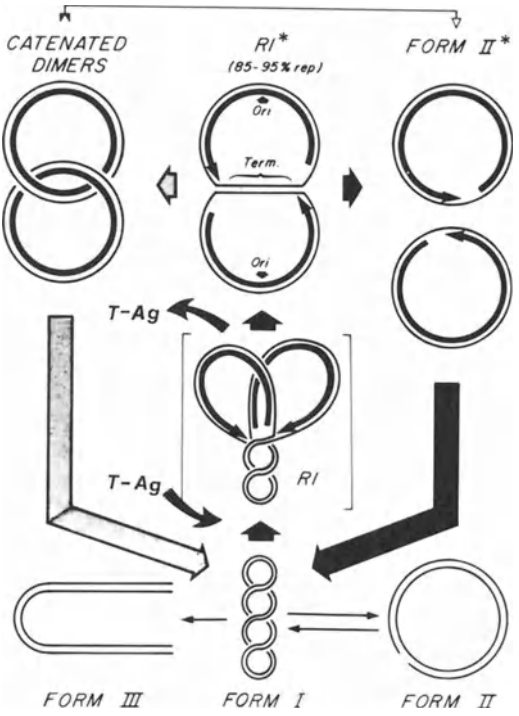


Figure 10A. Major DNA intermediates and products of SV40 and PyV replication.

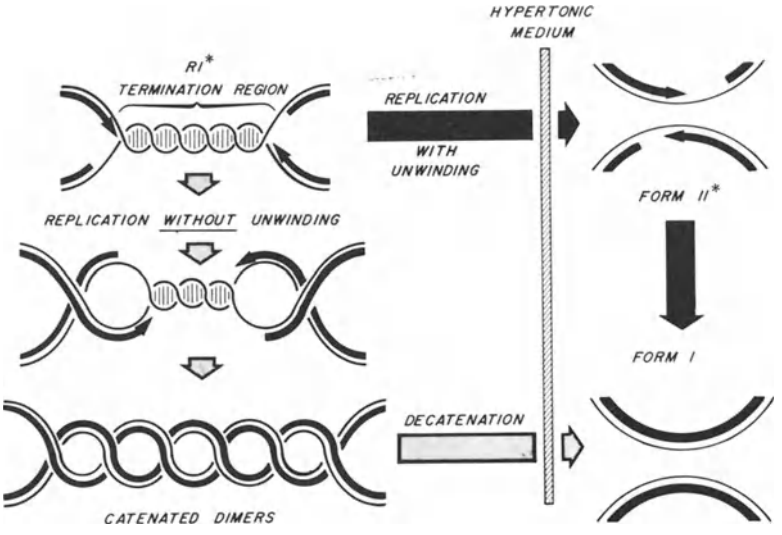


Figure 10B. Two alternative pathways for termination of DNA replication.

Termination Region

Separation of sibling molecules does not require a unique termination site on the genome because the normal termination site can be moved to other locations relative to ori or removed altogether without preventing replication; termination continues to occur approximately 180° from ori (reviewed in 156). Therefore, separation of sibling molecules appears to occur at whatever sequence the two oncoming replication forks happen to meet. Because all replication forks do not arrive simultaneously at the point 180° from ori, the actual sites where termination occurs represent a distribution about the mean that reflects variation from one RI molecule to the next. This would explain why replication fork arrest sites are distributed about the mean termination site by ±450 bp (145), and the gap in the nascent DNA strand of SV40(II*) DNA was distributed over a 730 bp region (166). These data define a "termination region" in which separation of sibling DNA molecules occurs that is sequence independent. However, although a unique sequence is not required for separation of sibling molecules, the sequence in the termination region does strongly affect the pathway used for separation of sibling molecules.

Separation of Sibling Molecules

In addition to RI and RI*, Form II* and catenated dimers have also been demonstrated to act as transient intermediates during viral DNA replication in virus-infected cells and subcellular DNA replication systems. SV40(II*) DNA consists of circular monomers with a short gap (≈50 nucleotides) in their nascent DNA strand in the termination region (8,12). Its structure, rapid synthesis, and disappearance of concomitant with the appearance of Form I DNA are properties consistent with a transient intermediate in viral DNA replication. Furthermore, SV40(II*) accumulates when formation of SV40(I) DNA is prevented by omitting a cellular protein required for completion of Okazaki fragments (48,162). Thus, separation can occur prior to the completion of the nascent DNA strand the termination region. Catenated dimers can also be rapidly labeled during the period of maximum DNA synthesis and disappear at a rate consistent with a transient intermediate in replication (152). Furthermore, most catenated dimers formed during a lytic SV40 infection result from DNA replication rather than recombination (8,12).

Sundin and Varshavsky (167) observed that when SV40-infected monkey cells are subjected to a hypertonic shock, all of the newly replicated DNA accumulated as catenated dimers, suggesting that formation of catenated dimers was an obligatory intermediate in the termination process. However, further examination of this phenomenon revealed two additional observations that significantly modify this interpretation (137,156). Hypertonic shock results in the reversible accumulation of *both RI* and catenated dimers*, suggesting that this treatment inhibits DNA unwinding specifically in the termination region and thus promotes the formation of catenated dimers as well as prevents their resolution. This inhibition appeared to be specific for the termination region of SV40 rather than a general inhibition of all replication forks because RI

continues replication while RI* accumulates. This point was proven by demonstrating that the hypertonic shock phenomenon depended on the DNA sequence at the termination region (156,168). The termination site was changed by cloning SV40 sequences into pML-1 or by deleting or adding sequences to SV40 DNA. These molecules, which did not produce T-Ag, replicated almost as efficiently as wild-type DNA when transfected into COS-1 or CMT-3 cells. However, while SV40 DNA still formed catenated dimers after hypertonic shock, molecules that terminated at other DNA sequences did not. Neither DNA size nor DNA sequences outside the termination region made any difference in the experiment. Therefore, some DNA sequences promote formation of catenated dimers in the termination region, and other sequences do not. The alternative possibility that some sequences promote resolution of catenated dimers would require that catenated intertwines are localized at the termination region, but this does not appear to be the case (55). These data demonstrate that catenated dimers are *not* an obligatory intermediate in termination, because conditions that prevent resolution of catenated dimers formed at the SV40 termination region do not trap catenated dimers when termination occurs at other DNA sequences.

These data are most easily explained on the basis of two alternative pathways for termination of replication. In isotonic medium where the rate of SV40 DNA replication is optimal, DNA replication and DNA unwinding continue concurrently until the two sibling molecules separate and become Form II* molecules (Fig. 10B). Alternatively, the final steps in DNA replication can continue in the absence of DNA unwinding, resulting in the production of catenated dimers with one intertwine of the two duplex molecules for each turn of the parental DNA helix that was not removed. Since SV40 catenated dimers have been shown to contain as many as 20 to 25 intertwines (167), this pathway can begin when 200 to 250 bp of unreplicated DNA remain at the termination region.

Topoisomerase II

Hypertonic medium has two effects on termination of replication. First, it inhibits unwinding of DNA (resulting in accumulation of RI*) and thus promotes formation of catenated intertwines. Second, it inhibits decatenation, resulting in accumulation of newly formed catenated dimers. Therefore, a likely target of hypertonic shock is topoisomerase II, the only enzyme known to be able to both relieve topological strain in superhelical DNA, and to unknot intertwined duplex DNA (168). A requirement for topoisomerase II in the termination process is strongly demonstrated by the fact that yeast topoisomerase II mutants (but not topoisomerase I mutants) accumulate catenated plasmid DNA (169) and interlocked nuclei (170) at the restrictive temperature. Furthermore, removal of topoisomerase II from cellular extracts capable of supporting SV40 DNA replication causes newly replicated DNA to accumulate as catenated dimers (114). Topoisomerase II is associated with PyV replicating chromosomes (172). **If topoisomerase II acts behind replication forks** to remove torsional strain by passing the

two daughter molecules through one another (the enzyme's decatenation activity), then unwinding of the parental unreplicated duplex can continue to completion without forming any catenated intertwines (171), whereas **if topoisomerase II acts only in front of replication forks**, then formation of catenated intertwines is an obligatory step in termination.

Topoisomerase II interacts efficiently with most, but not all, DNA sequences (153). Therefore, the SV40 termination region appears to represent a poor substrate for this enzyme so that inhibition by hypertonic medium accentuates the problem and thus promotes formation of catenated dimers. When a yeast centromere (cen-3) was placed 180° from SV40 ori, it promoted formation of catenated dimers even more effectively than SV40 when subjected to hypertonic shock, but cen-3 did not produce catenated dimers when placed 80° from ori (S. Fields-Berry and M. DePamphilis, manuscript in preparation). This sequence contains 86 bp that are 95% A/T. The SV40 termination region contains 213 bp that are 71% A/T , interrupted by a 25 bp G/C-rich sequence. Since cen-3 catenated dimers contained 7-8 intertwines, and SV40 catenated dimers contained 20-24 intertwines, it appears that A/T-rich sequences are difficult to unwinding, possibly because it is a poor substrate for topoisomerase II.

Termination of DNA Synthesis

In summary, there is no compelling evidence that formation of catenated dimers is an obligatory intermediate in termination of replication. In fact, under normal physiological conditions and at most termination regions it would appear that the major product of separation is Form II*. The gap is then filled in and the molecule sealed to produce Form I DNA. This final gap filling step, like the completion of Okazaki fragments (41), requires gap-filling proteins recovered in the cytosol fraction (48,162). The superhelical turns in Form I DNA result primarily from organization of DNA into nucleosomes, and since nascent DNA on both arms of replication forks is assembled rapidly into nucleosomes (8,12), superhelical turns can be expressed as soon as Form II* DNA is sealed.

References

1. DePamphilis, M.L. and Wassarman, P.M. (1980) Ann. Rev. Biochem. 49:627.
2. Ozer, H.L., Slater, M.L., Dermody, J.J. and Mandel, M. (1981) J. Virology 39:481.
3. Lebkowski, J.S., Clancy, S. and Calos, M.P. (1985) Nature 317:169.
4. Dubensky, T.W. and Villarreal, L.P. (1984) J. Virology 50:541.
5. Amati, P. (1985) Cell 43:561.
6. Levine, A.J. (1982) Curr. Top. Microbiol. Immunol. 101:1.
7. Acheson, N.H. (1980) In: Molecular Biology of Tumor Viruses, part 2, (J. Tooze, ed.), Cold Spring Harbor Laboratory, p. 125.
8. DePamphilis, M.L. and Wassarman, P.M. (1982) In: Organization and Replication of Viral

DNA (A.S. Kaplan, ed), Boca Raton, FL: CRC Press, p. 37.

9. Das, G.C. and Niyogi, D.K. (1981) In: Progress in Nucleic Acid Research and Molecular Biology (W.E. Cohn, ed), Academic Press, N.Y., p. 187.
10. Seidman, M. and Salzman, N.P. (1983) In: Replication of Viral and Cellular Genomes (Y. Becker, ed.), Martinus Nijhoff, Boston, p.29.
11. Richter, A. and Otto, B. (1983) In: Replication of Viral and Cellular Genomes (Y. Becker, ed.), Martinus Nijhoff, Boston, p. 53.
12. DePamphilis, M.L. and Bradley, M.K. (1986) In: The Papovaviridae, Vol. I (N.P. Salzmann, ed.), Plenum Press, N.Y., p. 99.
13. Kornberg, A. (1980) DNA Replication, W.H. Freeman.
14. Wirak, D.O., Chalifour, L.E., Wassarman, P.M., Muller, W.J., Hassell, J.A., and DePamphilis, M.L. (1985) Mol. Cell. Biol. 5:2924.
15. Hendrickson, E.A., Fritze, C., Folk., W. and DePamphilis, M.L. (1987) EMBO J., in press.
16. Arai, K. and Kornberg, A. (1981) Proc. Natl. Acad. Sci. USA 78:69.
17. Yamaguchi, M., Hendrickson, E.A., and DePamphilis, M.L. (1985) J. Biol. Chem. 260:6254.
18. Krokan, H., Schaffer, P. and DePamphilis, M.L. (1979) Biochemistry 18:4431.
19. Waqar, M.A., Evans, M.J., Burke, J.F., Tsubota, Y., Plummer, M.J., and Huberman, J.A. (1983) J. Virology 48:304.
20. Gourlie, B.B., Pigiet, V., Breaux, C.B., Krauss, M.R., King, C.R., and Benbow, R.M. (1981) J. Virology 38:862.
21. Lee, M.Y.W.T., Tan, C.K., Downey, K.M. and So, A.G. (1984) Biochemistry 23:1906.
22. Crute, J.J., Wahl, A.F. and Bambara, R.A. (1986) Biochemistry 25:26.
23. Lee, M.Y.W.T., Toomey, N.L. and Wright, G.E. (1985) Nucl. Acids Res. 13:8623.
24. Wobbe, C.R., Dean, F.B., Murakami, Y., Weissbach, L., and Hurwitz, J. (1986) Proc. Natl. Acad. Sci. USA 83:4612.
25. Decker, R.S., Yamaguchi, M., Possenti, R., Bradley, M.K., and DePamphilis, M.L. (1987), J. Biol. Chem., in press.
26. Khan, N.N., Wright, G.E., Dudycz, L.W. and Brown, N.C. (1984) Nucl. Acids Res. 12:3695.
27. Pritchard, C.G., Weaver, D.T., Baril, E.F. and DePamphilis, M.L. (1983) J. Biol. Chem. 258:9810.
28. Yamaguchi, M., Hendrickson, E.A., and DePamphilis, M.L. (1985) Mol. Cell. Biol. 5:1170.
29. Vishwanatha, J.K., Coughlin, S.A., Wesolowski-Owen, M. , Baril, E.F. (1986) J. Biol. Chem. 261:6619.
30. Skarnes, W., Bonin, P. and Baril, E. (1986) J. Biol. Chem. 261:6629.

31. Ottiger, H-P, and Hubscher, U. (1984) Proc. Natl. Acad. Sci. USA 81:3993.
32. Weinmann-Dorsch, C., Hedl, A., Grummt, I., Albert, W., Ferdinand, F.J., Friis,R.R., Pierron, G., Moll, W., and Grummt, F. (1984) Eur. J. Biochem. 138:179.
33. Krokan, H. (1980) In: "Chromosome Damage and DNA Repair" (K. Kleppe and E. Seeberg, eds.), Plenum Press.
34. Ogawa, T. and Okazaki, T. (1984) Mol. Gen. Genetics 193:231.
35. Pigiet, V., Eliasson, R. and Reichard, P. (1974) J. Mol. Biol. 84:197.
36. Hunter, T. and Francke, B. (1974) J. Mol. Biol. 83:123.
37. Eliasson, R. and Reichard, P. (1978) J. Biol. Chem. 253:7469.
38. Eliasson, R. and Reichard, P. (1979) J. Mol. Biol. 129:393.
39. Reichard, P. and Eliasson, R. (1979) Cold Spring Harbor Symp. Quant. Biol. 43:271.
40. Anderson, S., Kaufmann, G., and DePamphilis, M.L. (1977) Biochemistry 16:4990.
41. Anderson, S. and DePamphilis, M.L. (1979) J. Biol. Chem. 254:11495.
42. Kaufmann, G., Anderson, S. and DePamphilis, M.L. (1977) J. Mol. Biol. 116:549.
43. Kaufmann, G. (1981) J. Mol. Biol. 147:25.
44. DePamphilis, M.L., Anderson, S., Bar-Shavit, R., Collins, E., Edenberg, H., Herman, T., Karas, B., Kaufmann, G., Krokan, G.H., Shelton, E.R., Su, R.T., Tapper, D.P. and Wassarman, P.M. (1979) Cold Spring Harbor Symp. Quant. Biol. 43:679.
45. Hay, R.T. and DePamphilis, M.L. (1982) Cell 28:767.
46. Hay, R.T., Hendrickson, E.A., and DePamphilis, M.L. (1984) J. Mol. Biol. 175:131.
47. Li, J.J. and Kelly, T.J. (1984) Proc. Natl. Acad. Sci. USA 81:6973.
48. Tapper, D.P., Anderson, S. and DePamphilis, M.L. (1982) J. Virology 41:877.
49. Decker, R.S., Yamaguchi, M., Possenti, R., and DePamphilis, M.L. (1986) Mol. Cell. Biol. 6:3815.
50. Vishwanatha, J.K., Yamaguchi, M., DePamphilis, M.L., and Baril, E.F. (1986) Nucl. Acids Res. 14:7305.
51. Tseng, B.Y. and Ahlem, C.N. (1984) Proc. Natl. Acad. Sci. USA 81:2342.
52. Cusick, M.E., DePamphilis, M.L. and Wassarman, P.M. (1984) J. Mol. Biol. 178:249.
53. Sogo, J.M., Stahl, H., Koller, Th. and Knippers, R. (1986) J. Mol. Biol. 189:189.
54. Stillman, B. (1986) Cell 45:555.
55. Varshavsky, A., Sundin, O., Ozkaynak, E., Pan, R., Solomon, M. and Snapka, R. (1983) In: Mechanisms of DNA Replication and Recombination, (N.R. Cozzarelli, ed.), Alan R. Liss, N.Y., p. 463.
56. Weintraub, H. (1979) Nucl. Acids Res. 7:781.
57. Gargiulo, G., Razvi, F., Ruberti, I, Mohr, I. and Worcel, A. (1985) J. Mol. Biol. 181:333.
58. Cereghini, S. and Yaniv, M. (1984) EMBO J. 3:1243.
59. Murakami, Y., Toshihiko, E., Yamada, M., Prives, C., and Hurwitz, J. (1986) Proc. Natl.

Acad. Sci. USA 83:6347.

60. DeLucia, A.L., Deb, S., Partin, K. and Tegtmeyer, P. (1986) J. Virol. 57:138.
61. Lee-Chen, G-J., and Woodworth-Gutai, M. (1986) Mol. Cell. Biol. 6:3086.
62. Lee-Chen, G-J., and Woodworth-Gutai, M. (1986) Mol. Cell. Biol. 6:3077.
63. Deb, S., DeLucia, A.L., Baur, C-P., Koff, A., and Tegtmeyer, P. (1986) Mol. Cell. Biol. 6:1663.
64. Li, J.J., Peden, K.W.C., Dixon, R.A.F. and Kelly, T. (1986) Mol. Cell. Biol. 6:1117.
65. Stillman, B., Gerard, R.D., Guggenheimer, R.A. and Gluzman, Y. (1985) EMBO J. 4:2933.
66. Hassell, J.A., Muller, W.J., and Mueller, C.R. (1986) In: Cancer Cells, vol. 4 (Botchan, M., Grodzicker, T. and Sharp, P. (eds)), Cold Spring Harbor Laboratories, p. 561.
67. Veldman, G.M., Lupton, S. and Kamen, R. (1985) Mol. Cell. Biol. 5:649.
68. Mueller, C.R., Mes-Masson, A.M., Bouvier, M. and Hassell, J.A. (1984) Mol. Cell. Biol. 4:2594.
69. Herbomel, P. Bourachot, B. and Yaniv, M. (1984) Cell 39:653.
70. Clertant, P., Gaudray, P., May, E. and Cuzin, F. (1984) J. Biol. Chem. 259:15196.
71. Clark, R., Peden, K., Pipas, J.M., Nathans, D. and Tjian, R. (1983) Mol. Cell. Biol. 3:220.
72. Lane, D.P. and Crawford, L.V. (1979) Nature 278:261.
73. McCormick, F. and Harlow, E. (1980) J. Virology 34:213.
74. Stahl, H., Droge, P. and Knippers, R. (1986) EMBO J. 5:1939.
75. Hertz, G.Z. and Mertz, J. (1986) Mol. Cell. Biol. 6:3513.
76. Mercer, W.E., Nelson, D., Hyland, J.K., Croce, C.M. and Baserga., R. (1983) Virology 127:149.
77. Tegtmeyer, P. (1972) J. virology 10:591.
78. Francke, B. and Eckhart, W. 1973. Virology 55:127.
79. Chou, J.Y., Avila, J. and Martin, R.G. (1974) J. Virology 14:116.
80. Dinter-Gottlieb, G. and Kaufmann, G. (1982) Nucl. Acids Res. 10:763.
81. Stahl, H., Droge, P., Zentgraf, H. and Knippers, R. (1985) J. Virology 54:473.
82. Tack, L.C. and DePamphilis, M.L. (1983) J. Virology 48:281.
83. Tack, L.C., Wright, and Gurney, E.G. (1986) J. Virology 58:635.
84. Vogt, B., Vakalopoulou, E. and Fanning, E. (1986) J. Virology 58:765.
85. Tjian, R. (1978) Cell 13:165.
86. Tegtmeyer, P., Lewton, B.A., DeLucia, A.L., Wilson, V.G. and Ryder, K. (1983) J. Virology 46:151.
87. DeLucia, A.L., Lewton, B.A., Tjian, R. and Tegtmeyer, P. (1983) J. Virology 46:143.
88. Pomerantz, B.J., and Hassell, J.A. (1984) J. Virology 49:925.

89. Cohen, G.L., Wright, P.J., DeLucia, A.L., Lewton, B.A., Anderson, M.E. and Tegtmeyer, P. (1984) J. Virology 51:91.
90. Ryder, K., Vakalopoulou, E., Mertz, R., Mastrangelo, I, Hough, P., Tegtmeyer, P., Fanning, E. (1985) Cell 42:539.
91. Ryder, K., Silver, S., DeLucia, A.L., Fanning, E. and Tegtmeyer, P. (1986) Cell 44:719.
92. Dilworth, S.M., Cowie, A., Kamen, R.I., and Griffin, B.E. (1984) Proc. Natl. Acad. Sci. USA 81:1941.
93. Cowie, A. and Kamen, R. (1986) J. Virology 57:505.
94. Gottlieb, P., Nasoff, M.S., Fisher, E.F., Walsh, A.M. and Caruthers, M.H. (1985) Nucl. Acids Res. 13:6621.
95. DeMaio, D. and Nathans, D. (1982) J. Mol. Biol. 156:531.
96. Meyers, R.M. and Tjian, R. (1980) Proc. Natl. Acad. Sci. USA 77:6491.
97. Jones, K.A. and Tjian, R. (1984) Cell 36:155.
98. Margolskee, R.F. and Nathans, D. (1984) J. Virology 49:386.
99. Scheidtmann, K., Echle, B. and Walter, G. (1982) J. Virology 44:116.
100. van Roy, F., Fransen, L. and Fiers, W. (1983) J. Virology 45:315.
101. Smale, S.T. and Tjian, R. (1986) Mol. Cell. Biol. 6:4077.
102. Oren, M., Winocour, E., and Prives, C. (1980) Proc. Natl. Acad. Sci. USA 77:220.
103. Scheidtmann, K., Hardung, M., Echele, B. and Walter, G. (1984) J. Virology 50:636.
104. Scheidtmann, K., Schickedanz, J., Walter, G., Lanford, R. and Butel, J.S. (1984) J. Virology 50:636.
105. Simmons, D.T. (1984) J. Biol. Chem. 259:8633.
106. Jarvis, D.L., and Butel, J.S. (1985) Virology 141:173.
107. Bradley, M.K., Hudson, J., Villaneuva, M.S. and Livingston, D. (1982) Proc. Natl. Acad. Sci. USA 81:6574.
108. Goldman, N.D., Brown, M. and Khoury, G. (1981) Cell 24:567.
109. Klockmann, U., Staufenbiel, M. and Deppert, W. (1984) Mol. Cell. Biol. 4:1542.
110. Bradley, M.K., Griffin, J.D. and Livingston, D.M. (1982) Cell 28:125.
111. Fanning, E., Westphal, K-H., Brauer, D., and Corlin, D. (1982) EMBO J. 1:1023.
112. Gidoni, D., Scheller, A., Barnet, B., Hantzopoulos, P., Oren, M. and Prives, C. (1982) J. Virology 42:456.
113. Burger, C. and Fanning, E. (1983) Virology 126:19.
114. Yang, L., Wold, M.S., Li, J.J., Kelly, T.J. and Liu, L.F. (1987) Proc. Natl. Acad. Sci. USA 84:950.
115. Chia, W. and Rigby, P.W.J. (1981) Proc. Natl. Acad. Sci. USA 78:6638.
116. Cambell, B.A. and Villarreal, L.P. (1985) Mol. Cell. Biol. 5:1534.
117. Campbell, B.A. and Villarreal, L.P. (1986) Mol. Cell. Biol. 6:2068.
118. Bennett, E.R., Naujokas, M., Collins, C. and Hassell, J.A. (1986) Cold Spring Harbor

DNA Tumor Virus Meeting.

119. deVilliers, J., Olson, L., Tyndall, C. and Schaffner, W. (1982) Nucl. Acids Res. 10:7965.
120. deVilliers, J., Schaffner, W., Tyndall, C., Lupton, S. and Kamen, R. (1984) Nature 312:242.
121. Murakami, Y., Wobbe, C.R., Weissbach, L., Dean, F.B. and Hurwitz, J. (1986) Proc. Natl. Acad. Sci. USA 83:2869.
122. Weber, F. , deVilliers, J., and Schaffner, W. (1984) Cell 36:983.
123. Chalifour, L.E., Wirak, D.O., Wassarman, P.M., Hansen, U. and DePamphilis, M.L. (1986), J. Virology 59:619.
124. DeSimone, V., LaMantia, G., Lania, L., and Amati, P. (1985) Mol. Cell. Biol. 5:2142.
125. Melin, F., Pinon, C., Reiss, C., Kress, C., Montreau, N. and Blangy, D. (1985) EMBO J. 4:1799.
126. Barsoum, J. and Berg, P. (1985) Mol. Cell. Biol. 5:3048.
127. Piette, J., Kryszke, M-H, and Yaniv, M. (1985) EMBO J. 4:2675.
128. Fujimura, F.K. (1986) Nucl. Acids Res. 14:2845.
129. Muller, W.J., Mueller, C.R., Mes, A-M., and Hassell, J.A. (1983) J. Virol. 47:586.
130. Mechali, M. and Kearsey, S. (1984) Cell 38:55.
131. Blumenthal, A.B., Kriegstein, H.J. and Hogness, D.S. (1973) Cold Spring Harbor Symp. Quant. Biol. 38:205.
132. Newport, J. and Kirschner, M. (1982) Cell 30:675 and 687.
133. Clegg, K.B. and Piko, L. (1982) Dev. Biol. 95:331.
134. Pomerantz, B.J., Mueller, C.R. and Hassell, J.A. (1983) J. Virology 47:600.97.
135. Lewton, B.A., DeLucia, A.L. and Tegtmeyer, P. (1984) J. Virology 49:9.
136. Tenen, D.G., Taylor, T.S., Haines, L.L., Bradley, M.K., Martin, R.G. and Livingston, D.M. (1983) J. Mol. Biol. 168:791.
137. DePamphilis, M.L., Decker, R.S., Yamaguchi, M., Possenti, R., Wirak, D.O., Perona, R., and Hassell, J.A. (1986) In: Mechanisms of DNA Replication and Recombination (T. Kelly and R. McMacken, eds), Alan R. Liss, NY.
138. Yamaguchi, M. and DePamphilis, M.L. (1986) Proc. Natl. Acad. Sci. (USA) 83:1646.
139. Chen, S.S. and Hsu, M.-T. (1984) J. Virology 51:14.
140. Luchnik, A.N., Bakayev, V.V., Zbarsky, I.B. and Georgiev, G.P. (1982) EMBO J. 1:1353.
141. Dinter-Gottlieb, G. and Kaufmann, G. (1982) Nucl. Acids Res. 10:763.
142. Wang, T.S.-F., Hu, S.-Z., and Korn, D. (1984) J. Biol. Chem. 259:1854.
143. Plevalani, P., Badaracco, G., Augl, C. and Chang, L.M.S. (1984) J. Biol. Chem. 259:7532.
144. Tseng, B.Y. and Ahlem, C.N. (1982) J. Biol. Chem. 257:7280.

145. Tapper, D.P. and DePamphilis, M.L. (1980) Cell 22:97.
146. Buckler-White, A.J., Krauss, M.R., Pigiet, V. and Benbow, R.M. (1982) J. Virology 43:885.
147. Martin, R.G. and Setlow, V.P. (1980) Cell 20:381.
148. Bjursell, G., Munck, V. and Therkelsen, A.J. (1979) J. Virology 30:929.
149. Tapper, D.P. and DePamphilis, M.L. (1978) J. Mol. Biol. 120:401.
150. Seidman, M. and Salzman, N.P. (1979) J. Virology 30:600.
151. Herman, T.M., DePamphilis, M.L. and Wassarman, P.M. (1979) Biochemistry 18:4563.
152. Sundin, O. and Varshavsky, A. (1980) Cell 21:103.
153. Liu, L.F., Halligan, B.D., Nelson, E.M., Rowe, T.C., Chen, G.L. and Tewey, K.M. (1983) In: Mechanisms of DNA Replication and Recombination (N.R. Cozzarelli, ed), Alan R. Liss, NY, p. 43.; Sander, M. and Hsieh, T-S. (1985) Nucl. Acids Res. 13:1057.
154. Gourlie, B.B. and Pigiet, V.P. (1983) J. Virology 45:585.
155. Brockman, W.W., Gutai, M.W. and Nathans, D. (1975) Virology 66:115.
156. Weaver, D.T., Fields-Berry, S.C., and DePamphilis, M.L. (1985) Cell 41:565.
157. Bjursell, G., Skoog, L., Thelander, L. and Soderman, G. (1977) Proc. Natl. Acad. Sci. USA 74:5310.
158. Bjursell, G. and Magnusson, G. (1976) Virology 74:249.
159. Martin, R.F. (1977) J. Virology 23:827.
160. Danna, K.J. and Nathans, D. (1971) Proc. Natl. Acad. Sci. USA 69:3097.
161. Lai, C-J. and Nathans, D. (1974) J. Mol. Biol. 97:113.
162. Tapper, D., Anderson, S. and DePamphilis, M.L. (1979) Biochem. Biophys. Acta 565:84.
163. Buckler-White, A.J. and Pigiet, V. (1982) J. Virology 44:499.
164. Zannis-Hadjopoulos, M., Chepelinsky, A.B. and Martin, R.G. (1983) J. Mol. Biol. 165:599.
165. Weaver, D.T. and DePamphilis, M.L. (1984) J. Mol. Biol. 180:961.
166. Chen, M.C.Y., Birkenmeier, E. and Salzman, N.P. (1976) J. Virology 17:614.
167. Sundin, O. and Varshavsky, A. (1981) Cell 25:659.
168. Goto, T. and Wang, J.C. (1982) J. Biol. Chem. 257:5866.
169. DiNardo, S., Voelkel, K., and Sternglanz, R. (1984) Proc. Natl. Acad. Sci. USA 81:2626.
170. Uemura, T., and Yanagida, M. (1984) EMBO J. 3:1737.
171. Champoux, J.J. and Been, M.D. (1980) In: Mechanistic Studies of DNA Replication and Genetic Recombination, (B. Alberts, ed.) Alan R. Liss, N.Y, p. 809.
172. Krauss, M.R., Gourlie, B.B., Bayne, M.L. and Benbow, R.M. (1984) J. Virology 49:333.
173. Saragosti, S., Moyne, G. and Yaniv, M. (1980) Cell 20:65.

2

POLYOMAVIRUS SEQUENCES AFFECTING THE INITIATION OF TRANSCRIPTION AND DNA REPLICATION

WILLIAM R. FOLK, W. J. TANG, M. MARTIN, J. LEDNICKY, S. BERGER AND R. H. ADAMS

Department of Microbiology, The University of Texas at Austin, Austin, Texas 78712

ABSTRACT

The location of the polyomavirus early mRNA 5' termini is determined primarily by the ATA homology, cap sites and nearby upstream sequences. Regulation of mRNA levels occurs by large T-antigen acting at sites distal to the point of transcription initiation. As yet, there is little evidence for proximal upstream activator sequences affecting early transcription.

The polyomavirus origin contains at least four domains, a T-A rich region which borders the enhancer, a G·C rich palindrome with multiple large T-antigen binding motifs, the sites at which bidirectional replication is initiated, and a region to which large T-antigen and cellular proteins may bind. The relationship of each of these to the other, and to the process by which DNA replication is initiated remains unclear.

The polyomavirus enhancer contains at least three elements which act in a positive or negative manner upon transcription and DNA replication in various cell types. The limits of these elements as well as some of the proteins binding to them are being defined.

INTRODUCTION

Polyomavirus regulates its transcription and DNA replication in a precise, coordinated manner. Soon after infection of permissive cells, transcription of the viral early genes leads to the synthesis of large, middle and small T-antigens. Large T-antigen, a multifunctional protein, acts upon the viral DNA in concert with cellular proteins to initiate DNA replication. The role of the other two viral early proteins during vegetative growth of the virus is unclear.

As viral genomes and early proteins accumulate within the cell, expression of the viral DNA shifts to favor the late genes. By the end of the infectious cycle, approximately 50-fold more late transcripts are made than early transcripts (1); this balance is dependent upon large T-antigen acting as a repressor of early transcription (1-3), and as an activator of the early and late

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

promoters (R. Adams and W. R. Folk, unpublished). Synthesis of the late proteins leads to encapsidation of the viral DNA and ultimately, to cell death and release of progeny viruses.

The life cycles of polyomavirus and SV-40 are very similar, as are the proteins they encode (4-6). Remarkably, however, their regulatory sequences differ appreciably. It is too early to tell how these differences are reflected in function, but the analysis of each virus is providing new information about the role of individual sequence elements in regulating transcription and DNA replication.

The carefully orchestrated steps of the viral life cycle occur within a complex cell milieu which provides most of the proteins required for trancription and DNA replication, as well as the nucleotide precursors. In large part, polyomavirus and SV-40 provide DNA templates and one or two regulatory proteins. These are sufficient to abrogate the normal cellular processes and to induce the cell to synthesize a large amount of viral DNA and the proteins which encapsidate it. Studying how these viruses subvert the cell is providing numerous insights into how the animal cell regulates its own biosynthetic processes, apart from those of the virus.

EXPERIMENTAL RESULTS AND DISCUSSION

The Polyomavirus Early Promoter

After infection of permissive cells, at early times or at late times in the absence of a functional large T-antigen, two principal mRNAs are transcribed from the polyomavirus early region with 5' termini 25-35 nucleotides (nt) downstream from the ATA homology, at nt 134-140 (Strain A-3 numbering system of Friedmann et al. ; 4).

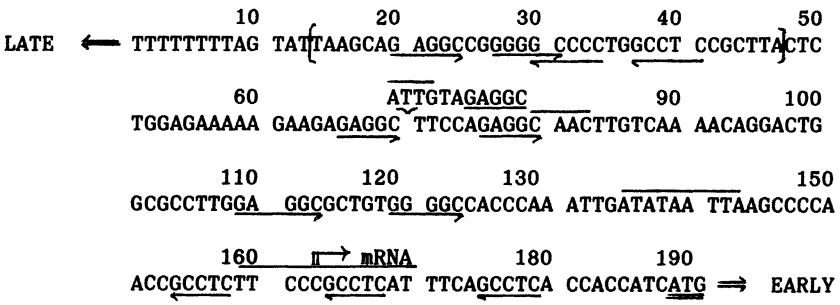


Fig. 1. Nucleotide Sequence of the Polyomavirus Origin and Early Promoter Region. Strain A-2 contains an 11 bp insertion at nt 70 (6). Large T-antigen binding sites are underlined. Origin palindrome is bracketed. Location of 5' termini of early mRNAs and initiation codon used for translation of early proteins are indicated. ATA and CAAT homologies are overlined.

The polyomavirus early promoter ATA homology and cap sites are physically separate from the origin of DNA replication, unlike those of SV-40 (7,8). We were able to isolate viable viruses lacking these transcriptional elements. Initial characterization of several such mutants indicated that their early transcription was not greatly impaired(9).Deletion of nt 85-187 (including the ATA homology and the cap site; Fig. 1) resulted in the formation of heterogeneous 5' termini extending throughout the early promoter-origin region, from nt 5260 to within the coding sequences at nt 188 (7). The amount of early mRNA initiated was not markedly affected by loss of these sequences. In other laboratories, indirect measurements of transcription which relied upon viral transformation or synthesis of large T-antigen gave similar results (10-12).

Using cell-free extracts from human cells, analysis of the transcription of DNAs with deletions throughout the early promoter region indicated that deletion of the ATA homology reduced the number of mRNAs initiated in vitro, but only slightly increased the heterogeneity of mRNA 5' termini. Only when the downstream cap region was also deleted were mRNAs with new 5' termini prominent (13). This contrasted with the expression of several viable deletion mutants in mouse fibroblasts where it appeared that the integrity of the ATA homology and nearby upstream sequences determined the homogeneity of 5' termini (14). The reason for this difference might lie in the sequence specificity of murine and human transcription factors interacting with the ATA homology region (15,16).

There is little evidence suggesting that other proximal sequences upstream of the ATA homology are required for function of the polyomavirus early promoter. A CAACT sequence occurs at nt 80-85, approximately 80-90 bp upstream of the major wild-type mRNA 5' termini (Fig. 1), and the polyomavirus strain A-2 contains a second similar motif, a CAAT sequence, on the opposite strand within the 11 bp insertion in this region. One or both of these elements might serve as binding sites for transcriptional activators. These sequences impinge upon the origin of DNA replication (9,12,17,18) hence it has not been possible to fully evaluate their importance with viable deletion mutants. Indirect measurements of expression (transformation or T-antigen expression in Hela cells) failed to detect a significant effect of deleting these sequences in DNA constructs (10-12). However, insertion of 134 bp of foreign DNA between these sequences and the ATA homology induced marked heterogeneity of in vivo early mRNA 5' termini in a viable insertion mutant (19) implying that an upstream activator may exist. We are

currently determining whether transcription factors which recognize homologs to the CAACT sequence (20,21) will footprint this region.

To further explore whether proximal upstream activator sequences are required for polyomavirus transcription, we are using a general strategy that should detect a variety of sequences capable of activating RNA polymerase II promoters. It is based conceptually upon the "enhancer trap" experiments of Weber et. al. (22) except that the sequences trapped should be promoter-proximal activators. An SV-40 construct lacking the 21 bp repeats (and in some cases, one of the 72 bp repeats) was generated by joining fragments from two SV-40 deletion mutants S312 and X113 (23). Without a proximal activator such as the 21 bp repeats, SV-40 transcription is greatly reduced (23-25), resulting in the virus forming very small plaques at delayed times. Although the 21 bp repeats also affect DNA replication in vivo (23,26) their effect is not pronounced when an enhancer element is present (27,28).

Transient expression assays have implicated several groups of sequences on the late side of the origin of replication as being possible promoter elements (11,29). However, their significance in vivo is at present unclear. Up to the present, we have failed to detect any polyomavirus proximal activators between nt 5150-nt 80 using the promoter activator trap procedure (J. Lednicky and W. R. Folk, unpublished). We are currently screening other parts of the polyomavirus regulatory region for proximal activator sequences.

Large T-antigen plays a major role in modulating the expression of the early promoter. An approximately 50-fold shift in the transcription of the early and late promoters occurs during the viral life cycle within cells containing a functional large T-antigen. Some of this is due to repression of transcription; however, how such repression occurs is not totally clear. Large T-antigen binds to multiple sites throughout the early promoter-origin region (30-33). Deletion of the binding sites closest to the ATA homology had only a small effect upon the capacity of T-antigen to repress transcription (1,12). Thus, if repression is exerted by T-antigen interacting with the viral DNA, the upstream binding sites must also play an important role.

Two models for how T-antigen might repress transcription by binding to the upstream sites seem plausible. First, T-antigen might block activation of the early promoter by impairing translocation of protein(s) from the enhancer to the ATA homology/cap site region, or by preventing a DNA conformational change induced by the enhancer. Alternatively, T-antigen might sterically block an upstream site required for a positive-acting transcription factor.

Determining which of these, or any other mechanism(s), operate is difficult with viable deletion mutants, for the sequences implicated in this repression extend through the origin of DNA replication. With transient expression assays, the extent of repression by T-antigen observed is usually only several-fold, and does not approach the level observed *in vivo* during virus growth (R. Adams and W. R. Folk, unpublished).

The polyomavirus origin of DNA replication

Sequences that are essential for the initiation of DNA replication have been roughly mapped to an 60 bp region immediately to the early side of the PvuII 1/4 junction (nt 1-60, Fig. 1) (9,11,12,17,34). Sequences to the late side of this region, which include the enhancer, also contribute to replication (11,34-36,37; W. J. Tang, S. L. Berger, S. J. Triezenberg and W. R. Folk, Mol. Cell. Biol., in press, 1987).

Essential sequences near the PvuII 1/4 junction include a string of eight T/A bp. Deletion of part of these inactivates the origin (34). This sequence may serve as a recognition site for a protein, or its physical structure may be important to origin function. Another prominent landmark in the origin region is a 40 bp imperfect palindrome which is highly conserved in genomes of all members of the polyomavirus genus. Polyomavirus large T-antigen binds to repeated $(Pu)_2GGC$ motifs within this sequence (30-33,38). Point mutants in these sequences between nt 25-40 inactivate viral replication (39,40). *In vitro*, T-antigen binding to several of these mutated origins is reduced several fold (A. Cowie, R. Kamen, S. Triezenberg and W. Folk, unpublished).

Between nt 40-70 is a third region of importance for DNA replication (9,12,17). Although the few single point mutants throughout this region which have been tested display normal, or near-normal origin function, a mutant with a cluster of five G→A changes between nt 54-68 is replication defective (M. Sullivan, W. J. Tang and W. Folk, unpublished). The transition between continuous and discontinuous strand synthesis occurs within this region (E. Hendrickson, C. Fritze, W. Folk and M. DePamphilis, unpublished).

The origin sequences between nt 1-70 define a functional unit whose orientation with respect to adjoining sequences is relatively flexible. This region can be inverted without destroying viability of the virus (M. Sullivan and W. Folk, unpublished), indicating that sequences outside the origin which are required for DNA replication must not act upon the origin in a polar fashion.

The polyomavirus enhancer

The polyomavirus enhancer sequences are located on the late side of the origin (Fig. 2).

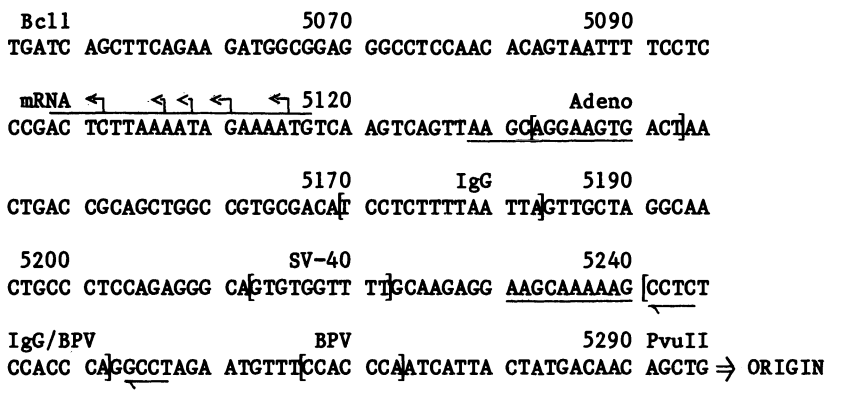


Fig. 2. Nucleotide sequence of the polyomavirus enhancer and late promoter region. Locations of sequences with homologies to the enhancers of Adenovirus 5E1A, immunoglobulin heavy chain, SV-40 and BPV enhancers are bracketed. DNAse I hypersensitive sites are underlined, and region specifying major late mRNA 5' termini is overlined. Large T-antigen binding sites are indicated by arrows.

It is structurally unique, yet it contains functionally redundant elements (11,35,41-43). For convenience, the BclI-PvuII and PvuII-4 fragments have been used to delimit the two major functional regions. Within each of these fragments are located sequences with homology to other viral and cellular enhancers. Most notable among these are homologies to the Adenovirus 5 E1A enhancer (nt 5132- nt 5143) to the IgG enhancer (nt 5165-nt 5187), to the SV-40 enhancer (nt 5213-nt 5221) and to the BPV enhancer (nt 5245-nt 5252; nt 5266-nt 5273) (44-48).

Some laboratory strains of polyomavirus contain repeated copies of sequences including the adenovirus enhancer homology (49,50). Mutants of polyomavirus selected to grow on non-fibroblastic cells frequently duplicate this region as well, or may transpose it and delete sequences in the PvuII-4 fragment (51-54). Mutants selected for growth on F9 embryonal carcinoma cells have, as a common denominator, a single A→G change at nt 5258 (55,56). These rearrangements must reflect the jockeying for position among positive- and negative-acting sequence elements which determine enhancer activity in different cell types.

Several extensive deletion analyses of the polyomavirus enhancer indicate that the essential elements within the Bcl-PvuII fragment which in fibroblasts act in concert with the origin of DNA replication lie between nt 5132-nt 5155 (42,57). This region is hypersensitive to DNAse 1 in polyomavirus chromatin (58,59). It is the binding site for one or more cellular protein(s) (J. Piette and M. Yaniv, personal communication) We have been unable to detect protection of these sequences against DNAse I in chromatin, which suggests that only a small fraction of the polyomavirus chromosomes contain associated proteins (M. Martin and W. R. Folk, unpublished).

Less precise limits have been placed upon the essential elements for DNA replication and transcription in fibroblasts within the PvuII-4 fragment, although most analyses indicate that the sequences between nt 5176-nt 5229 (which contain homologies to the IgG and SV-40 enhancers) are of singular importance (29,42,57). Several cellular proteins binding to sequences within this region have been detected (60-62). These sequences may interfere with virus growth in non-fibroblastic cells, as deletions in this region assist virus infection of mouse trophoblasts and lymphoid cells (52,63).

A third functional domain within the polyomavirus enhancer must include the A→G substitution at nt 5258 which permits DNA replication and transcription in undifferentiated F9 cells (56). This nucleotide change makes the surrounding sequences more homologous to the SV-40 core enhancer element, and to the TGGCA protein (64) binding sequence. However, it is presently unclear if this sequence homology is of functional significance.

Our approach to analyzing the multitudinous elements and functions of the polyomavirus enhancer has been to generate and analyze mutants with G·C→A·T substitutions in this region, so as to define critical nucleotides important for function and relate the effects of those changes to alterations in protein binding and/or DNA structure. Using bisulfite mutagenesis (65,66), we isolated over 70 mutants with single or multiple base changes in the enhancer region (67). Analysis of the transcription and replication of a small number of these mutants, and of viable revertants derived from them, has allowed us to draw the following conclusions (W. J. Tang, S. L. Berger, S. Triezenberg and W. R. Folk, Mol. Cell. Biol, in press): 1) The sequence homologous to the Adenovirus E1A enhancer is critical for the function of the enhancer element(s) in the Bcl-PvuII fragment. Viral replication and transcription in fibroblasts are depressed when nt 5134 and nt 5140 are altered; 2) within the PvuII-4 fragment, the SV-40-homologous element is important for

replication and transcription, as when nt 5215 and nt 5218 are altered, enhancer function is reduced. Either or both nt 5192 and nt 5227 in the IgG enhancer homology are important for its function; 3) When both the Bcl-PvuII and PvuII-4 enhancer domains are inactivated by point mutations, enhancer function can be restored by a nucleotide change at 5258. We believe that a positive-acting element is created by this change (which is the same as that observed in polyomavirus variants adapted to F-9 cells); however, we have not excluded that this mutation inactivates a repressor-binding site (68). Our analysis of these mutants is incomplete, and additional mutants need to be characterized before we can fully delineate the important nucleotides required for enhancer function in fibroblasts.

Resolving how the enhancer activates both transcription and DNA replication will require the development of *in vitro* transcription and replication systems which depend upon enhancer sequences. Then, we will be able to test whether proteins binding to the enhancer stimulate the formation of preinitation complexes assembling at the origin and promoters (W. J. Tang, S. L. Berger, S. Triezenberg and W. R. Folk, Mol. Cell. Biol., in press).

The enhancer region contains, in addition to these positive- and negative-acting elements, the sequences responsible for the initiation of late mRNA synthesis (69,70). The essential elements of the late promoter remain to be defined, and the question of how the virus modulates expression of the late genes remains to be answered.

ACKNOWLEDGMENTS

We thank Susan Crossland for her expert help in preparing this manuscript. Support from the U.S.P.H.S. and the American Cancer Society is gratefully acknowledged.

REFERENCES

1. Farmerie, W. G. and Folk, W. R. Proc. Natl. Acad. Sci. USA 81:6919-6923, 1984.
2. Cogen, B. Virology 85:222-230, 1978.
3. Fenton, R. G. and Basilico, C. Virology 121:384-392, 1982.
4. Friedmann, T., Esty, A., LaPorte, P. and Deininger, P. Cell 17:715-724, 1979.
5. Deininger, P. A., Esty, A., LaPorte, P. and Friedmann, T. Cell 18:771-779, 1979.
6. Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. and Griffin, B. E. Nature 283:445-453, 1980.
7. Kamen, R., Jat, P., Treisman, R., Favaloro, J. and Folk, W. R. J. Mol. Biol. 159:189-224, 1982.
8. Heiser, W. C. and Eckhart, W. J. Virol. 44:175-188, 1982.
9. Bendig, M. M., Thomas, T. and Folk, W. R. Cell 20:401-409, 1980.
10. Katinka, M. and Yaniv, M. Mol. Cell. Biol. 2:1238-1246, 1982.
11. Mueller, C. F., Mes-Masson, A. M., Bouvier, M. and Hassell, J. Mol. Cell. Biol. 4:2594-2609, 1984.
12. Dailey, L. and Basilico, C. J. Virol. 54:739-749, 1985.
13. Jat, P., Novak, V., Cowie, A., Tyndall, C. and Kamen, R. Mol. Cell. Biol. 2:737-751, 1982.
14. Farmerie, W. G. and Folk, W. R. Virology 150:518-523, 1986.
15. Sawadogo, M. and Roeder, R. G. Cell 43:165-175, 1985.
16. Carthew, R. W., Chadoch, L. A. and Sharp, P. A. Cell 43:439-448, 1985.
17. Katinka, M. and Yaniv, M. J. Virol. 47:244-248, 1983.
18. Muller, W. J., Mueller, C. R., Mes, A. M. and Hassell, J. A. J. Virol. 47:586-599, 1983.
19. Clark, K. L., Bendig, M. M. and Folk, W. R. J. Virol. 52:1032-1035, 1984.
20. Jones, K. A., Yamamoto, K. R. and Tjian, R. Cell 42:559-572, 1985.
21. Graves, B. J., Johnson, P. F. and McKnight, S. L. Cell 44:565-576, 1986.
22. Weber, F., deVilliers, J. and Schaffner, W. Cell 36:983-992, 1984.
23. Fromm, M. and Berg, P. J. Mol. Appl. Genet. 1:457-481, 1982.
24. Benoist, C. and Chambon, P. Nature 290:304-310, 1981.
25. Gruss, P., Dhar, R. and Khoury, G. Proc. Natl. Acad. Sci., USA 78:943-947, 1981.
26. Bergsma, D. F., Olive, D. M., Hartzell, S. W. and Subramanian, K. N. Proc. Natl. Acad. Sci. USA 79:381-385, 1982.
27. Li, J., Peden, K. W., Dixon, R. and Kelly, T. Mol. Cell. Biol. 6:1117-1128, 1986.
28. DeLucia, A., Deb, S., Partin, K. and Tegtmeyer, P. J. Virol. 57:138-144, 1986.
29. Bohnlein, E., Chowdhury, K. and Gruss, P. Nucl. Acids Res. 13:4789-4809, 1985.
30. Gaudray, P., Tyndall, C., Kamen, R. and Cuzin, F. Nucl. Acids Res. 9:5697-5710, 1981.
31. Pomerantz, B. J., Mueller, C. and Hassell, J. J. Virol. 47:600-610, 1983.

32. Dilworth, W. M., Cowie, A., Kamen, R. and Griffin, B. Proc. Natl. Acad. Sci. 81:1941-1945, 1984.
33. Cowie, A. and Kamen, R. J. Virol. 52:750-760, 1984.
34. Luthman, H., Nilsson, G. and Magnusson, G. J. Mol. Biol. 161:533-550, 1982.
35. Tyndall, C., LaMantia, G., Thacker, M., Favaloro, J. and Kamen, R. Nucl. Acid. Res. 9:6231-6249, 1981.
36. Fujimura, F. and Linney, E. Proc. Natl. Acad. Sci. USA 79:1479-1483, 1982.
37. deVillers, J., Schaffner, W., Tyndall, C., Lupton, S. and Kamen, R. Nature 312:242-246, 1984.
38. Cowie, A. and Kamen, R. J. Virol. 57:505-514, 1986.
39. Triezenberg, S. and Folk, W. R. J. Virol. 51:437-444, 1984.
40. Luthman, H., Osterlund, M. and Magnusson, G. Nucl. Acid. Res. 12:7503-7515, 1984.
41. Herbomel, P., Bourachot, B. and Yaniv, M. Cell 39:653-662, 1984.
42. Veldman, G., Lupton, S. and Kamen, R. Mol. Cell. Biol. 5:649-658, 1985.
43. deVillers, J. and Schaffner, W. Nucl. Acid. Res. 9:6251-6264, 1981.
44. Hearing, P. and Shenk, T. Cell 33:695-703, 1983.
45. Hen, R., Borelli, E., Sassone-Corsi, P. and Chambon, P. Nucl. Acid. Res. 11:8747-8759, 1983.
46. Weiher, H., Konig, M. and Gruss, P. Science 219:626-631, 1983.
47. Weiher, H. and Botchan, M. R. Nucl. Acid. Res. 12:2901-2916, 1984.
48. Banerji, J., Olson, L. and Schaffner, W. Cell 33:729-755, 1983.
49. Ruley, H. E. and Fried, M. J. Virol. 47:233-237, 1983.
50. Rothwell, V. and Folk, W. R. J. Virol. 48:472-480, 1983.
51. Katinka, M., Yaniv, M., Vasseur, M. and Blangy, D. Cell 20:393-399, 1980.
52. Tanaka, K., Chowdhury, K., Chang, K. S., Israel, M. and Ito, Y. EMBO Journal 1:1521-1527, 1982.
53. Melin, F., Pinon, H., Reiss, C., Kress, C., Montreau, N. and Blangy, D. EMBO Journal 4:1799-1803.
54. Amati, P. Cell 43:561-562, 1985.
55. Sekikawa, K. and Levine, A. J. Proc. Natl. Acad. Sci. USA 78:1100-1104, 1981.
56. Fujimura, F., Deininger, P. L., Friedmann, T. and Linney, E. Cell 23:809-814, 1981.
57. Hassell, J., Muller, W. J. and Mueller, C. R. in Cancer Cells Vol. 4:561-569 (M. Botchan, T. Grodzicker and P. Sharp, eds.). Cold Spring Harbor Press, 1986.
58. Herbomel, P., Saragosti, S., Blangy, D. and Yaniv, M. Cell 25:651-658, 1981.
59. Bryan, P. and Folk, W. R. Mol. Cell. Biol. 6:2249-2252, 1986.
60. Piette, J., Kryszke, M. H. and Yaniv, M. EMBO Journal 4:2675-2685, 1985.
61. Fujimura, F. K. Nucl. Acid. Res. 14:2845-2861, 1986.
62. Bohnlein, E. and Gruss, P. Mol. Cell. Biol. 6:1401-1411, 1986.
63. Campbell, B. A. and Villareal, L. P. Mol. Cell Biol. 6:2068-2079, 1986.
64. Nowock, J., Borgmeyer, V., Puschell, R. W., Rupp, R. A. and Sippel, A. E. Nucl. Acid. Res. 13:2045-2061, 1985.

65. Shortle, D. and Nathans, D. Proc. Natl. Acad. Sci. USA 75:2170-2174, 1978.
66. Folk, W. R. and Hofstetter, H. Cell 33:585-593, 1983.
67. Triezenberg, S. Ph.D. Thesis, Univ. of Michigan, 1984.
68. Hen, R., Borelli, E., Fromental, C., Sassone-Corsi, P. and Chambon, P. Nature 321:249-251, 1986.
69. Cowie, A., Tyndall, C. and Kamen, R. Nucl. Acid. Res 9:6305-6322, 1981.
70. Kern, F. G., Dailey, L. and Basilico, C. Mol. Cell. Biol. 5:2070-2074.

3

THE SV40 EARLY PROMOTER

M. ZENKE*, A. WILDEMAN** and P. CHAMBON

Laboratoire de Génétique Moléculaire des Eucaryotes du CNRS,
Unité 184 de Biologie Moléculaire et de Génie Génétique de l'INSERM,
Faculté de Médecine - 11, rue Humann - 67085 Strasbourg-Cédex - France

ABSTRACT

The SV40 early promoter consists of the TATA-box, which directs the transcriptional machinery/RNA polymerase to the specific startsites of transcriptional initiation, and the 21-bp repeat region and enhancer, both of which are indispensable for efficient early transcription. Specific sequence motifs in the different promoter elements interact with specific trans-acting transcription factors to form a large nucleoprotein complex at the promoter. It is the coordinate functioning of the various promoter elements and their cognate protein factors which is required to promote transcription.

INTRODUCTION

The simian virus 40 (SV40) early promoter controls the expression of the early transcription unit of SV40 both in the lytic cycle of viral infection and in virus-transformed cells. To date the SV40 early promoter is one of the best characterized control regions for initiation of transcription by RNA polymerase class B (II), and has proven to be an extremely useful model system to study the molecular mechanisms controlling the expression of protein-coding genes at the transcriptional level in higher eukaryotes.

In vivo (1-4 and refs. therein) and in vitro (2, 5, 6 and refs. therein) studies have shown that the SV40 early promoter region consists of two overlapping promoters, controlling initiation of transcription at the early-early (EE) and late-early (LE) start sites (see Figure 1).

---

* Present address: European Laboratory of Molecular Biology, Postfach 1022.40, 6900 HEIDELBERG, Germany
**Present address: Dept. of Molecular Biology and Genetics, University of Guelph, GUELPH, Ontario, Canada

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

Early in viral infection transcription of the early genes (large and small T antigens) initiates chiefly at the early-early start sites (EES, Figure 1), whereas late in infection early transcription starts further upstream at the late-early start sites (LES, Figure 1; see 2, 7-10 and refs. therein). The EE promoter (EEP) comprises (i) the TATA box sequence, which ensures accurate and efficient initiation of transcription from EES, (ii) an upstream element, here referred to as the 21-bp repeat region, and (iii) the enhancer. Both the upstream element and the enhancer are indispensable for efficient early transcription. The LE promoter (LEP) shares both the 21-bp repeat region and the enhancer element with the EEP.

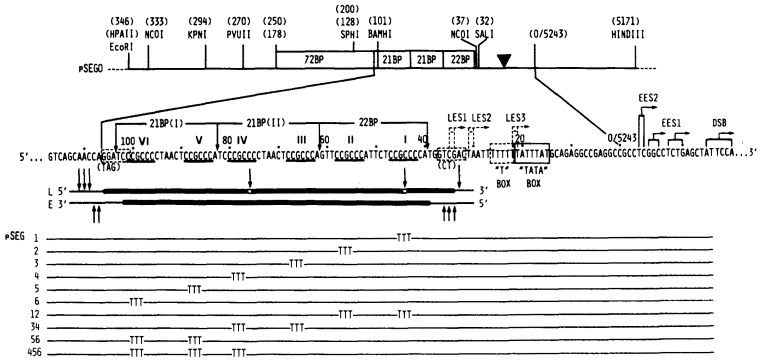


Fig. 1. General organization and sequence of the SV40 early promoter region and of its mutant derivatives in the 21-bp repeat region. The top diagram represents the organization of the SV40 early promoter region from HpaII (346) to HindIII (5171) in recombinant pSEGO (10)). Some SV40 DNA restriction sites, the 72-bp sequence, 21-bp repeat region and the TATA box are indicated. Unless otherwise stated, all recombinants described in this review contain one copy of the 72 bp-sequence only. The second line shows the sequence of pSEGO (non-coding early strand) between coordinates 5224 and 117. The SV40 wild-type sequences changed during the creation of the BamHI (coordinate 101) and SalI (coordinate 32) sites (10) are depicted in parentheses below the main sequence. The position of the DSB (downstream bands, Ref. 2), early-early startsites EES1 and EES2 (arrows, see Ref. 4) and late-early startsites LES (LES1, LES2 and LES3, dashed arrows) are indicated. The six GC-motifs 5'-CCGCCC-3' are underlined and identified by the roman numerals I-VI. The two directly repeated 21-bp sequences (I and II) and the 22-bp sequence are delimited by vertical arrows pointing downwards. The segments of the two DNA strands which are protected by a nuclear extract against DNase I digestion (see Figures 3 and 4, and text) are indicated by thick bars below the sequence, with arrows pointing to DNase I hypersensitive sites. The sequence mutated in the pSEG mutant series (Table I) is indicated below the pSEGO sequence. All nucleotide numbers follows the BBB system (9). This figure was taken from Ref. 10.

Although little is known about the sequences within and around the TATA box element and the cap site that are required for efficient early transcription, the upstream sequence and the enhancer element have been extensively characterized. A systematic analysis of these promoter elements revealed that six GC-motifs I-VI (Figure 1) are involved in the function of the 21-bp repeat region *in vivo* (10-12 and refs. therein) and that within the enhancer element there are at least two domains, A and B (Figures 5 and 6; see ref. 13) indispensable for efficient early transcription. The transcriptional activity of each enhancer domain is due to the presence of several specific sequence motifs which also occur in various assortments in other viral and cellular enhancers.

In vitro transcription studies have shown that different *trans*-acting factors bind specifically to these *cis*-acting regulatory sequences of the SV40 early promoter. The TATA box factor binds to the TATA box to form a stable complex; this is a prerequisite for specific initiation of transcription (14-15; Tamura et al., in preparation). The cellular transcription factor Sp1 attaches specifically to the six GC-motifs I-VI of the 21-bp repeat region and stimulates transcription from the EES and LES (10, 16-18 and refs. therein). Different *trans*-acting factors bind to the multiple sequence motifs within the enhancer and stimulate transcription from both the SV40 early and heterologous promoter elements (19-24 and refs therein). DNase I footprinting and dimethylsulfate methylation protection experiments have shown that binding *in vitro* of the Sp1 factor (10, 16-18) and enhancer factor(s) (23, 24) to the 21-bp repeat and the enhancer regions, respectively, is severely impaired by mutations known to be detrimental to the *in vivo* function of these promoter elements. In addition, a stereospecific alignment of the various promoter elements is required for efficient stimulation of transcription, which strongly suggests that protein-protein interactions are involved in the activation of transcription initiation from this RNA polymerase class B promoter (4).

The TATA box region

The SV40 early promoter contains, as part of an AT-rich sequence at a distance of 20-25 nucleotides upstream of the major EES1, the TATA box element of the early transcription unit (5'-TATTTAT-3', SV40 position 21-15, Figures 1 and 2), which directs the transcriptional machinery to

initiate at the EES1 (1, 2, 25, 26). Initially, from studies using deletion mutants, it has been concluded that this TATA sequence is dispensable and not required for efficient early transcription (1, 27-29). However, a double point mutation in the SV40 TATA box has been shown to decrease the overall transcription from the EES by approximately 2.5-fold (2). In this case initiation of transcription from EES1 is decreased by more than 95%, but there is compensatory increase in initiation of transcription from the LES and sites located downstream from the EES1 (DSB, in ref. 2; see Figure 2). The existence of an additional functional TATA box-like sequence, present in the cluster of T residues located immediately upstream from the major TATA-box (T box in Figures 1 and 2), is suggested by the disappearance of the minor start sites EES2 (Figures 1 and 2 and ref. 4) when the series of T residues from position 22 to 26 (T box) is mutated (M.Pauly et al., in preparation), whereas they are not affected by a mutation in the major TATA box sequence (see ref. 2 and Figure 2). Furthermore, efficient transcription initiation at EES1 and

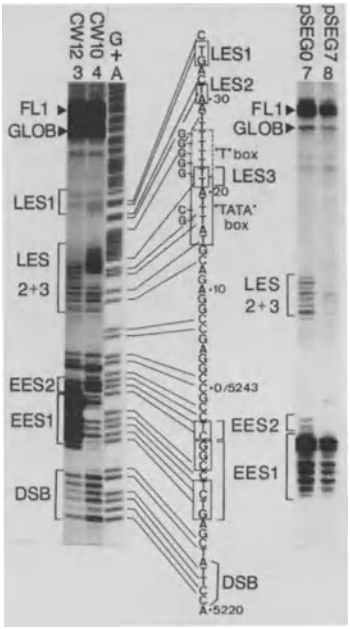


Fig. 2. Effect of mutations within the TATA-box and T-box, respectively, on transcription from the SV40 early promoter in vivo. Recombinant plasmids were transfected into HeLa cells and after transient expression analysed for their transcriptional activity by quantitative S1 nuclease mapping as described (2, 10). In the construction CW10 (TATA-box mutant) the two T residues at position 17 and 18 have been mutated to G and C residues, respectively, (2) as indicated. Lanes 3 and 4 show the transcriptional activity of CW10 with respect to the wild type promoter present in CW12 (2). A+G sequence ladder is shown. In the recombinant pSEG7 the cluster of T residues (position 22-26, T-box) has been mutated to a series of G residues (M. Pauly et al., in preparation), as indicated, and analysed for its effect on transcription (lane 8). The transcriptional activity of the corresponding wild type recombinant pSEG0 (10) is shown in lane 7. DSB, EES1 and EES2, LES1-3, "TATA"-box and "T"-box are as in Fig. 1. For GLOB and FL1 see (2) and (10). This figure was modified from Ref. 2.

EES2 requires a stereospecific alignment of the upstream sequence element (the 21-bp repeat region, see below) and the TATA box region (4).

In conclusion, the TATA box sequence(s) of the SV40 early transcription unit play(s) the same role as the TATA box elements of other class B promoter regions, in that it selects the precise site of transcriptional initiation and contributes to the overall efficiency of transcription from the early promoter. However, the SV40 early promoter appears to be different from most class B promoters in that it contains several TATA box substitute elements which can compensate to a large extent for a mutation in the main TATA box.

The 21-bp repeat region.

The initial observation that progressive deletions extending upstream from the EES dramatically reduce T-antigen expression of the SV40 early transcription unit (1) identified the GC-rich region upstream to the TATA sequence as an essential element of the SV40 early promoter. This region comprises two tandemly repeated 21-bp sequences and a related 22-bp sequence, with each of these three sequences containing two GC-rich motifs, 5'-CCGCCC-3'. The 21-bp repeat region can stimulate transcription when present in the reverse orientation (10-12), indicating that it constitutes a bidirectional promoter element for both the EE and LE promoters. An analysis by means of deletion and multiple random point mutations in the 21-bp repeat region demonstrated that the GC-rich motifs were the important sequence elements in this region (11, 12). To further analyze the contribution of each GC-motif in EE and LE promoter function *in vivo* and *in vitro* a systematic site-directed mutagenesis of all GC-motifs, either individually or in combination, has been performed (10; Figure 1). The transcriptional activity of the various recombinants was determined by measuring the amount of RNA synthesized after transient expression in Hela cells. This study reveals that GC-motif I (pSEG1, Figure 1 and Table 1) represents a key component of the EEP. Although less crucial, GC-motifs II and III (pSEG2, pSEG3; Figure 1 and Table 1) also play an important role in efficient initiation from the EES, whereas individual mutations in GC-motifs IV, V and VI (pSEG4-pSEG6, Figure 1 and Table 1) appear to be less detrimental to transcription. This result is also seen using recombinants where the two GC-motifs of a given 21-bp repeated sequence are mutated simultaneously, since EE transcription is

almost abolished in pSEG12, whereas it is decreased to a lower extent in pSEG34 and even more so in pSEG56 (Figure 1 and Table 1). Nevertheless, the distal GC-motifs IV, V and VI are also important elements of the EEP, since their simultaneous mutation (pSEG456, Figure 1 and Table 1) results in a 10-fold decrease of initiation from the EES. The reason why GC-motif I plays such a crucial role for promoter activity is unknown, but may be related to the fact that it resides immediately adjacent to the TATA box, if one assumes for instance that the transcription factor(s) (see below) attached to the 21-bp repeat region interact(s) with that bound to the TATA box and that this interaction is mediated in some way by the factor bound at GC-motif I. It is noteworthy that GC-motifs I and II (pSEG1, pSEG2; Figure 1 and Table 1) are not required for efficient RNA initiation from the LEP, whereas GC-motifs III-VI (pSEG3- pSEG6, Figure 1 and Table 1) appear to be weak elements of this promoter (10, 12).

Table 1. Effect of mutations within the GC-motifs on RNA initiated at the early-early (EES) and late-early (LES) startsites.

RECOMBINANTS	MUTATED GC-RICH MOTIF	RELATIVE TRANSCRIPTION	
		EES	LES
pSEG 0		100	100
1	I	4,8 [1,9-7,3]	90 [81-100]
2	II	20 [13-27]	156[140-177]
3	III	15 [11-20]	80 [64-97]
4	IV	42 [33-54]	93 [88-101]
5	V	33 [21-43]	91 [87-94]
6	VI	56 [49-62]	88 [72-109]
12	I + II	1,2 [0,5-1,8]	78 [53-99]
34	III + IV	10 [6-14]	59 [53-72]
56	V + VI	27 [20-30]	51 [41-65]
456	IV + V + VI	10 [8-14]	39 [35-46]

The relative amount of RNA initiated at early-early (EES) and late-early (LES) sites was estimated by scanning autoradiograms from five different quantitative S1 nuclease mapping experiments and correction for transcription from the reference rabbit β-globin gene, taking pSEG0 values as 100% (10). For each recombinant, the average value and the two extreme values (in brackets) are given. This table was taken from Ref. 10.

The effect of the various GC-motif mutations on in vivo transcription can be reproduced to a large extent in vitro in a whole cell or nuclear extract (6, 10) with the overall pattern of transcription being more specific in the nuclear extract (10). Furthermore, mutations affec-

ting transcription in vivo and in vitro interfere with binding of a specific cellular transcription factor Sp1, which is required for efficient transcription from the SV40 early promoter in vitro (10, 16-18, 30). To examine the interaction between the GC-motifs and their cognate trans-acting factor(s), footprinting experiments have been performed using both nuclear extracts of Hela cells (10) and nearly homogeneous Sp1 preparations (18). As shown in Figure 3 (panel A, compare -Extract with +Extract), the early-coding strand of pSEG0 is protected from DNase I digestion in the nuclear extract throughout all six GC- motifs, with GC-motif I being less efficiently protected than the other five. A similar protection is seen on the late-coding strand (Figure 4, panel A). On both early- and late-coding strands (e.g. Figure 4, panel B, nucleotide position 44, compare lanes 1 and 2 of pSEG0), the protection over GC-motif I is increased as more extract is used in the footprinting reactions. The footprint of pSEG0 is also characterized by the presence of hypersensitive sites at several discrete positions flanking the 21-bp region and downstream of the TATA box (see arrows at sites 108, 35, 2 and 5234 on the early strand in Fig. 3A, and sites 33 and 110 on the late strand in Fig. 4A, see also Figure 1).

Individual point mutations in the various GC-motifs I-VI lift off the in vitro footprint principally over that particular motif, both on the early- and late-coding strand (Figures 3A and 4A, respectively), suggesting very little cooperativity in protein-binding among the various GC-motifs. Both DNase I footprinting and DMS methylation protection experiments using purified Sp1 transcription factor suggest that each of the six tandemly repeated GC-motifs of the 21-bp repeat region can interact individually and non-cooperatively with a protomer of Sp1 factor in vitro (18). GC-motif IV is not completely protected from DNase I digestion, particularly on the late coding strand (nucleotide 76, pSEG0 in Figures 4A and B), perhaps because of steric hindrance to protein binding in this region, since of all six GC-motifs I-VI, the distance between GC-motif IV and GC-motif V is the smallest, being only 3-bp (Figure 1). When the effect of point mutations in the neighbouring GC-motifs III and V on factor-binding to GC-motif IV was investigated (Figure 4B), the mutations in GC-motif V increased protection of GC-motif IV (pSEG5 in Figure 4B, see also pSEG56 in Figure 4A), a result confirmed by using purified Sp1 transcription factor (18). Efficient binding of Sp1 factor to GC-motif IV

in vitro, therefore, appears to occur only when binding to GC-motif V is prevented or in the presence of a high concentration of Sp1 factor (I. Davidson, unpublished observation).

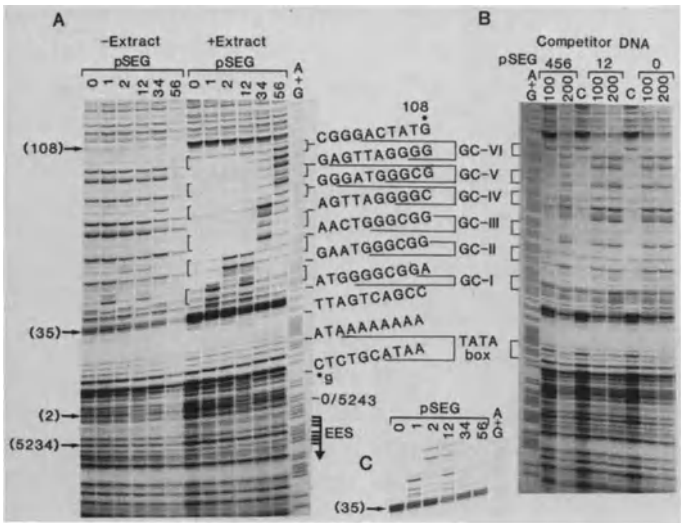


Fig. 3. Effect of mutations within the GC-motifs on the binding of proteins to the early strand of the 21-bp repeat in nuclear extracts. The early strand of the templates were end-labeled by digestion of the pSEG series with HindIII, phosphorylation with polynucleotide kinase, and subsequent digestion with EcoRI (Fig. 1). These fragments were then purified on 6% polyacrylamide gels and electroeluted. All footprinting reactions were performed at 30°C using 6 μl of total nuclear extract in a 12 μl reaction volume, as described (10, 23), using DNase I at an approximate final concentration of 5 μg/ml. The reactions were stopped with SDS, phenol/$CHCl_3$ extraction, and the samples analysed on 8% polyacrylamide salt gradient gels.

Panel A. DNase I footprints on pSEG0 and pSEG mutants. DNA fragments (as indicated at the top of the lanes were digested after incubation in the absence (-Extract) or presence (+Extract) of a nuclear extract prepared as previously described (21). Samples without extract had the same ionic strength buffer as those with extract (the -Extract pSEG56 which was overdigested in this experiment, normally gave a pattern similar to that of the other fragments). Arrows indicate the position of sites made hypersensitive to DNase I in the presence of extract. The position of the early-early startsites (EES) and the sequence (early-coding strand) of the 21-bp repeat region are indicated. Square brackets indicate the position of the six GC-motifs.

Panel B. The effect of mutations present in pSEG12 and pSEG456 on the ability of the 21-bp region to compete for the DNase I footprint. Competition footprinting reactions were carried out as described (23). The nuclear extract was pre-incubated with 25 ng of linearized pBR322 DNA, and either no competitor fragment (C) or 100 or 200 ng of purified BamHI to SalI fragment (see Figure 1) from the pSEG recombinant indicated

at the top of the figure. Following addition of the end-labeled template (the early-coding strand of pSEG0 labeled at the HindIII site), standard footprinting reactions were performed.
Panel C. Effect of mutations on the hypersensitivity to DNase I of positions flanking the 3'-side of the 21-bp repeat region. Shown here is a lighter exposure of the region around position 35 in panel A, +Extract, to more clearly demonstrate the effect of GC-motif mutations on DNase I cleavage at this site.
In panels A, B and C, A+G lane is a sequencing track of pSEG0 template (early-coding strand). This figure was taken from Ref. 10.

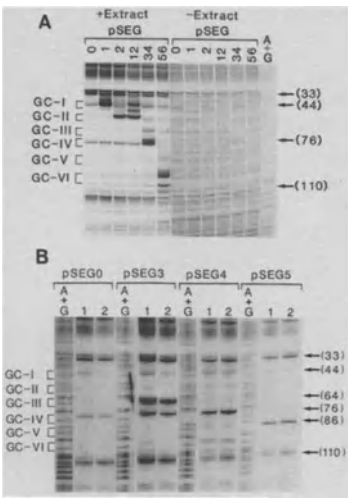


Fig. 4. The effect of mutations within the GC-motifs on the binding of proteins to the late strand of the 21-bp region in nuclear extracts. The pSEG templates (as indicated at the top of the lanes) were end-labeled at the EcoRI site (Fig. 1), digested with HindIII, and purified as decribed in legend to Fig. 3.
Panel A. DNase I footprint on pSEG0 and pSEG mutants. All footprinting reactions were carried out at 30°C, in the presence (+Extract) or absence (-Extract) of nuclear extract (4 μl), and the samples analyzed as described in legend to Fig. 3A. The arrows indicate the position of sites which become DNase I hypersensitive upon incubation with the nuclear extract. An A+G Maxam and Gilbert sequencing track of the pSEG0 fragment (late-coding strand) is shown.
Panel B. Footprinting reactions were carried out as in panel A, with the pSEG templates as indicated, using either 4 μl (lane 1 in each of the four series) or 6 μl (lane 2 in each series) of nuclear extract. An A+G sequencing track of each pSEG template (late-coding strand) is shown, and nucleotide positions particularly susceptible to DNase I are indicated in parentheses. This figure was taken from Ref. 10.

In vitro competition experiments such as the one shown in Figure 3B (the functional GC-motifs I-III and III-VI present in pSEG456 and pSEG12, respectively, as well as the wild-type 21-bp repeat region of pSEG0 can efficiently compete for binding of Sp1 factor to all GC-motifs I-VI) support the notion that the proteins which bind to the six tandemly repeated GC-motifs I-VI are identical. Taken together with the observation that each GC-motif appears to behave as an individual binding site, these results suggest that as many as six molecules of Sp1 can bind simultaneously to the six GC-motifs I-VI of the 21-bp repeat region.

The SV40 enhancer

An additional cis-acting element, the enhancer, is located further upstream from the 21-bp repeat region, and stimulates transcription from the early promoter by approximately three orders of magnitude (13). It was initially identified using deletion mutants (1, 27), and became very popular as the "72-bp repeat", since it occurs in various SV40 isolates as a 72-bp tandemly repeated sequence. These "far located" cis-acting promoter elements have been termed enhancers, since they dramatically stimulate transcription from homologous or heterologous RNA polymerase class B promoters, in an orientation-independent manner and over long distances (31, 32). However the activity of the SV40 enhancer appears to decrease rapidly when it is moved away from the elements of the SV40 early promoter (33) or from some heterologous promoter elements such as the conalbumin or adenovirus-2 major late promoters (31, 33-35). Although enhancers were first identified in SV40, many other viral and several cellular enhancers have since been described (for reviews see 36-41). In contrast to the SV40 enhancer which is active in a variety of different cell lines, some viral enhancers exhibit a pronounced host cell preference (42-44), and stricter cell lineage specificity is associated with cellular enhancers, most notably immunoglobulin gene enhancers (45-49 and refs. therein).

The sequence requirement within the SV40 enhancer element has been determined at the nucleotide level by a systematic study of a series of deletion and point mutations (13). Various recombinants have been constructed employing in vitro site-directed mutagenesis (50) and assayed for their transcriptional activity in vivo after transient expression in Hela cells and quantitative S1 nuclease analysis (13). This study demonstrates that, in addition to the 72-bp sequence, sequences located further upstream are essential for enhancer activity, and delineates the DNA sequences required for the enhancer function in HeLa cells to a region of approximately 100-bp (position 185-275 for an enhancer containing one copy of the 72-bp sequence, Figures 1, 5 and 9). Two domains A and B, which by themselves activate transcription only at a very low level, have been identified within this region (Figures 5 and 6, see ref. 13). The existence of a third domain C (position 298-347) located upstream from the KpmI site, whose activity is barely apparent when domains A and B are

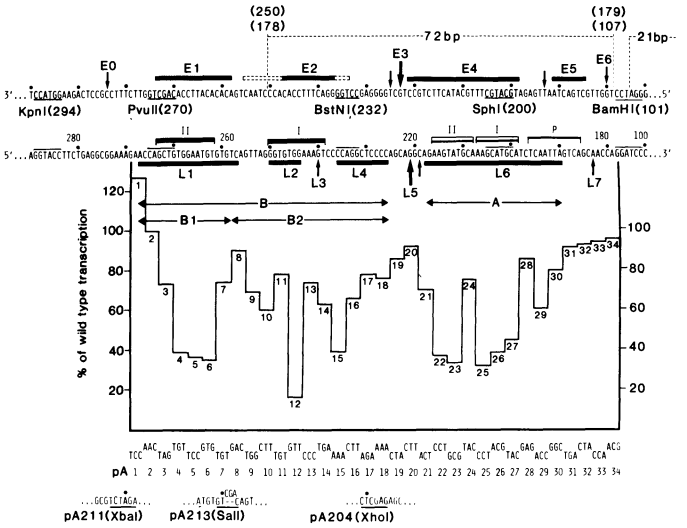
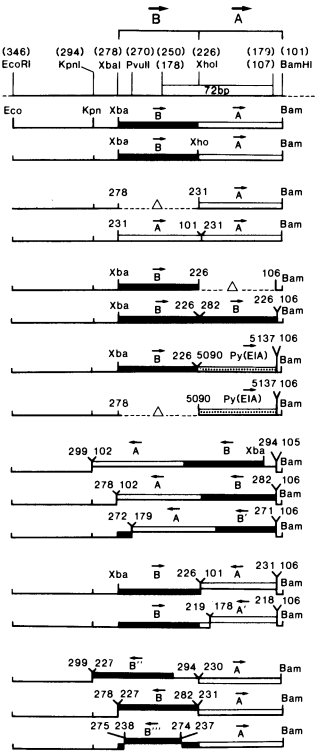


Fig. 5. Comparison of the location of regions of DNase I protection and hypersensitivity observed in vitro with the results of the in vivo activity of enhancer point mutants. The early (upper) and late (lower) coding strand of the SV40 72-bp and 5'-flanking sequence of pA0 (13) are shown (together with some relevant restriction sites) at the top of the figure (the wild type sequence at positions 101 to 103 has been mutated to generate the BamHI site in pA0, as described (13, 50). The regions of DNase I protection identified in Figure 7, and confirmed by competition experiments (not shown) and point mutation analysis (Fig. 8), are shown by a solid line. The broken lines flanking region E2 indicate those residues which based on "competition" footprinting studies (not shown) are not clearly protected, but based on point mutation footprinting (Figure 8) appear to be included in this region. Nucleotides that have sensitivity to DNase I that is increased strongly (large arrows) and moderately (small arrows) are indicated. The locations of the SV40 GT-motifs I and II (solid lines, I and II), the Sph-motifs I and II (double lines, I and II) and P-motif (single line, P) are indicated (see text and ref. 13). Below the enhancer sequence is shown the profile of the in vivo effect of point mutations on transcriptional efficiency of the enhancer as measured by quantitative S1 nuclease mapping after transient expression in HeLa cells (13) (expressed relative to pA0), and below the profile are shown the nucleotide sequences present in each mutated template of the pA series (pA1 to pA34). The location of XbaI, SalI and XhoI restriction sites introduced by site-directed mutagenesis (mutants pA231, pA213 and pA204, see Ref. 13) and used for the experiments shown in Figure 6, are indicated at the bottom of the figure. This figure was taken from (23).

intact (Figure 6, pA211 and pA260), is suggested because of the ability of this DNA segment to rescue the activity of a truncated enhancer devoid of most of the sequences of enhancer domain B (13). That this segment contains some potential enhancer activity is further supported by the observations of Weber et al. (51) and Swimmer and Shenk (52), who reported that duplication of sequences from position 298 to position 376 or 357, respectively, generates a functional enhancer element.

Both domains A and B have genuine enhancer properties, since mutations in either domain are equally detrimental to enhancer activity when the enhancer is either in close apposition to the 21-bp repeat region or moved 600-bp away from the remaining part of the early promoter (13). Since isolated domains A and B exhibit very little enhancer activity on their own (pA223, pA224, pA233, pA234 in Figure 6), the 400-fold stimulation of transcription (13) brought about by their juxtaposition must involve some efficient synergistic mechanism. Surprisingly, this synergy does not appear to depend critically on the relative orientation of either domain. Domain A or B can function bidirectionally (see pA302, pA203, pA310 and pA305, pA303, pA202, pA311, respectively, in Figure 6), like the entire enhancer (pA301, pA306, pA201, pA308, in Figure 6), and even the distance between the two domains can be varied to some extent without severely affecting enhancer function (13). Thus, the two enhancer domains appear to behave as individual units, even though they functionally cooperate to generate enhancer activity. However, maximal enhancer activity, which is achieved when the enhancer is in close apposition to the remaining part of the SV40 early promoter (33), requires a stereospecific alignment between some element(s) of domain A and the 21-bp repeat region (see below and ref. 4).

Enhancer activity can be generated by the association of domains A and B (pA211, pA212, pA260, pA261, in Figure 6), by duplication of either domains A or B (pA235, pA225 and pA237, pA227, respectively, in Figure 6), or even by replacing SV40 domain A by the polyoma virus enhancer domain A (pA411, pA401, in Figure 6). Generation of enhancer activity by duplication of a non-functional enhancer domain has also been reported by others for both the SV40 and polyoma virus enhancers (51-55). Once the basic SV40 enhancer activity has been achieved by associating domains A and B, increasing the number of 72-bp sequences within the enhancer or



	Transcription relative to pA0(EES)			
	with Eco-Xba region		without Eco-Xba region	
	percent		percent	
pA211	92		pA260	46
pA212	102		pA261	38
pA233	2.3		pA223	0.4
pA235	5.2		pA225	1.6
pA234	3.3		pA224	0.8
pA237	15.5		pA227	8
pA411	51		pA401	26
pA410	3.0		pA400	0.6
pA301	79			
pA306	62		pA308	45
pA201	62			
pA302	42		pA310	18
pA203	53			
pA305	41			
pA303	44		pA311	8.5
pA202	33			

Fig. 6. Effect of rearrangement of enhancer domains A and B. The SV40 enhancer region of pA0 from EcoRI (position 346) to BamHI (position 101) is depicted at the top of the figure. Enhancer domains A and B emcompass sequences from BamHI to XhoI (position 101-226, open box) and from XHoI to XbaI (position 226-278, filled box), respectively (see Fig. 5). The arrows indicate the relative orientation of the two domains within the wild type and rearranged enhancer. The various recombinants were constructed using synthetic linkers and adaptors, taking advantage of restriction sites engineered in the SV40 enhancer region : XhoI site (position 226, pA204); XbaI site (position 278, pA211); combination of both sites (pA212) (Fig. 5). Two classes of recombinants with or without the sequences further upstream from the XbaI site up to the EcoRI site (position 278-346) were constructed (with Eco-Xba and without Eco-Xba region, respectively) (for a more detailed description of the various constructions, see Ref. 13). Transcription initiated at the EES of each recombinant was determined by quantitative S1 nuclease analysis and expressed relative to pA0, taken as 100% (13). In view of the requirement for stereospecific alignment between the 21 bp repeat region and the enhancer (4, and see below), the relative enhancer activities of the various constructions must be considered as minimal values. This figure was taken from (13).

polymerization of the entire enhancer (position 179-272) results in a further linear increase of enhancer activity (13). This observation suggested that the underlying mechanism is different from that responsible for the "burst" of enhancer activity generated by juxtaposing domains A and B.

A systematic point mutation scanning analysis (Figure 5) reveals mutiple sequence motifs required for enhancer function. One of these, defined by the scanning mutants pA21-pA30 (position 219-190, Figure 5) in enhancer domain A, contains a tandem repetition of the sequence 5'-AAG(C/T)ATGCA-3' at positions 207-199 and 216-208 (Figure 5), termed the Sph- motifs I and II, since the 3' repeated sequence harbors the SphI site. A second region important for enhancer function, which is defined by mutants pA3-pA18 (domain B, position 273-226, Figure 5) consists of sequences located both within and upstream from the 72-bp sequence. The repeated sequence 5'-G(C/G)TGTGG-3', which belongs to a longer repeated sequence 5'-G(C/G)TGTGGAA(A/T)GT-3' (called hereafter the GT-motifs I and II, Figure 5), represents an essential motif of this enhancer domain (see mutants pA3-pA6, pA10 and pA12). GT-motif I is followed by a repetition of the motif 5'-TCCCCAG-3' (TC-motif), whose equivalent bases are 9 nucleotides apart, but only mutations within the more upstream (positions 239 - 233, Figure 5) of these two motifs, appear to efficiently decrease transcriptional activity (pA15). Interestingly, the mutation present in pA1 generates an additional TC-motif, which is apparently accompanied by an increase in transcription. The scanning mutant pA29 and pA30 (Figure 5) identify an additional sequence important for enhancer function, which, because of its homology to part of the polyoma virus enhancer, has been termed the P-motif (13).

GT-motif I partially overlaps with the so-called 'core sequence' GTGG(A/T)(A/T)(A/T)G which, based on sequence homology between various enhancer elements and on the result of a random point mutagenesis, has previously been identified as an important sequence of the SV40 enhancer (56). Both GT-motifs I and II are clearly required for transcriptional activity of enhancer domain B (13). The fact that the sequence 5'-GGTGTGG-3' present in the two GT-motifs (Figure 5) is also repeated at a distance of 20-23 nucleotides in other enhancers [bovine papillomavirus (57, 58), Adenovirus E1A (59), immunoglobulin κ-light chain (60)] further supports this conclusion. While the distance between the centers of the

two GT-motifs (23-bp) corresponds to approximately two DNA helical turns, there appears to be no stringent distance requirement between them, since an insertion of 5-bp between them does not abolish the transcriptional activity of domain B (13). Furthermore, deleting GT-motif II can be compensated for by juxtaposing the potential enhancer domain C to GT-motif I (13). Thus domain B may be composed of two subdomains B1 and B2, both of which are required to generate an active domain B, but which can function, at least to some extent, independently of one another.

It has been suggested that two clusters of alternating purine and pyrimidine residues (positions 265-258 and 205-198, Figure 5) are important sequence features for SV40 enhancer function, because they have the potential to form Z-DNA structures (for refs. see 61, 62). The scanning mutation analysis (13) indicates that these alternating purine-pyrimidine motifs do in fact belong to important sequence elements of the SV40 enhancer (see mutants pA5, pA6, pA7, pA25, pA26 and pA27, Figure 5). However, it has also been shown that transition mutations in either motif, as well as a combination of transition mutations in both motifs are as detrimental to enhancer function as the corresponding transversion mutations (13). Furthermore, when the sequence at positions 205-198 is mutated to either alternating GC or GT residues (13), the enhancer activity is also decreased. These results suggest that the contribution to enhancer activity of the two motifs containing the alternating purine-pyrimidine residues cannot be ascribed simply to their potential to form Z-DNA structures.

In conclusion, it appears that the SV40 enhancer is made up of multiple sequence motifs which are contained in two distinct enhancer domains A and B. Individual point mutations in important sequence motifs reduce the transcriptional activity of the enhancer by at most 8-fold (pA12, Figure 5), which still represents a 50-fold enhancement of transcription relative to an enhancerless promoter (13). To completely abolish enhancer activity requires multiple mutations situated at key positions within the enhancer (13). This supports the idea that the SV40 enhancer represents a large DNA segment composed of multiple important sequence motifs, the association of which is indispensable for the generation of full enhancer activity.

Results of *in vitro* (19-22, 63) and *in vivo* (64, 65) studies have indicated that specific *trans*-acting factors are involved in enhancer

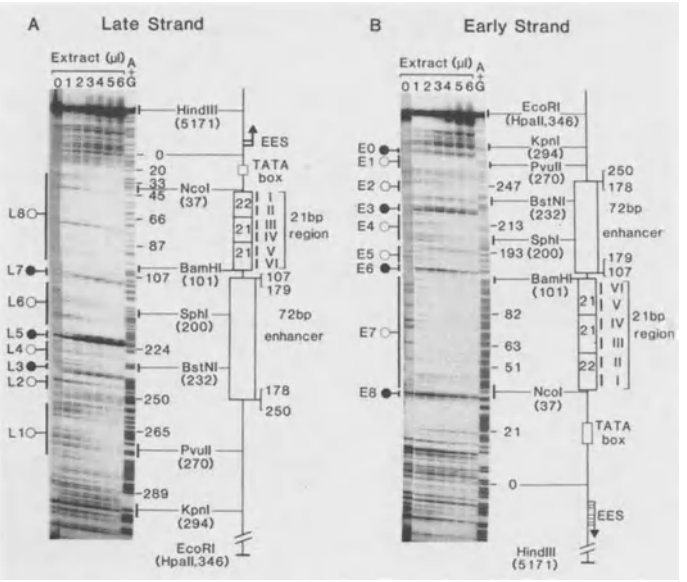


Fig. 7. DNase I footprinting over the SV40 promoter region in nuclear extracts. The HindIII-EcoRI promoter template was derived from the wild-type recombinant pAO which contained only one 72-bp sequence (13, and Figs. 1, 5 and 9). The results obtained with the L- and E-coding strands are shown in panels A and B, respectively. DNase I footprinting was carried out as described in (23), with the amount (in microliters) of HeLa cell nuclear extract indicated above each panel. A+G sequence ladders of the labeled template were run in parallel. The positions (in parentheses) of the enhancer sequence, the 21-bp repeat region with its six GC-rich motifs (I to VI), the TATA box, the early-early start sites (EES), and some key restriction enzyme sites are indicated. To the left of each autoradiogram is indicated the regions that have increased (•) or decreased (○) sensitivity to DNase I after incubation in the nuclear extract. This figure was taken from Ref. 23.

function. In vitro DNase I footprinting experiments strongly suggest that these trans-acting factors are exerting their function by binding to the various enhancer motifs identified by the point mutation scanning analysis (23). A typical in vitro DNase I footprinting experiment on either the late- or early-coding strand is shown in Figure 7, where the SV40 early promoter region (position 5171-346) has been incubated with increasing amounts of nuclear extract (1-6 µl) and subsequently treated with DNase I. As summarized in Figure 5, regions of increased DNase I sensitivity (L3, L5, L7 and E0, E3, E6) reside in positions at which point

mutations have little effect in vivo on the transcriptional activity of the enhancer. On the other hand, several sequences become protected : L1, L2, L4, L6 and E1, E2, E4, E5 (Figure 7) on the late- and early-coding strand, respectively. The amount of extract required to protect regions L4 and E2 was higher than that necessary to protect the other regions of the enhancer and to generate the hypersensitive sites (23). As illustrated in Figure 5, the protected sequences correspond to enhancer motifs which have been identified in vivo as being important for enhancer activity (13). In addition, in vitro competition experiments performed by

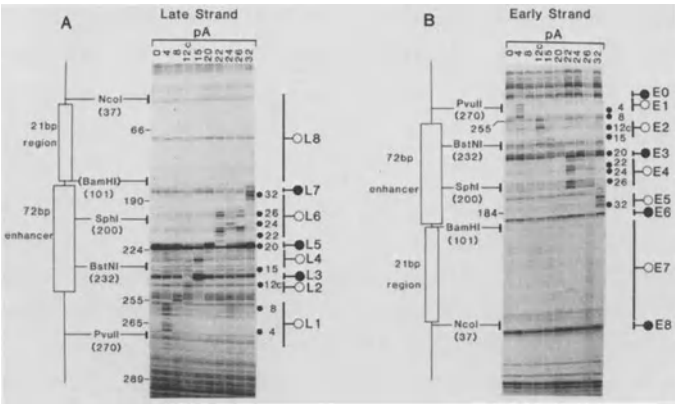


Fig. 8. DNase I footprinting over SV40 promoter regions containing point-mutated enhancers. Footprinting reactions were performed as described in (23) and Legend to Fig. 7, using 6 µl of nuclear extract per reaction. The first lane in each panel is the pA0 template, and subsequent lanes are the mutated pA templates, as indicated at the top of each lane (see Fig. 5 for the nucleotide changes present in these mutants). pA12c contains the mutation in the same position as pA12 except that the sequence was changed to ACC (see Ref. 13). The position of each mutation is shown to the right of each panel as a small solid circle, and the location of the regions of DNase I protection and hypersensitivity are indicated as in Fig. 7. In control experiments all mutant DNAs were digested with DNase I in the absence of nuclear extract to check that their pattern of digestion is similar to that seen for pA0 (see Fig. 7 and Ref. 23). Only mutant pA24 showed an additional cut site at the position of the mutation on the late strand of its naked DNA, which results in a band which is also present in the presence of extract (see lane pA24 in panel A). The early strand of the pA26 template was overdigested in this particular experiment, but the deprotection in E4 caused by this mutation was identical in other experiments where the extent of DNase I digestion was lower. This figure was taken from Ref. 23.

Wildeman et al. (23) suggest, that the DNase I footprint of the enhancer is due to proteins which recognize specific sequences. This idea is further supported by the observation that mutations which are detrimental to enhancer function in vivo interfere with protein binding to the corresponding enhancer motif in vitro. Enhancer sequences carrying point mutations at key positions within the enhancer were subjected to DNase I footprinting analysis (23; Figure 8). Point mutations present in pA4 and pA12c lift off the in vitro footprint only over GT-motifs I and II, respectively. Mutations present in either pA22 or pA26 result in general deprotection of E4 and L6 (corresponding to the Sph-motifs I and II), but did not affect protection of the other regions. Although pA32 has very little effect on enhancer activity in vivo, there was deprotection of the residues at that position on both early- and late-coding strand (E5 and part of L6), corresponding to the P-motif (Figure 5). In marked contrast, the point mutations pA8, pA20 and pA24, which have little or no effect on enhancer function in vivo (13), result only in marginal alterations of the pattern of protection and hypersensitivity. None of the enhancer mutations affects the protection observed over the 21-bp repeat region (L8 and E7), nor the hypersensitive sites L7 and E6 which are generated by the 21-bp repeat region (23).

These results support the notion that the enhancer consists of two independent domains A and B, since none of the mutations located in domain A affect the footprint on domain B and vice versa. Similar argument can be put forward to divide domain B into two almost independent subdomains B1 and B2, corresponding to the motifs TC-II and GT-I, and GT-II, respectively (23). The fact that both the pA22 and pA26 mutations lift off the footprint in a very similar fashion suggests that there is either a single protein bound to the two Sph-motifs I and II or a highly cooperative binding of two proteins (see also ref. 24). It appears, therefore, that there are a minimum of four protein molecules bound to the enhancer : one protein each for GT-motif I, GT-motif II, the Sph-motifs, and the P-motif. Consistent with the ability of enhancer domains A and B to generate enhancer acitivity in vivo irrespective of their orientation (Figure 6) and, to some extent, of their spacing (13), is the observation that the binding of protein factors to the sequence motifs in domain A and B, is not affected when either domain A or B have been inverted or moved apart (23). This, taken together with the fact that

none of the mutations in domain A affects the footprint on domain B and vice versa, makes it very likely that domains A and B independently bind proteins in vitro.

It has been reported that late in viral infection a subpopulation of SV40 minichromosomes exhibits a nucleosomal gap over the enhancer region (66-69 and refs. therein). This gap is generated over enhancer sequences even when they are separated from the remaining part of the SV40 early promoter (68). The results of the in vitro footprinting analyses described above raise the possibility that it is the binding of specific protein factors to the enhancer motifs which prevents the formation of a nucleosome. Genomic footprinting of wild-type and mutated SV40 enhancer sequences in vivo, using techniques such as those of Church and Gilbert (70), should provide further insight into this problem.

The SV40 enhancer is active in HeLa and lymphoid B-cells, unlike the IgH enhancer, which is preferentially active in B-cells (45, 46, 71-72), suggesting that both HeLa and B-cells might contain similar protein factors which recognize the various SV40 enhancer motifs. This notion is further supported by the finding that the IgH enhancer can efficiently compete with the SV40 enhancer in vitro in HeLa (20) or lymphoid cell extracts (63) and in vivo in B-cells (65). However, experiments performed by Davidson et al. (24) indicate that it is not the same set of sequence motifs and proteins which are responsible for the activity of the SV40 enhancer in the two cell types. As summarized in Figure 9, BJA-B lymphoid B cell nuclear extracts contain protein factors which, similarly to HeLa cell extracts (see also Figure 5), protect both domains A and B of the SV40 enhancer against DNase I digestion; these protections coincide with regions of specific DNA-protein interactions as demonstrated by DMS methylation protection experiments (Figure 9). However, there exist marked differences in the pattern of protection obtained with BJA-B and HeLa cell extracts (Figure 9, and ref. 24), suggesting that some of the enhancer motifs are cell-specifically recognized. This is particularly evident for the proteins which bind to enhancer domain A. Using methylation protection experiments, the late strand G residues G205 and G214 which reside at identical positions within Sph-motifs I and II, respectively, were protected in HeLa cell extracts, while G210 was hypermethylated (Figure 9). This result is in accord with the finding that the Sph-motifs I and II constitute the important sequence motifs of enhancer

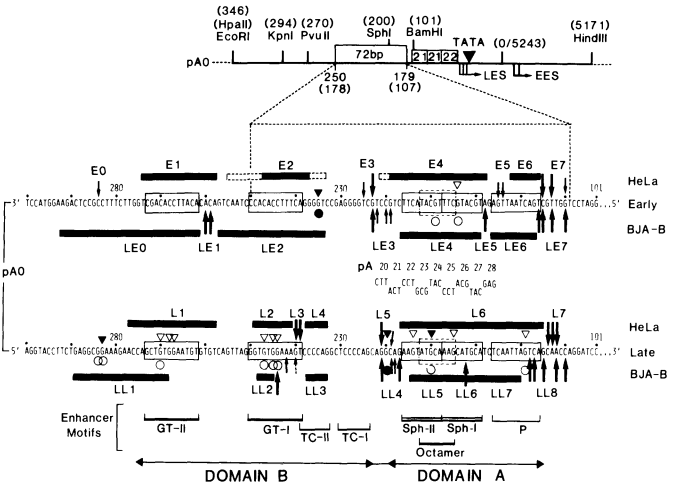


Fig. 9. Schematic representation of the results of DNase I footprinting and DMS methylation protection experiments over the SV40 enhancer in HeLa and lymphoid cell nuclear extracts. The SV40 early promoter region containing a single copy of the 72-bp sequence from plasmid pAO (13) is shown at the top of the figure together with some relevant restriction sites. The early and late mRNA coding strands of the 72-bp and 5'-flanking sequences are presented. The regions of protection from DNase I digestion in HeLa and lymphoid cell nuclear extracts are shown by the solid black lines above and below the sequence of each strand respectively. For the broken lines flanking regions E2 and E4, see legend to Fig. 5 and Ref. 23. The sites which were rendered hypersensitive to DNase I digestion by each extract are indicated by the large (strong) and small (weaker) arrows (dotted arrow, see Ref. 24). The location of the G residues which were protected from methylation by DMS in the HeLa (open triangles) and BJA-B (open circles) cell extracts are shown along with the hypermethylated G residues (filled triangles and circles for HeLa and BJA-B extracts, respectively). It should be noted that, using the HeLa cell extract, efficient protection of G189 within the P motif of pAO, was observed only when methylation was performed at 0°C (24). The nomeclature of each region is that used in Fig. 5 and Refs. 23 and 24. The locations of the enhancer sequence motifs which are essential for in vivo activity in HeLa cells (13) (GT-I, GT-II, TC-II, Sph-I, Sph-II and P) are shown along with those of the other enhancer motifs (TC-I and octamer) and of domains A and B. Motifs GT-I, GT-II, Sph-I, Sph-II and P are boxed with solid lines and the octameric motif with broken lines. The locations, and DNA sequence (late strand) of the mutations present in pA20-28 mutants of pAO (13) are shown between the early and late mRNA coding strands. Only those DNase I and DMS "footprints" which were consistently observed in several experiments using different preparations of HeLa or BJA-B cell extracts are indicated. This figure was taken from Ref. 24.

domain A (13, 23). Interestingly, the repetition of the two Sph-motifs generates the sequence 5'-ATGCAAAG-3', which is homologous to the "octameric" motif found in the IgH enhancer (13, 73 and refs. therein), the upstream elements of the promoters of the immunoglobulin genes (74-76) and the distal elements of the Xenopus U1 and U2 RNA gene promoters where they have been shown to exhibit enhancer properties (77, 78). This SV40 octameric motif appears to be the binding site of specific protein factor(s) present in the BJA-B cells, since the G residues G209 and G210 on the early- or late-coding strand, respectively, are protected in the BJA-B cell extract, whereas in the HeLa cell extract they are either not protected (G209) or hypermethylated (G210) (see Figure 9). Other G residues (G189, G204, G220) show the same pattern of protection or hypermethylation, respectively, in extracts of both cell types. Further evidence that it is the octameric motif rather than the Sph-motifs I and II which is important for enhancer activity in B-cells, stems from the observation that in B-cells only point mutations located within the octameric motif (pA23, pA24, pA25; Figures 5 and 9) reduce the activity of the enhancer in vivo, whereas mutations which affect exclusively the Sph-motifs (pA22, pA26, pA27; Figures 5 and 9) had little effect (24). In addition, exclusively those point mutations which reside in the octameric motifs lift off the in vitro DNase I footprint observed over this particular motifs in BJA-B cells. Thus, the differential pattern of protein binding observed in vitro with HeLa and BJA-B cell nuclear extracts accurately reflects the activity of the SV40 enhancer in vivo in HeLa and B cells, suggesting that the proteins interacting with the octameric motif act as transcriptional factors. Differences in the sequence motifs recognized by the HeLa and B-cell proteins are also evident in the region corresponding to domain B of the SV40 enhancer (Figure 9 and ref. 24); for example, the DNAse I region of protection denoted LE0 and LL1 (Figure 9) in BJA-B cell extracts extends further upstream than the corresponding region (E1 and L1) in HeLa cell extracts (Figures 5 and 9). The DMS methylation protection pattern is also different for the two cell extracts in this region (Figure 9 and ref. 24).

In summary, it appears that the multiple, sometimes overlapping, sequence motifs which constitute the SV40 enhancer can cell-specifically interact with trans-acting protein factors to create different specific nucleoprotein complexes which are presumably involved in the cell-speci-

fic generation of enhancer activity. Note that in addition the SV40 enhancer can be specifically "repressed" (79) which further increases the combinatorial possibilities of regulating its activity.

A stereospecific alignment of the various promoter elements is required for efficient initiation of transcription.

The mutational dissection in vivo and in vitro of the SV40 early promoter discussed above indicates that at least three different proteins (or sets of proteins) interact with the three different promoter elements to allow RNA polymerase B to initiate transcription. To investigate the possible cooperativity between the nucleoprotein complexes corresponding to these elements, the distances between the various promoter sequences have been altered by inserting short DNA sequences that generate odd and even multiples of half a helical DNA turn. Marked differences in the in vivo effects of these two types of insertions on transcriptional initiation have been observed (4), suggesting that protein-protein interactions between the trans-acting factors bound to the different promoter elements are involved in activation of transcription. Introduction of 10-bp or 21-bp, both near- integral multiples of the 10.5-bp per turn of B-DNA in vitro (80), between the 21-bp repeat region and the enhancer causes a 50% decrease in RNA initiated from all EES and LES (pSE10 and pSE21 mutants; Figures 10 and 11), whereas alterations of the distance between these two promoter elements by odd multiples of half a DNA turn result in dramatic 80-90% decreases in transcription initiated from all start sites (pSE5, pSE15 and pSE25 mutants; Figures 10 and 11). In marked contrast to the differential variations observed with insertions between the 21-bp repeat region and the TATA box (see below), the level of all RNA species initiating downstream from the enhancer (EES1, EES2, LES 2+3; see Figure 10) varies in a similar manner (pSE series in Figure 11). In addition, there is no rapid overall decrease of RNA synthesis as the length of the "even inserts" is increased and no decrease at all with increasing "odd inserts". These results indicate that efficient transcription from both the EES and and LES requires stereospecific alignments between some elements of the enhancer and of the 21-bp repeat region (4).

It has been demonstrated (10, 12) that insertion of DNA segments of increasing length between the 21-bp repeat and the TATA box regions leads to a rapid and drastic decrease in RNA initiating at the major early-

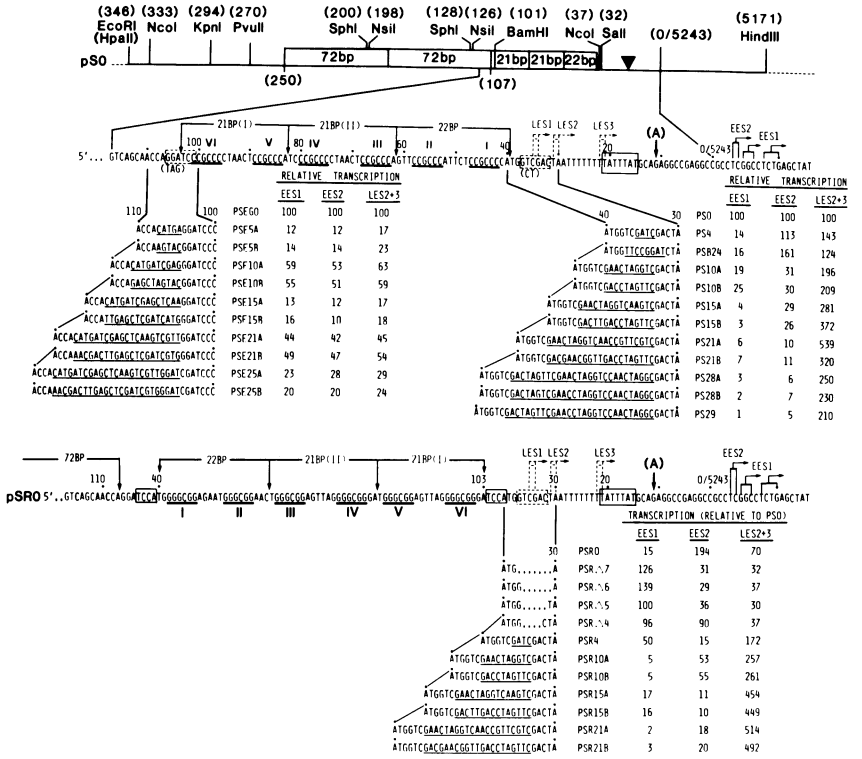


Fig. 10. General organization and sequences of the SV40 early promoter region and of the pS, pSE and pSR mutant series. The top diagram shows the organization of the SV40 early promoter region in recombinant pS0 (Ref. 4). Some key natural or engineered (BamHI, SalI) restriction sites are depicted as in Fig. 1. The triangle indicates the location of the TATA-box. The sequence of pS0 (non-coding early strand) between coordinates 5227 and 117 is shown. The TATA sequence is boxed. The position of the early-early startsites (EES1 and EES2, arrows) and late-early startsites (LES1, LES2 and LES3, dashed arrows), the directly repeated 21-bp sequences (I and II), the 22-bp sequence and the six GC motifs 5'-CCGCCC-3' are indicated as in Fig. 1. The sequences inserted (underlined) in the pSE and pS mutant series are represented on the left- and right-hand sides below the pS0 sequence, respectively. The pSE mutant series is derived from pSEG0 (10), which contains a single 72-bp sequence, whereas the pS mutant series is derived from pS0 which contains a complete 72-bp repeat (4). The bottom diagram shows the organization of the SV40 early promoter in recombinant pSR0 in which the 21 bp repeat region is in reverse orientation. pSR0 contains a complete 72 bp repeat. The sequences deleted or inserted (underlined) in the pSR mutant series are shown below the pSR0 sequence. The relative amount of RNA initiated at EES1, EES2 and LES2+3 after transfection of the various recombinants into HeLa cells was estimated by scanning autoradiograms from at least three different quantitative S1 nuclease mapping experiments using different plasmid preparations. After correction for transcription from the reference β-globin gene (4) the results are expressed relative to pS0 (for the pS and pSR series) or pSEG0 (for the pSE series), taken as 100%. This figure was taken from Ref. 4.

early start sites EES1. However, Takahashi et al. (4) showed (pS series in Figures 10 and 11) that 10- and 21-bp are consistently less detrimental to transcription than 4- and 15-bp inserts, respectively. Transcription initiating at the minor start sites EES2 (Figure 10) varied in exactly the opposite way (pS series, Figures 10 and 11), presumably because it is the T box (Figures 1 and 2) located immediately upstream from the major TATA box which acts as a substitute TATA sequence to

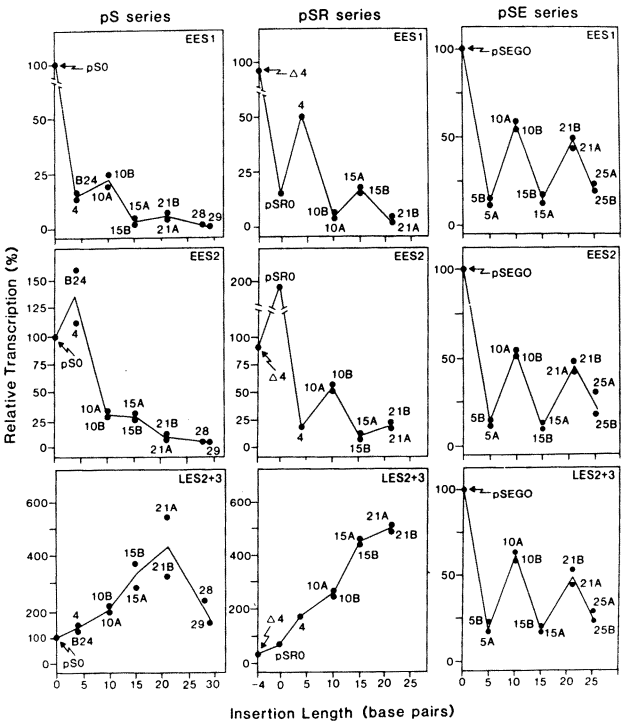


Fig. 11. Diagramatic comparison of the results obtained with the various insertion mutants. The length of the insertions (or deletion, pSR series) between the 21-bp repeat region and the TATA box region or the enhancer is indicated on the abscissa and the relative efficiency of transcription is given on the ordinate. The left-hand panels represent the results of the pS mutant series for transcription starting at EES1, EES2 and LES2+3, whereas the right-hand and center panels show the results for the pSE (insertions between the TATA box and inverted 21 bp repeat region) and pSR (insertion between the TATA box and inverted 21 bp repeat region mutant series, respectively. The values which are taken from Fig. 10 are expressed relative to EES1, EES2 and LES2+3 of pSO (for the pS and pSR series), pSEGO (for the pSE series). This figure was taken from Ref. 4.

direct initiation of transcription at EES2 (see above). Even in the reverse orientation, the 21-bp repeat region has to be correctly aligned with respect to the TATA-box element to preferentially activate transcription from the EES1, and similarly will preferentially stimulate initiation from the EES2 when aligned with the T-rich substitute TATA box (4; pSR series in Figures 10 and 11). Moving the 21-bp repeat region away from the TATA box region by > 20-bp results in a dramatic decrease of transcription from both EES1 and EES2, whereas initiation from the LES is concomitantly increased, indicating that the location of the 21-bp repeat region is now optimal for initiation from LES (Figure 11). Using longer DNA inserts leads to the appearance of new start sites located within 30-50-bp downstream from the 21-bp repeat region (4, 10), suggesting that the region at which transcription initiation occurs is primarily determined by the location of the 21-bp repeat region, and that the TATA box element selects the precise site of initiation.

From these experiments, it has been concluded that the variations in the transcriptional activity of the early promoter observed by inserting either odd or even multiples of half a helical DNA turn between the various promoter elements can be most easily explained by assuming that protein-protein interactions between the nucleoprotein complexes corresponding to the various promoter elements are involved in activation of transcriptional initiation (4).

CONCLUSIONS

The SV40 early promoter constitutes the prototype of a eukaryotic promoter : (i) The TATA box establishes directionality of transcription by deciding where the RNA polymerase starts and what direction it goes. (ii) The more upstream promoter elements and enhancer sequences determine the rate of transcriptional initiation. The question arises how do TATA box, upstream sequence elements and enhancers "talk to each other" to promote transcription ?

Several lines of evidence indicate that the various sequence elements of the promoter bind specific transcription factors to form a large nucleoprotein complex and that the multiple protein-protein interactions between its components generate an active transcriptional initiation complex at the promoter. From results obtained both in vivo and in vitro the following model has been proposed (4; see Figure 12, for a cartoon).

The TATA box factor binds to the TATA box to form a stable complex; this is a prerequisite for specific initiation of transcription (14, 15, Tamura et al., in preparation). The transcription machinery (presumably RNA polymerase B) interacts primarily with at least the first (GC-motif I) of the Sp1 factor molecules bound to 21-bp repeat region, which are all located on the same side of the DNA helix. The TATA box factor, when correctly positioned by a properly aligned TATA box, interacts with the RNA polymerase and directs it to initiate at the corresponding startsite (EES1 in the wild-type situation). Misalignment caused by inserting half a DNA turn between the TATA box and 21-bp repeat regions leads to the preferential use of the T box (substitute TATA box) which is located further upstream on the opposite side of the helix and controls initiation at the EES2.

In this model Sp1 factor molecules which are bound to GC-motifs II and III and are important for efficient initiation of transcription, may interact either directly with the transcription machinery or alternatively interact with one another and mediate their effect via the Sp1 factor bound to GC-motif I. Whether the Sp1 factor molecule bound to GC-motif I and the TATA box factor interact directly with each other as well as with RNA polymerase, for instance to stabilize their respective interactions with their cognate binding sites, remains to be investigated. However, stable binding of Sp1 factor to the 21-bp repeat region in vitro does not require the presence of a TATA box (81). Increasing the distance between the 21-bp repeat region and the TATA box (or its substitute), may either prevent the transcription machinery from interacting efficiently with the bound TATA box factor (even when properly aligned) or prevent the TATA box factor from binding efficiently to the TATA box, resulting in the observed rapid decrease in initiation from the EES1 and EES2 (4). Weak substitute TATA box elements may then direct the transcription machinery to preferentially initiate at the various LES (4). With longer inserts, similar weak multiple substitute TATA box elements (35) may direct the transcription machinery to initiate at multiple sites in the 30-50-bp region located immediately downstream from the 21-bp repeat region (10). Alternatively, RNA polymerase bound to the 21-bp repeat region may initiate at multiple sites without the assistance of the TATA box factor. Such a model readily accounts for the observation that there is no marked overall decrease in RNA initiated from the SV40

early promoter region in the insertion mutants of the pS series (Figure 10, see ref. 4).

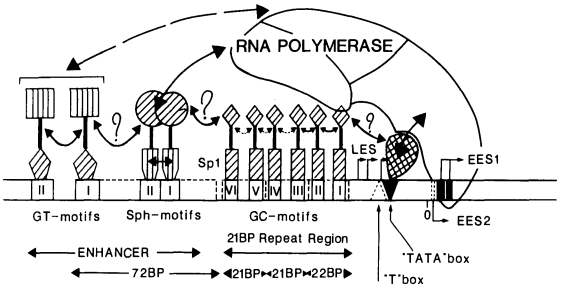


Fig. 12. A cartoon of the possible interactions (arrows) between the various transcription factors bound to their cognate sequence motif in the SV40 early promoter and their interaction with RNA polymerase. The different promoter elements, early-early (EES1 and EES2) and late-early (LES) startsites, TATA-box and T-box, the 21-bp repeat region (containing the GC-motifs I-VI) and the enhancer (Sph-motifs I and II, GT-motifs I and II) are indicated as in figures 1, 5, 9 and 10, together with a symbolic illustration of their cognate protein factors (e.g. TATA-box factor, Sp1 factor).

The in vivo studies of Takahashi et al. (4) indicate that equivalent protein-protein contacts can be established between the transcription machinery and the factors bound to the GC-motifs, irrespective of the orientation of the 21-bp repeat region. Since there is no apparent symmetry in the sequence, several explanation have been proposed to account for this paradox (4): (1) symmetry may be present in the GC-motif, even though it is not readily apparent; (2) factor Sp1 could be an assymmetrical molecule with equivalent protein interaction sites at either end; (3) the protein with which Sp1 factor interacts (for instance, RNA polymerase which has multiple subunits) may possess symmetrical protein interaction sites; (4) Sp1 factor may be composed of two functional domains linked by a flexible stem : one DNA-binding domain which would specifically recognize the GC-motifs, and one protein-interacting domain which would interact with some component of the transcription machinery. The two-domain hypothesis appears to be particularly attractive, since such separate domains for DNA binding and activation of transcription by protein interaction with RNA polymerase are known to exist in repressor proteins of E.coli λ phages (82-84).

Efficient initiation of transcription from the entire SV40 early promoter also requires a stereospecific alignment between some element(s) contained in domain A of the enhancer and the 21-bp repeat region (4). Enhancer specific transcription factor(s) are known to bind in vitro to domain A (20, 21, 23, 24). Thus, it is likely that it is these proteins which have to be stereospecific aligned with Sp1 factor molecules bound to the 21-bp repeat region for the enhancer to efficiently stimulate transcription. Since the important sequence elements of enhancer domain A (Sph-motifs I and II, Figure 5) are separated by approximately one helical turn from one another and by a multiple of 10 bases from the GC-motifs of the 21-bp repeat region, protein(s) bound to the Sph-motifs and the GC-motifs may be situated on the same side of the DNA helix, and may specifically interact with each other and/or independently with elements of the transcription machinery (for example RNA polymerase B). In this respect, it is interesting to note that the Sp1 factor binds to the 21-bp repeat region in vitro in the absence of the enhancer sequence (16, 17, 81), whereas enhancer-specific proteins seem to be less stably bound in vitro to the enhancer in the absence of the 21-bp repeat region (Wildeman, unpublished observation).

It appears, therefore, that the assembly of proteins bound to the various elements of the SV40 early promoter could constitute a broad protein interaction region located mainly on one side of the DNA helix, where it could act as a specific recognition region for RNA polymerase B. The function of the transcriptional factors bound to the 21-bp repeat and the TATA box region would be to anchor stably RNA polymerase to enable it to initiate at the start sites, and perhaps also to have a role in its activation. The protein(s) bound to enhancer domain A would participate in activation of transcription initiation, either directly by interacting with RNA polymerase, or indirectly through contacts with the protein(s) bound to the 21-bp repeat region. All these interactions may contribute to increase the transcription machinery concentration at a given promoter region which has to compete successfully with the rest of the very large genomic DNA for efficient initiation of transcription to occur. That the enhancer, and not only the 21-bp repeat region, can activate transcription irrespective of its orientation, can be explained by the hypothesis that proteins bound to it are composed of two distinct functional domains linked by a flexible stem (see above). A DNA looping mechanism (33, 85)

may account for the persistence of the enhancer activity (33, 35) as it is moved away from the other promoter elements.

In conclusion, our results clearly support the view that multiple protein-protein interactions have a key role in the generation, at a promoter site, of an active transcriptional initiation complex, which is a very large specific nucleoprotein structure composed of multiple sequence elements recognized by different cooperating factors. It is likely that this concept can be generalized to other class B promoters, as the same DNA turn dependence has been observed when the SV40 enhancer activates transcription from the adenovirus-2 major late promoter (our unpublished results).

ACKNOWLEDGEMENTS

We are most grateful to our colleagues whose names appear as authors of the papers published by our laboratory for their generous support and communication of results prior to publication. We thank the secretarial and illustration staffs of the LGME-U184 in Strasbourg and A. Walter and B. Blanasch of the EMBL in Heidelberg for their assistance with the preparation of the manuscript. The work performed at the LGME-U184 in Strasbourg was supported since 1979 by grants from the CNRS, INSERM, MRT, ARC and FRMF.

REFERENCES

1. Benoist, C. and Chambon, P. Nature 290: 304-310, 1981.
2. Wasylyk, B., Wasylyk, C., Matthes, H., Wintzerith, M. and Chambon, P. EMBO J. 2: 1605-1611, 1983.
3. Buchman, A.R., Fromm, M. and Berg, P. Mol. Cell Biol. 4: 1900-1914, 1984.
4. Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. Nature 319: 121-126, 1986.
5. Hansen, U. and Sharp, P.A. EMBO J. 2: 2293-2303, 1983.
6. Vigneron, M., Barrera-Saldana, H.A., Baty, D., Everett, R.E. and Chambon, P. EMBO J. 3: 2373-2382, 1984.
7. Ghosh, P.K. and Lebowitz, P. J. Virol. 40: 224-240, 1981.
8. Hansen, U., Tenen, D.G., Livingson, D.M. and Sharp, P.A. Cell 27: 603-612, 1981.
9. Tooze, J., ed. In: DNA Tumor Viruses, Cold Spring Harbor Lab. Press, N.Y., 1982.
10. Barrera-Saldana, H., Takahashi, K., Vigneron, M., Wildeman, A., Davidson, I. and Chambon, P. EMBO J. 4: 3839-3849, 1985.
11. Everett, R.D., Baty, D. and Chambon, P. Nucl. Acids Res. 11: 2447-2464, 1983.
12. Baty, D., Barrera-Saldana, H.A., Everett, R.D., Vigneron, M. and Chambon, P. Nucl. Acids Res. 12: 915-932, 1984.
13. Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. EMBO J. 5: 387-397, 1986.

14. Davison, B.L., Egly, J.M., Mulvihill, E.R. and Chambon, P. Nature 301: 680-686, 1983.
15. Parker, C.S. and Topol, J. Cell 36: 357-369, 1984.
16. Dynan, W.S. and Tjian, R. Cell 32: 669-680, 1983a.
17. Dynan, W.S. and Tjian, R. Cell 35: 79-87, 1983b.
18. Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P. and Tjian, R. Science 230: 511-517, 1985.
19. Sassone-Corsi, P., Dougherty, J.P., Wasylyk, B. and Chambon, P. Proc. Natl. Acad. Sci. USA 81: 308-312, 1984.
20. Sassone-Corsi, P., Wildeman, A.G. and Chambon, P. Nature 313: 458-463: 1985.
21. Wildeman, A.G., Sassone-Corsi, P., Grundström, T., Zenke, M. and Chambon, P. EMBO J. 3: 3129-3133, 1984.
22. Sergeant, A., Bohmann, D., Zentgraf, H., Weiher, H. and Keller, W. J. Mol. Biol. 180: 577-600, 1984.
23. Wildeman, A.G., Zenke, M., Schatz, Ch., Wintzerith, M., Grundström, T., Matthes, H., Takahashi, K. and Chambon, P. Mol. Cell. Biol. 6: 2098-2105, 1986.
24. Davidson, I., Fromental, C., Augereau, P., Wildeman, A., Zenke, M. and Chambon, P. Nature, 1986, in press.
25. Ghosh, P.K., Lebowitz, P., Frisque, R.J. and Gluzman, Y. Proc. Natl. Acad. Sci. USA 78: 100-104, 1981.
26. Mathis, D.J. and Chambon, P. Nature 290: 310-316, 1981.
27. Benoist, C. and Chambon, P. Proc. Natl. Acad. Sci. USA 77: 3865-3869, 1980.
28. Fromm, M. and Berg, P. J. Mol. Biol. Appl. Genet. 1:, 457-481, 1982.
29. Fromm, M. and Berg, P. J. Mol. Appl. Genet. 2: 127-135, 1983.
30. Gidoni, D., Dynan, W.S. and Tjian, R., Nature 312: 409-413, 1984.
31. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. Nucl. Acids Res. 9: 6047-6068, 1981.
32. Banerji, J., Rusconi, S. and Schaffner, W. Cell 27: 299-308, 1981.
33. Wasylyk, B., Wasylyk, C. and Chambon, P. Nucl. Acids res. 12: 5589-5608, 1984.
34. Hen, R., Sassone-Corsi, P., Corden, J., Gaub, M.P. and Chambon, P. Proc. Natl. Acad. Sci. USA 79: 7132-7136, 1982.
35. Wasylyk, B., Wasylyk, C., Augereau, P. and Chambon, P. Cell 32: 503-514, 1983.
36. Chambon, P., Dierich, A., Gaub, M.P., Jakowlev, S.B., Jongstra, J., Krust, A., LePennec, J.P., Oudet, P. and Reudelhuber, T. Recent Prog. Horm. Res. 40, 1-42, 1984.
37. Yaniv, M. Biol. Cell. 50: 203-216, 1984.
38. Gruss, P. DNA 3: 1-5, 1984.
39. Wasylyk, B. In: "Maximizing Gene Expression", W. Reznikiff and L. Gold (Eds)., Butterworths Publ. Group, Stoneham (Massa.) USA, pp. 79-99, 1986.
40. Serfling, E., Jasin, M. and Schaffner, W. Trends Genet. 1: 224-230, 1985.
41. Khoury, G. and Gruss, P. Cell 33: 313-314, 1983.
42. DeVilliers, J. and Schaffner, W. Nucl. Acids Res. 9: 6251-6264, 1981.
43. Laimins, L.A., Khoury, G., Gorman, C., Howard, B. and Gruss, P. Proc. Natl. Acad. Sci. USA 79: 6453-6457, 1982.
44. Kriegler, M. and Botchan, M. Mol. Cell. Biol. 3: 325-339, 1983.
45. Banerji, J., Olson, L. and Schaffner, W. Cell 33: 729-740, 1983.
46. Gilles, S.D., Morrison, S.L., Oi, V.T. and Tonegawa, S. Cell 33: 717-728, 1983.

47. Neuberger, M.S. EMBO J. 2: 1373-1378, 1983.
48. Picard, D. and Schaffner, W. Nature 307: 80-82, 1984.
49. Queen, C. and Baltimore, D. Cell 33: 741-748, 1983.
50. Grundström, T., Zenke, M., Wintzerith, M., Matthes, H.W.D., Staub, A. and Chambon, P. Nucl. Acids Res. 13: 3305-3316, 1985.
51. Weber, F., DeVilliers, J. and Schaffner, W. Cell 36: 983-992, 1984.
52. Swimmer, C. and Shenk, T. Proc. Natl. Acad. Sci. USA 81: 6652-6656, 1984.
53. Veldman, G., Lupton, S. and Kamen, R. Mol. Cell. Biol. 5: 649-658, 1985.
54. Herr, W. and Gluzman, Y. Nature 313: 711-714, 1985.
55. Herr, W. and Clarke, J. Cell 45: 461-470, 1986.
56. Weiher, H., Koenig, M. and Gruss, P. Science 219: 626-631, 1983.
57. Lusky, M., Berg, L., Weiher, H. and Botchan, M. Mol. Cell. Biol. 3: 1108-1122, 1983.
58. Weiher, H. and Botchan, M.R. Nucl. Acids Res. 12: 2901-2916, 1984.
59. Hen, R., Borrelli, E., Sassone-Corsi, P. and Chambon, P. Nucl. Acids Res. 11: 8747-8760, 1983.
60. Queen, C. and Stafford, J. Mol. Cell. Biol. 4: 1042-1049, 1984.
61. Nordheim, A. and Rich, A. Nature 303: 674-679, 1983.
62. Azorin, R. and Rich, A. Cell 41: 365-374, 1985.
63. Schöler, H.R. and Gruss, P. EMBO J. 4: 3005-3013, 1985.
64. Schöler, H.R. and Gruss, P. Cell 36: 403-411, 1984.
65. Mercola, M., Goverman, J., Mirell, C. and Calame, K. Science 227: 266-270, 1985.
66. Jakobovitz, E.B., Bratosin, S. and Aloni, Y. Nature 285: 263-265, 1980.
67. Saragosti, S., Moyne, G. and Yaniv, M. Cell 20: 65-73, 1980.
68. Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C.B., Jeltsch, J.M., Mathis, D. and Chambon, P. Nature 307: 708-714, 1984.
69. Weiss, E., Ruhlmann, C. and Oudet, P. Nucleic Acids Res. 14: 2045-2058 (1986).
70. Church, G.M. and Gilbert, W. Proc. Natl. Acad. Sci. USA 81: 1991-1995, 1984.
71. Wasylyk, C. and Wasylyk, B. EMBO J. 5: 553-560, 1986.
72. Mosthaf, L., Pawlita, M. and Gruss, P. Nature 315: 597-600, 1985.
73. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. Science 227: 134-140, 1985.
74. Falkner, F.G. and Zachau, H.G. Nature 310: 71-74, 1984.
75. Mason, J.O., Williams, G.T. and Neuberger, M.S. Cell 41: 479-487, 1985.
76. Grosschedl, R. and Baltimore, D. Cell 41: 885-897, 1985.
77. Mattaj, I.W., Lienhard, S., Jiricny, J. and DeRobertis, E.M. Nature 316: 163-164, 1985.
78. Citiberto, G., Buckland, R., Cortese, R. and Philipson, L. EMBO J. 4: 1537-1543, 1985.
79. Borrelli, E., Hen, R. and Chambon, P. Nature 312: 608-612, 1984.
80. Wang, J. Proc. Natl. Acad. Sci. USA 76: 200-203, 1979.
81. Miyamoto, N.G., Moncollin, V., Egly, J.M. and Chambon, P. Nucl. Acids Res. 12: 8779-8799, 1984.
82. Hochschild, A., Irwin, N. and Ptashne, M. Cell 32: 319-325, 1983.
83. Wharton, R.P., Brown, E.L. and Ptashne, M. Cell 38: 361-369, 1984.
84. Ptashne, M. Trends Biochem. Sci. 9: 142-145, 1984.
85. Ptashne, M. Nature 322: 697-701, 1986.

4

THE POLYOMA ENHANCER

J. PIETTE, M.-H. KRYSZKE and M. YANIV

Department of molecular Biology, Pasteur Institute, 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

ABSTRACT

We have analyzed the interaction of mouse 3T6 cellular proteins with the polyoma enhancer in vivo and in vitro. Four interacting domains are disclosed corresponding to functional domains defined in vivo. The interaction with the B enhancer domain was analyzed in more detail. Strong base-specific contacts were detected with the early proximal half of the GC-rich palindrome on the late coding strand. A 25 bp region including this sequence is sufficient to provide binding specificity : this region coincides with the core of the B element. The implication of these findings for polyoma enhancer function are discussed.

INTRODUCTION

Although the presence of an enhancer element in polyoma virus was revealed at about the same time as in simian virus 40 (SV40) (1, 2), most of the efforts were directed towards the understanding of the structure and function of the SV40 enhancer. Nevertheless, in addition to properties typical of enhancer elements, the polyoma enhancer possess some particular features, making it a very attractive and powerfull model for the study of the regulation of gene activity, DNA replication, and even differentiation in mammalian cells (3, 4).

The organization of the polyoma non-coding regulatory region is depicted in Fig. 1. Although the overall genome organization and the coding sequences of polyoma and SV40 viruses are similar (5), this is not the case for the control region : in polyoma the early promoter is further removed from the DNA replication origin, and it lacks the 21 bp GC-rich repeats found in SV40. The enhancer region itself does not display a repeat of 72 bp, nor is there any extensive homology with the SV40 enhancer. The main functional difference, possibly a direct

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

consequence of this different organization, is the absolute requirement in polyoma for the presence of the enhancer sequence in cis to allow replication of viral DNA (1, 6, 7). This is particularly clear, since replacement of the polyoma enhancer by an Immunoglobulin (Ig) heavy chain gene enhancer leads to tissue-specific replication of the virus in lymphocytes, cells where the Ig enhancer is normally active (7).

Another interesting particularity of the polyoma virus is the existence of a large number of host range mutants. Polyoma virus grows normally in mouse fibroblasts but it does not replicate in mouse embryonal carcinoma (EC) cells like F9 or PCC4 cells (8). In appropriate conditions, host range mutants can be selected that are adapted for growth on this type of cells (9) : they all carry point mutations or rearrangements in the enhancer region (10, 4), probably allowing appropriate interactions with factors present in EC cells in such a way that both transcription and replication of the viral DNA can occur. Similarly many differentiated cell lines of the mouse like erythroleukemia cells, neuroblasterian cells or trophoblasts are partially refractive to virus growth. Variants that are selected for fast growth on these cells contain rearrangements (deletions and duplications) in the enhancer region (4).

A characteristic of the polyoma enhancer, shared with other enhancers, is the organization in functional domains : some domains conserve partial activity, combination of different domains or even duplication or polymerization of a single domain restores full activity (see figure 2). Herbomel et al. showed that the polyoma enhancer sequences first defined by de Villiers and Schaffner (2) can be subdivided into two active redundant enhancers, A and B, corresponding to the BclI-PvuII and PvuII-PvuII fragments (11). A minimal core was defined for each of these enhancers. Sequences outside these cores were considered as auxillary sequences, the presence of which is required for full enhancer activity. These results and parallel experiments by the groups of Kamen and Hassel permit the division of the entire polyoma enhancer into four domains, A to D, in accordance with the suggestion of Veldman et al. (12). Domain A corresponding to the core of enhancer A contains a sequence homologous to the adenovirus (Ad) enhancer (14). It corresponds to the core or part of segment 2 defined by Hassel et al. (13). Domain B corresponds roughly to the core of

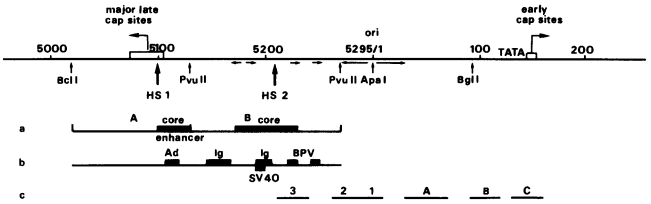


Fig. 1. : the regulatory region of polyoma virus. The numbering of nucleotides is according to Tyndall et al. (1). The restriction sites for BclI, PvuII, ApaI and BglI are indicated by small vertical arrows ; major late cape sites (39) and early cap sites (40) are indicated by boxes with arrows, palindromes or direct repeats by small horizontal arrows ; the two DNase I hypersensitive sites mapped in the chromatin (16) by large vertical arrows termed HS1 and HS2. Also indicated are the origin region (ori) and the TATA box sequence of the early promoter. (a) the 244 bp polyoma BclI-PvuII enhancer region is subdivided into the A (BclI-PvuII) and B (PvuII-4) enhancer elements with the enhancer core sequences identified by Herbomel et al. (11) represented by black bars. (b) The homologies between the polyoma enhancer and the enhancers of adenovirus (14), Ig heavy chain gene enhancer (41), BPV (42) and SV40 (15) are represented by black bars. (c) Binding sites of large T antigen as deduced from footprinting experiments are shown (43, 44). From Piette et al. (26), with permission.

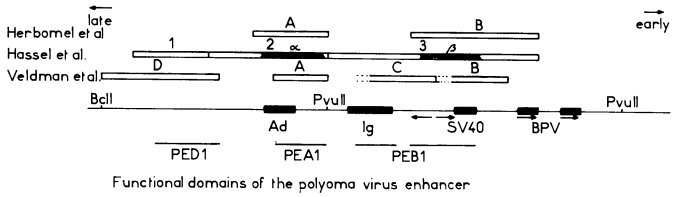


Fig. 2 : Organization of the polyoma enhancer in functional domains. The BclI-PvuII enhancer fragment is represented with symbols as in Fig.1. The functional domains as determined in the indicated references (respectively 11, 13 and 12) are represented by boxes. The DNase I protected domains detected in this work are indicated by lines with the names of the postulated factors in the lower part of the figure. See text for further details.

enhancer B (11). It is part of segment 3, roughly the β-core of Hassel et al. (13). It includes the Weiher and Gruss consensus sequence found in the SV40 enhancer among others (15) and a GC-rich palindrome. Domain C is adjacent to the core of B and contains a sequence homologous to the immunoglobulin heavy chain enhancer. Domain D is on the late side of the adenovirus homology (domain A) and is required for the full enhancer activity of the BclI-PvuII fragment. The nucleotide cores of either one of the two enhancers : domain A (α) or B (β) are sufficient to complement the DNA replication origin core to allow replication of the viral DNA (6, 13). Also typical for enhancer elements is the presence of DNase I hypersensitive sites in the polyoma minichromosomes: precise mapping revealed two major sites, one near the A domain and the Ad homology, the other near the B domain and the SV40 homology (ref 16 and Fig. 1). Finally, differential activity of the polyoma enhancers is noted in EC cells : in PCC3 cells for example, the B element is as active as in 3T6 fibroblasts, while the A element is clearly less active (11). This could be brought in parallel to the recent observation, that only the B enhancer and not the A enhancer allows replication of the viral DNA in microinjected mouse embryos (17).

We hope to show that in fact the different domains represent sites of action of cellular proteins with the viral enhancer. In one case at least, the interaction site can be reduced to a "core" corresponding to the functional core defined in vivo. The formation of DNase I hypersensitive sites is caused by the binding of proteins, possibly causing structural alterations in the DNA double helix.

RESULTS

We have used two approaches to characterize DNA-protein interactions with the polyoma enhancer. In the first, in vivo approach, contact sites were mapped in intact cells or nuclei by use of DNaseI digestion or dimethylsulfate alkylation followed by the genomic sequencing technique according to Church and Gilbert (18). In the second, in vitro approach, a combination of gel retardation assays (19, 20) and enzymatic or chemical footprinting techniques was used (21, 22).

Localization of protein-DNA contacts in vivo

A major problem in the in vivo approach is the non-homogeneity of the viral minichromosome population (5). As clearly shown for SV40, only a limited fraction of the minichromosomes displays the typical nucleosomal gap (maximal 20 %), possesses torsional stress or is associated with nascent transcriptional complexes (23, 24, 25). Thus, it is possible that only part of the minichromosomes will be interacting with "enhancer" binding proteins, this interaction will not necessarily be the same for all the molecules. The global in vivo picture of DNA-protein interactions may thus be the result of a superposition of different configurations. Aware of these problems, we choose to study the interaction with the polyoma minichromosomes 24 hours after viral infection. Most of the viral DNA molecules are then engaged in the late phase of the viral cycle and some are actively replicating. This allows us to obtain large quantities of material. In this phase, the potential cellular factors are still expected to interact with the enhancer region since the latter is necessary for replication and for the continued synthesis of early RNA ; DNase I hypersensitivity and the presence of a nucleosomal gap are still observed at this stage of infection (23, 16).

A DNase I footprinting experiment is displayed in Fig. 3. Four large domains of protection separated by small non protected regions can be discerned ; because they largely coincide with the domains defined by functional tests we call them in a similar way D, A, C and B. A detail of a methylation protection experiment is also given : closer interaction with specific base pairs can be detected in this way. The B domain is represented, emphasizing interactions with the early part of the GC-rich palindrome as will be confirmed by the in vitro analysis. To summarize, four interacting domains are detected on the polyoma enhancer, roughly corresponding to the functional domains ; the protections are only partial indicating that only part of the minichromosomes are interacting with specific enhancer–binding proteins. It is also not sure if the enhancer region is protected in an all or none fashion, although the precise juxtaposition of the different domains suggests that this could be the case. Nor is it known if the interactions disclosed here reflects the active state of the minichromosomes (thus positive factors) or the repressed state of

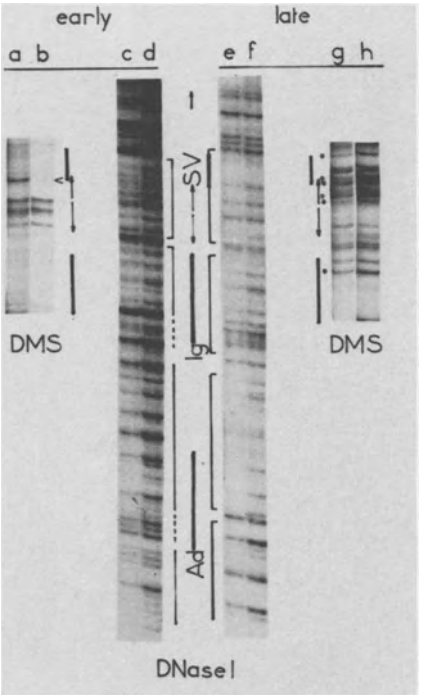


Fig. 3 : In vivo footprinting of the polyoma enhancer. In the DNase I protection experiments, 3T6 nuclei were prepared 24 hours after infection with the virus, as described in Cereghini and Yaniv (45). They were treated with DNase I in 130 mM (lanes c and e) or 300 mM NaCl (lanes d and f), DNA was subsequently isolated, and the genomic sequencing technique of Church and Gilbert was further followed (18). Lane (c) represents the footprint of the early, lane (e) of the late coding strand. Lanes (d) and (f) are taken as control respectively for the early and late coding strand. In the methylation protection experiments, 3T6 cells were treated with DMS 24 h after infection, the DNA was then extracted and the same procedure was followed as for the DNase I experiments. The protection pattern for the early strand (a) and late strand (g) are shown. Naked DNA is taken as control for the early coding strand in (b) and for the late strand in (h). Typical features of the sequence are represented next to the lanes with symbols as in Fig. 1. DNase I protections are indicated by brackets, methylation protections by o and enhancement by >.

minichromosomes (thus negative factors). It is clear that an in vitro approach is necessary to extend these findings and to confirm that the factors are really of cellular origin.

Characterization of cellular proteins that interact with the polyoma enhancer in vitro

If the hypothesis that cellular factors are interacting with the polyoma enhancer, and are in this way initiating the viral cycle, is correct, then protein extracts prepared from uninfected host cells like mouse 3T6 fibroblasts should contain such factors. We have found it useful to follow different methods to obtain a complete picture of the protein-DNA interactions involved : mainly DNase I footprinting and gel retardation (or bandshift) assays. Mouse 3T6 cells were collected before confluence, nuclei prepared and proteins extracted at different

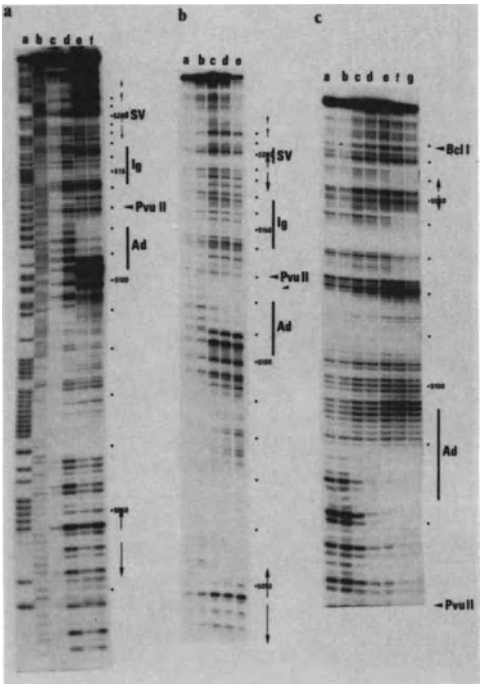


Fig. 4 : In vitro footprinting of the polyoma enhancer. A. 20 µg of a 0.4 M nuclear extract was incubated with a few ng of 3' labelled DNA, in the presence of 150 ng pdIdC carrier for 10 minutes at 20°C. The mixture was then treated with two DNase I concentrations in lanes (e) and (f) for 1 minute. After phenol extraction the DNA was loaded on a 6% sequencing gel. In lanes (c) and (d) no protein was added as control. B and C. Respectively 2.5 (lanes c), 5 (d), 10 (e), 20 (f) or 30 µg (g) of a dialyzed 0.4M nuclear extract was incubated with a few ng of 3' labelled DNA, in the presence of 150 ng pdIdC carrier during 10 minutes at 20°C. The mixture was treated with DNase I for 1 minute. After deproteinization with protease K and phenol extraction the DNA was loaded on a 6 or 8 % sequencing gel. In lanes (a) and (b), no protein was added as control. In B, the early coding strand of the BclI-ApaI fragment containing the entire enhancer is shown. In C the late coding strand of the BclI-PvuII fragment containing enhancer domains A and D is shown. Symbols are as in figure 1.

ionic strengths (26). For example, a 0.4 M nuclear extract means an extract containing proteins eluted at 0.4 M NaCl from the nuclei.

The first method we will describe is direct DNase I footprinting (21). Critical in these experiments is the use of high protein concentrations, low salt conditions and use of an appropriate carrier DNA, here poly(dI-dC).poly(dI-dC). In Fig. 4A a footprint of the enhancer region with a 0.4 M salt nuclear extract is shown. At least three, almost contiguous partially protected regions are seen : the first one overlaps partially the adenovirus (Ad) homology, the second one covers the Ig homology, the third one the GC-rich palindrome and the SV40 homology. Strong hypersensitive sites are bordering the late side of this region, weaker hypersensitive sites the early side. A weaker protected region is seen in the late part of the PvuII-BclI

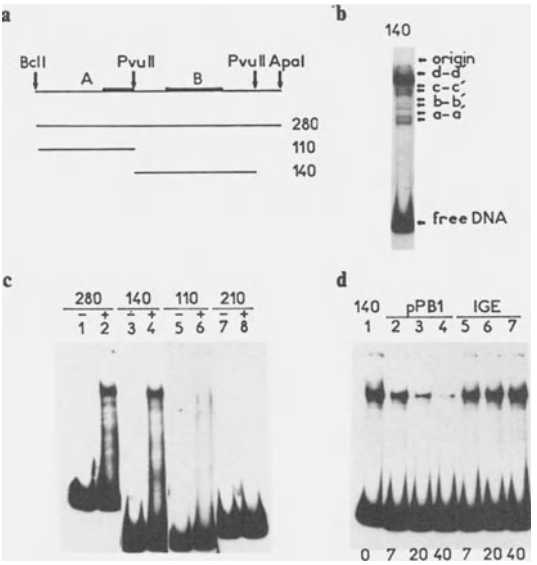


Fig. 5 : Band shifting experiments with the polyoma enhancer. A. Fragments used in the band shifting experiments. The region extending from the BclI site at position 5022 to the ApaI site at position 5291 containing the A and B enhancers is indicated : the enhancer core sequences are represented by a black bar. B. Band shifting pattern obtained with the B enhancer. The 140 bp B enhancer fragment together with an excess of non radioactive carrier DNA (sonicated salmon sperm) were incubated with a 0.4 M nuclear extract of 3T6 cell and loaded on a 7.5 % polyacrylamide gel. Indicated are the origin of the gel, the DNA fragments complexed by proteins (labelled a-a' to d-d') and the free DNA band. The four bands including d-d' were frequently observed as doublets. C. Band shifting experiments with different DNA polyoma fragments given in panel a. 1 and 2 : fragment 280 ; 3 and 4 : fragment 140 ; 5 and 6 : fragment 110, lanes 7 and 8 : 210 bp fragment of prokaryotic origin. The fragments were incubated without (-) or with (+) a 0.4 M nuclear extract of 3T6 cells. D. Competition experiments. A nuclear extract of 3T6 was incubated with the B enhancer in the absence (lane 1) or in the presence of linearized pPB1 (a plasmid containing the B enhancer) (lanes 2-4) or linearized IGE (a plasmid containing the mouse heavy chain gene enhancer). A 7 to 40 molar excess of competitor was used as indicated in the figure. From Piette et al. (46), with permission.

fragment. This protection is clearly visible in figure 4B and 4C when a dialyzed extract was used for the footprinting. It should be noted that the protection of the Ig homology, palindrome and SV40 homology are lost upon dialysis of the extract. These results are schematized in Fig. 8. A close correlation with the functional domains mapped in vivo, or with the protected domains mapped by in vivo footprinting is observed : domain A is strongly protected, B, C and D more weakly. Moreover, the two hypersensitive regions mapped here coincide with those mapped in vivo by the indirect end-labelling technique (16). In the D, A, and B-C domains, the protections we observe are due to

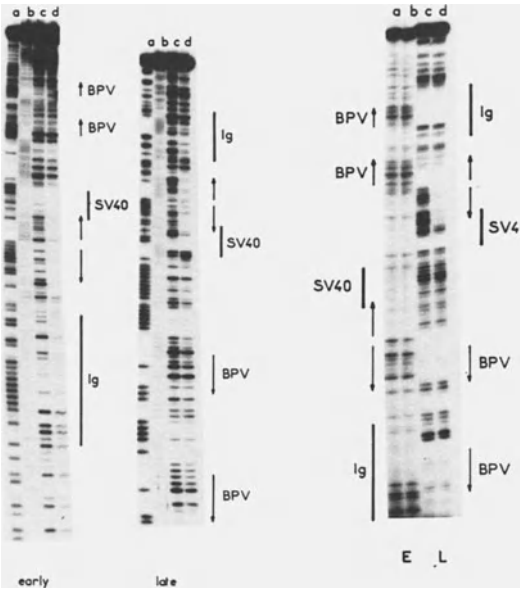


Fig 6. : Footprint of the DNA-PEBI complex The complex was allowed to form as described in Piette et al. (26) (see also Fig. 5). In the DNase I footprint experiments (A) it was treated with DNase I for 1 minute at 20°C, and the digestion was slowed down by the addition of 10 µg of salmon sperm DNA. This was immediately loaded on a preparative gel and the free DNA band (lane c) was separated from the retarded band (d), eluted and loaded on a 8 % sequencing gel. In the DMS interference experiments (B), the DNA was methylated before complex formation (47) and further processed as above. (A) lanes (a) and (b) : GA and TC sequence reaction products ; lane (c) : free DNA ; lane (d) : complexed DNA. Both early and late strands are shown. (B) lanes (a) and (c) : free DNA ; lanes (b) and (d) : complexed DNA of respectively the early and late coding strands of the PvuII-4 fragment containing enhancer domain C and B.

different factors, since the protecting activity of D and A elutes already at 0.3 M while the B-C activity elutes around 0.4 M NaCl from the nuclei. The availability of point mutations which inactivate the A domain (kindly provided by G. Magnusson) allowed us to test the correlation between binding and enhancer activitity in vivo. We observed clearly less protection of the mutated A domain when the DNase I footprinting reaction was done in exactly the same conditions as for the wild type. Thus at least one domain involved in interaction with a cellular protein is functionally participating in the process of enhancement.

We wished to obtain a more detailed analysis of the different interactions disclosed here. For this purpose we turned to gel retardation assays (19, 20). This method is based on the slower

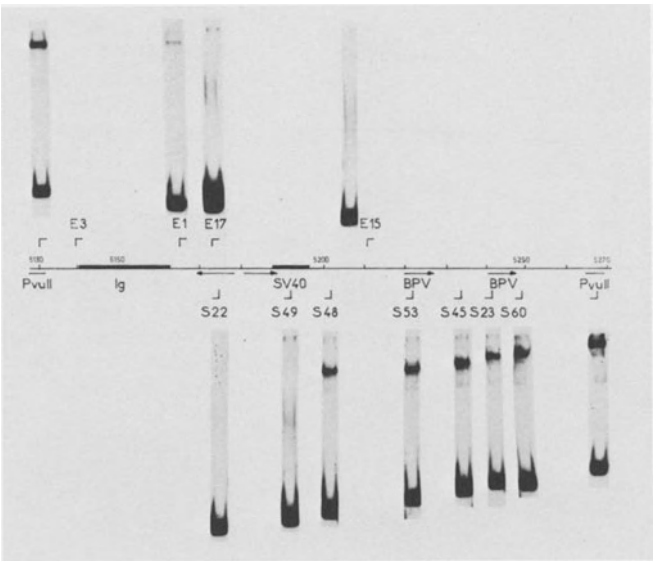


Fig. 7 : Deletion analysis of the PvuII-4 fragment. The PvuII-4 fragment was digested with Bal31 nuclease from respectively the late proximal extremity (ΔE series) or the early proximal extremity (ΔS series). The obtained fragments were tested for their capacity to form a complex with a 0.5 M nuclear extract by gel retardation assays (26). The PvuII-4 fragment is represented with the typical features symbolized as in Fig. 1. Above, the new extremity of the ΔE series fragments is shown with the pattern of the cognate fragment in the retardation assay. Below, the ΔS series is shown. All the tests were run on the same 7.5 % polyacrylamide gel ; origin and lower retarded band are indicated by an arrow.

migration in polyacrylamide gels of protein-DNA complexes with respect to free DNA. Two essential conditions have to be met, the interaction must be stable and specific. The latter requirement has to be carefully checked. In Fig. 5 such an analysis is presented: the specificity is confirmed by (i) lack of interaction with a fragment of plasmid DNA of prokaryotic origin (Fig. 5c), (ii) competition with the

homologous Py fragment present on a plasmid (Fig. 5d). The interaction is mapped within the PvuII-4 fragment (Fig. 5c). Surprisingly no interaction is detected with the BclI-PvuII fragment, in contrast to what was observed in the preceding direct DNase I footprinting experiments (Fig. 5c). A reproducible pattern of retarded bands that are regularly spaced on the gel is obtained (Fig. 5b): the significance of this is not yet clear.

A finer mapping of the interaction between proteins and DNA can be

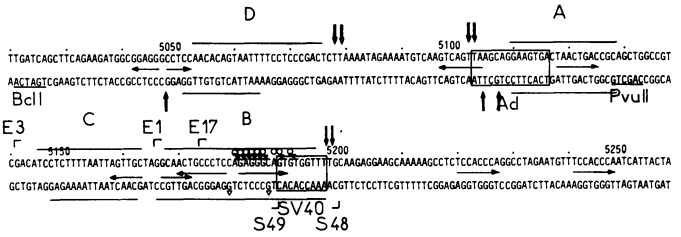


Fig. 8 : In vitro protein-DNA contacts with the polyoma enhancer. The sequence of the BclI-PvuII fragment containing the whole enhancer region is shown. Palindromes or repeats are represented by horizontal arrows. Homologies to the Adenovirus and SV40 enhancers are boxed. DNase I protection are indicated by horizontal lines, hypersensitive sites by vertical arrows, methylation interference by o, ethylation interference by ▼ and methylation enhancement by ▽. The end point of the relevant deletions is represented by ⌐. See text for further details.

obtained by combined use of the retardation assay and DNase I footprinting. After treatment of the protein-DNA complex by DNase I, the complex is separated from free DNA in a preparative gel and the DNase I pattern obtained from both are compared on a sequencing gel. The footprint of the factor will thus be visible on the retarded DNA. Such an experiment is shown in Fig. 6A. Both B and C domains are protected in a similar way as in the preceding footprinting experiments. Thus, either only one factor is recognizing the two domains, or two factors are closely associated. Destruction of the interaction by protease K treatment indicates that the factors are at

least in part proteins. The same rationale can be followed now for methylation or ethylation interference experiments to obtain a picture of the close contacts between DNA and protein. In the methylation experiments the DNA probe is first treated with dimethylsulfate (DMS) to methylate the 3' position of guanine (G) in the major groove or the 7' position of adenine (A) in the minor groove of the DNA helix. The complex is then allowed to form and separated from free DNA on gel. Methylated bases that interfere with the binding will not be present in the retarded DNA (see Fig 6B). In ethylation experiments with ethylnitrosourea (ENU) ionic interactions with the phosphates of the DNA backbone are revealed (22). The results of such experiments are schematized in Fig. 8. It is striking that the interaction is exclusively observed with the purine stretch of the early part of the GC-rich palindrome on the late coding strand. No strong interference or protection is observed with the other strand that includes the other half palindrome. The importance of this core sequence for the formation of the DNA-protein complex is confirmed by deletion analysis.

A series of deletions were constructed by Bal31 digestion : the fragments obtained were then tested for their capacity to bind the nuclear factor(s). When no more binding was observed the deleted DNA was replaced by plasmid DNA to check if the requirement for the deleted DNA was specific. The relevant deletions are represented in Fig. 7. We concluded that ΔE1 and ΔE17 remove specific sequences from the late side that contribute to complex formation. Deletion S49 removes such sequences from the early side. The sequences missing in ΔE1 and ΔE17 can be fully or partially replaced by non-specific plasmid DNA as judged by the yield of specific complex. On the contrary replacement of the sequences deleted in ΔS49 by plasmid sequences did not restore complex formation. We conclude that the essential sequences are contained in the 25 bp stretch between the end points of ΔE17 and ΔS48: the same region where the strong protein-DNA contacts were mapped. This sequence corresponds nearly exactly to the β-domain core required for DNA replication as mapped in vivo by Hassel et al. (13). We propose that only one factor, that we call PEB1 (polyoma enhancer-binding protein 1) is interacting with both B and C domains. Indeed, there is no shift in the migration of the complex when C is removed, or is replaced by plasmid DNA. Thus, the factor would recognize and interact

strongly with the B core domain, and subsequently make weaker contacts (detectable by DNase I footprinting) with the C domain : these contacts are still specific, since no protection of the plasmid DNA replacing the C domain is seen in DNaseI footprinting experiments. When run on a glycerol gradient, the binding activity migrates as a discrete 3.5 S peak, indicating that the factor could be in the molecular weight range of 50 Kd.

DISCUSSION

A first important point that can be deduced from our results is that the enhancer of polyoma virus is interacting with cellular proteins already present in uninfected cells. Such proteins may thus have a role in normal cells, as is the case with Sp1 transcription factor which interacts with SV40 or herpes promoters and with the regulatory region of several cellular genes (27). The accumulation of such regulatory protein binding sites in viral enhancers may contribute to make the virus a very powerful competitor for the cellular transcription and replication machinery. Both our in vivo and in vitro results indicate that a large portion of the enhancer region might be covered by proteins in an almost continuous fashion (nt 5000 to nt 5200). This region is recognized by at least three distinct activities: PEA1 and PEB1 binding to respectively domain A and B-C defined in vivo, a third factor PED1 binding to domain D. PEB1 is probably a single factor recognizing the B domain core, and interacting weakly with the C domain. This strongly suggests that the domains defined in vivo by functional tests correspond to distinct protein binding domains, and that these proteins play a crucial role in enhancer function. This is confirmed for the PEA1 factor, since this factor showed a drop in its binding affinity for an enhancer negative point mutation. The binding of specific factors is also the cause of hypersensitive sites to DNase I at loci similar to those detected in vivo. These sites are localized at the border of protected regions. In the case of PEB1 there is evidence that this site coincide with a local alteration in the DNA structure ; indeed the proximity of this site to the extremity of the DNA fragment resulted in a marked increase in migration mobility of the complex in polyacrylamide gels, a phenomenon that could be caused by DNA bending (28, 29). As already suggested

before, DNaseI hypersensitive sites may thus be caused by the presence of specific DNA binding factors and structural alterations in the DNA (30, 31). One of the DNA-protein interactions was characterized more precisely, i.e. that of PEB1 with the PvuII-4 fragment. It follows from this analysis that the essential contacts are contained in a 25 bp stretch that was defined as the β-core element by Hassel et al. (13) : the minimal sequence able to complement the DNA replication origin core to allow viral DNA replication. Remarkably, strong contacts are restricted to a purine-stretch on the late coding strand. The DNaseI protection extends further to the late side, covering about 50 bp of DNA. The C domain can nevertheless be removed with only a small drop in binding affinity. There is a striking similarity of this type of interaction, with that of TFIIIA and the 5 S RNA gene, although we do not observe significant sequence homology (32, 33). TFIIIA-5S gene like interactions may thus be involved in enhancer function. It is intriguing in that respect, that several eukaryotic regulatory proteins sequenced up to now show a so called finger-like domain organization similar to that of TFIIIA (34), what could indicate a quite general use of such interactions for gene regulation in eukaryotes.

Are there any other proteins interacting with the polyoma enhancer ? A factor different from those we described here was described by Fujimura (35). It binds to a small palindrome between nt 5158 and nt 5172. We have detected a similar activity after fractionation of a 3T6 nuclear extract on a heparine-agarose column. It is probably obscured in our current assays by the stronger activity of PEB1. The two binding sites are overlapping, and a finer regulation of enhancer activity could be achieved in this way : the balance between the two factors could vary, for example in different tissues or during the cell cycle. The real importance of the homologies to the Ad and SV40 enhancers needs also to be elucidated. Both sites are located at the extremity of binding domains in DNase I hypersensitive regions. May be, other factors are recognizing these sequences only after binding of PEA1, or PEB1. Such factors may perform some crucial steps in enhancer function as suggested by the presence of similar sequences in other enhancers and the detrimental effect on enhancer activity of point mutations in these sequences in the SV40 72 bp repeat (15).

The exact role of the factors described in this work in enhancer

function is unknown. In fact, their existence is compatible with the possible mechanisms proposed for the enhancer function. The presence of a large nucleo-protein complex could be responsible for nucleosome exclusion and the creation of a proper RNA polymeraseII (or associated factor) entry site. This nucleo-protein complex could also provide a strong nuclear matrix attachment site where transcription and replication processes would take place, or trigger DNA topoisomerase action to change the linking number of the minichromosomes. The binding of these factors by itself could induce a torsional stress in the DNA creating transcription and replication competent minichromosomes. All these alternatives are of course, not mutually exclusive. Real understanding of the mechanisms of enhancement should await the set-up of a proper in vitro system and the purification of the enhancer binding factors. Preliminary in vitro results were obtained with the SV40 (36, 37, 38) or Ig enhancer (37) with an in vitro trancription assay, however the components of this system are far from being pure. The particularity of polyoma virus that requires an enhancer also for the replication of its DNA may offer a more promising alternative. In any case, a complete understanding of the enhancer mechanism may well require among others the use of specific antibodies to enhancer factors to investigate the localization and function of these factors in the intact cell.

ACKNOWLEDGEMENTS

We thank F. Arnos and A. Doyen for technical assistance, J. Ars for her help in the preparation of the manuscript and D. Böhmann for advice in DNaseI footprinting. This work was supported by grants from the CNRS (UA 041149 and ATP 955515), the INSERM (C.R.E. n°852025, the ARC and the Fondation pour la Recherche Medicale Française. J. Piette was supported by fellowships from EMBO and the ARC.

REFERENCES

1. Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J. and Kamen, R. Nucl. Acids Res. 9:6231-6250, 1981.
2. de Villiers, J. and Schaffner, W. Nucl. Acids Res. 9:6251-6264, 1981.
3. Yaniv, M. Nature 297:17-18, 1982.
4. Amati, P. Cell 43:561-562, 1985.
5. Tooze, J. DNA tumor viruses. Molecular Biology of tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1981.
6. Muller, W.J., Mueller, C.R., Mes, A.M. and Hassell, J.A. J. Virol 47:586-599, 1983.
7. de Villiers, J., Schaffner, W., Tyndall, L., Lipton, S. and Kamen, R. Nature 312:242-246, 1984.

8. Swartzendruber, D.E. and Lehman, J.M. J. Cell Physiol. 85:179-188, 1975.
9. Vasseur, M., Kress, C., Montreau, N. and Blangy, D. Proc. Natl. Acad. Sci. USA 77:1068-1072, 1980.
10. Katinka, M., Yaniv, M., Vasseur, M. and Blangy, D. Cell 20:393-399, 1980.
11. Herbomel, P., Bourachot, B. and Yaniv, M. Cell 39:653-662, 1984.
12. Veldman, G.M., Lupton, J. and Kamen, R. Mol. Cell. Biol. 5:649-658, 1985.
13. Muller, W.J., and Hassel, J.A. Cancer cells 4, in press, 1986.
14. Hearing, P. and Shenk, T. Cell 33:695-703, 1983.
15. Weiher, H., Konig, M. and Gruss, P. Science 219:626-631, 1983.
16. Herbomel, P., Saragosti, S., Blangy, D. and Yaniv, M. Cell 25:651-658, 1981.
17. Wirak, D.O., Chalifour, L.E., Wassarman, P.M., Muller, W.J., Hassel, J.A. and de Pamphilis, M.L. Mol. Cell Biol. 5:2924-2935, 1985.
18. Church, G.M. and Gilbert, W. Proc. Natl. Acad. Sci. USA 81:1991-1995, 1984.
19. Garner, M.M. and Revzin, A. Nucl. Acids Res. 9:3047-3060, 1981.
20. Fried, M. and Crothers, D.M. Nucl. Acids Res. 9:6505-6525, 1981.
21. Galas, B.J. and Schmitz, A. Nucl. Acids Res. 5:3157-3170, 1978.
22. Siebenlist V., Simpson, R.B. and Gilbert, J. Cell 20:269-281, 1980.
23. Saragosti, S., Moyne, G. and Yaniv, M. Cell 20:65-73, 1980.
24. Luchnik, A.N., Bakayer, V.V., Zbarsky, I.B. and Georgiev, J.P. EMBO J. 1:1353- , 1982.
25. Choder, M., Bratosin, S. and Aloni, Y. EMBO J. 3:2929-2936, 1984.
26. Piette, J., Kryzske, M.-H. and Yaniv, M. EMBO J. 4:2675-2685, 1985.
27. Dynan, W.S., and Tjian, R. Nature 316:774-778, 1985.
28. Marini, J., Levene, S., Crothers, D.M. and Englund, P.T. Proc. Natl. Acad. Sci. USA 79:7664-7668, 1982.
29. Wu, H.M. and Crothers, D.M. Nature 308: 509-513, 1984.
30. Elgin, S. Nature 309:213-214, 1984.
31. Emerson, B.M. and Felsenfeld, G. Proc. Natl. Acad. Sci. USA 81: 95-99, 1984.
32. Sakonju, S. and Brown, D.D. Cell 31:395-405, 1982.
33. Smith, D.R., Jackson, I.J. and Brown, D.D. Cell 37:645-652, 1984.
34. Berg, J.M. Nature 319:264-265, 1986.
35. Fujimura, F.K. Nucl. Acids Res. 14:2845-2861, 1986.
36. Sassone-Corsi, P., Wildeman, A. and Chambon, P. Nature 313:458-463, 1985.
37. Schöller, H.R. and Gruss, P. EMBO J. 4:3005-3013, 1985.
38. Sargeant, A., Bohmann, D., Zentgraf, H., Weiher, H. and Keller, W. J. Mol. Biol. 180:577-600, 1984.
39. Cowie, A., Tyndall, C. and Kamen, R. Nucl. Acids Res. 9:6251-6264, 1981.
40. Cowie, A., Jat, P. and Kamen, R. J. Mol. Biol. 152:225-255, 1982.
41. Banerji, J., Olson, L. and Schaffner, W. Cell 33:729-740, 1983.
42. Weiher, H. and Botchan, M.R. Nucl. Acids Res. 12:2901-2916, 1984.
43. Cowie, A. and Kamen, R. J. Virol 52:750-760, 1984.
44. Dilworth, S.M., Cowie, A., Kamen, R. and Griffin, B.E. Proc. Natl. Acad. Sci. USA 79:1959-1963, 1984.
45. Cereghini, S. and Yaniv, M. EMBO J. 3:1243-1253, 1984.
46. Piette, J., Cereghini, S., Kryszke, M.-H. and Yaniv, M. Cancer Cells 4, in press, 1986.
47. Maxam, A. and Gilbert, W. Meth. Enzymol. 65:499-559, 1980.

5

IN VITRO POLYADENYLATION OF SV40 EARLY PRE-mRNA

LISA C. RYNER, MURARI CHAUDHURI AND JAMES L. MANLEY

Department of Biological Sciences, Columbia University, New York, N.Y. 10027

ABSTRACT

Using a pre-mRNA containing the SV40 early introns and poly(A) addition site, we have investigated several requirements for accurate and efficient mRNA 3' end cleavage and polyadenylation in a HeLa cell nuclear extract. 3' end formation and splicing occur under the same conditions in vitro, but do not appear to be coupled in any way. The ATP analog 3'dATP (cordycepin triphosphate) inhibited both cleavage and polyadenylation even in the presence of ATP, perhaps reflecting an interaction between cleaving and polyadenylating activities. A 5' cap structure appears not to be required for mRNA 3' end processing in vitro, because neither the presence nor absence of a 5' cap on the pre-mRNA nor the addition of cap analogs to reaction mixtures had any effect on the efficiency of 3' end processing. Micrococcal nuclease pretreatment of the nuclear extract inhibited cleavage and polyadenylation. However, restoration of activity was achieved by addition of purified E. coli RNA, suggesting that the inhibition caused by such a nuclease treatment was due to a general requirement for mass of RNA, rather than to the destruction of a particular nucleic acid-containing component, such as a snRNP.

INTRODUCTION

The development of cell-free systems that accurately and efficiently process pre-mRNA by splicing and 3' end cleavage and polyadenylation has enabled investigators to begin detailed analyses aimed at determining the biochemical char-

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

acteristics and precise mechanisms of these reactions (1-6). Previous studies have indicated that eucaryotic mRNA 3' end formation involves an endonucleolytic cleavage of a longer precursor RNA followed by polymerization of approximately 200 adenylate residues to the newly formed 3' end, and that these two reactions might normally be coupled (7-11). The highly conserved nucleotide sequence, 5'-AAUAAA-3', found approximately 10 to 30 nucleotides upstream of the poly(A) addition site (12), plays an essential role in both the cleavage and polyadenylation reactions (13-19). Other less conserved sequences found downstream of the poly(A) addition site also appear to be required for accurate and efficient 3' end processing in vivo (20-26).

In vitro RNA processing data suggests that a cap structure at the 5' end of the pre-mRNA molecule may be required for efficient splicing (27, 28). The efficiency of splicing was found to be reduced when uncapped pre-mRNAs were used and when cap analogs, which might compete for factors that recognize the 5' cap structure, were added to reaction mixtures. Recent in vivo data argues against a 5' cap structure being required for efficient mRNA 3' end processing, however. When a hybrid gene comprising a RNA polymerase III promoter attached to the protein-coding sequences of a RNA polymerase II gene was transiently expressed in a human cell line, accurate and efficient 3' end formation of the chimeric transcript was observed (29). This implies that a 5' cap structure is not required for mRNA 3' end formation, because all RNA polymerase III transcripts thus far examined lack such structures.

Recently, the involvement of small nuclear ribonucleoproteins (snRNPs) in eukaryotic mRNA processing has been demonstrated for mRNA splicing (U1, U2, U4/U6 and possibly U5; ref. 30-34) and for 3' end formation of a histone transcript (U7; Ref. 35). Whether snRNPs play a role in 3' end cleavage and polyadenylation remains an open question, although several lines of evidence suggest their involvement. Polyclonal antisera against U1 RNP and La nuclear antigen

(an antigen associated with RNA polymerase III transcripts) have been shown to inhibit the cleavage and polyadenylation reactions in vitro (4, 10). In addition, RNase T1 fragments of pre RNAs containing an AAUAAA consensus sequence were immunoprecipitated from HeLa cell nuclear extracts with anti-Sm sera, which precipitates snRNPs containing U RNAs, and with anti-trimethylguanosine cap sera, which precipitates snRNPs containing the cap structure found at the 5' ends of most U RNAs (36). Although these results are consistent with an involvement of snRNPs in mRNA 3' end processing, they do not prove it. Neither these studies nor others (33) have been able to provide evidence for the involvement of a particular snRNP in cleavage and/or polyadenylation.

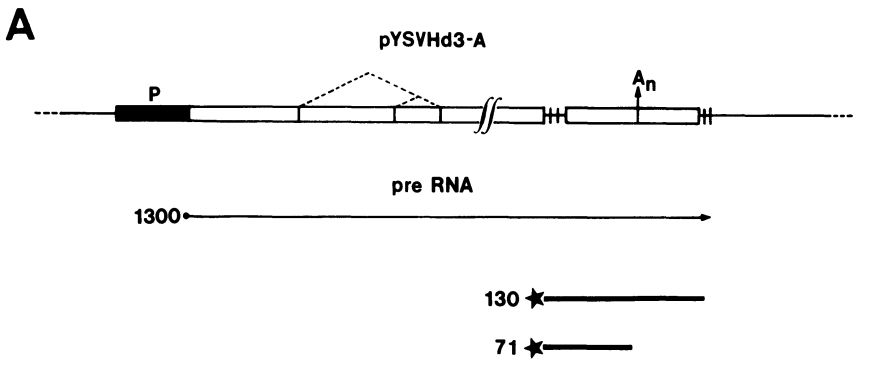
To investigate further the requirements for mRNA 3' end cleavage and polyadenylation, we have studied 3' end processing of a pre-RNA containing the SV40 early introns and poly(A) addition site in a HeLa cell nuclear extract. Here we present data demonstrating that accurate and efficient cleavage and polyadenylation of this pre-RNA can occur in this extract under the same conditions in which splicing occurs. In addition, we present evidence that the 3' end cleavage and polyadenylation activities are linked and that these reactions do not require a 5' cap structure on the pre-mRNA.

One way the requirement for an RNA component in mRNA splicing was established was by inactivation of splicing in a HeLa cell nuclear extract by micrococcal nuclease pretreatment and subsequent restoration of activity by addition of an snRNP-containing fraction (30, 31). Here we describe the results of similar experiments, which demonstrate that micrococcal nuclease pretreatment inhibits 3' end cleavage and polyadenylation, as well as splicing, in a HeLa cell nuclear extract. However, accurate 3' end processing was completely restored by addition of E. coli RNA, whereas splicing was not. Conversely, splicing was restored by addition of an snRNP-containing fraction, but 3' end cleavage was not.

RESULTS

Both splicing and polyadenylation of SV40 early region transcripts have been previously studied and display many of the features found to be typical of mRNA processing (37-39). Hence, we have used in these studies a precursor RNA generated by run-off transcription from the recombinant plasmid pYSVHd3-A, which contains both the introns and poly(A) site of the SV40 early region (Fig. 1A). When capped pYSVHd3-A pre mRNA was incubated in the presence of a HeLa cell nuclear extract plus ATP, creatine phosphate and Mg^{2+}, accurate splicing and 3' end processing were detected. S1 nuclease analysis of the RNA 3' ends indicates that at optimal conditions approximately 20% of the RNA recovered after processing was cleaved at the correct poly(A) site utilized in vivo (Fig. 1B). Oligo d(T) selection followed by S1 nuclease analysis showed that greater than 80% of the uncleaved pre-mRNA was polyadenylated at its 3' end and that all the RNA cleaved at the correct poly(A) site was polyadenylated (data now shown). When a pre-mRNA that did not contain the poly(A) site sequences, generated by cleaving the pYSVHd3-A template upstream of these sequences, was processed in the nuclear extract, no polyadenylation was detected (Fig. 1B). This is consistent with our previous results which showed that end-polyadenylation is dependent on the AAUAAA signal sequence (3, 15). Splicing was routinely monitored by the presence of the large T antigen intron lariat, (Fig. 1B) which can be easily detected because it is not polyadenylated and thus migrates as a sharp band during agarose gel electrophoresis (40). Other splicing products and intermediates were detected by S1 nuclease analysis (data not shown). We note that both splicing and 3' end processing occurred optimally under the same conditions. Furthermore, splicing efficiency was not detectably affected by the occurrence of 3' end processing on the same pre mRNA.

As noted by us and others (41), the SV40 early pre mRNA is 3' end processed considerably less efficiently than several other pre-mRNA's tested to date (e.g., adenovirus L3; Ref. 10).



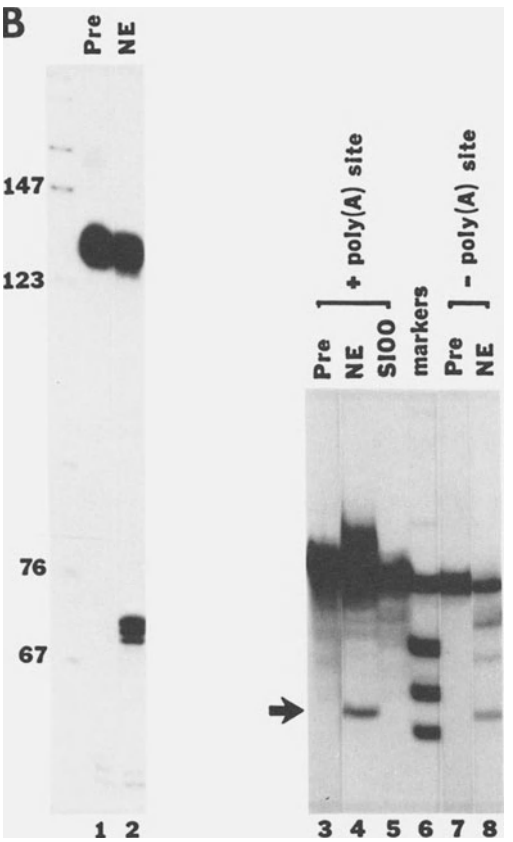


Figure 1. Processing of pYSVHd3-A pre RNA.

A, Schematic diagram of pYSVHd3-A pre RNA and of S1 nuclease analysis of processed RNA. The DNA template, pYSVHd3-A, was constructed by inserting a fragment containing the SV40 early poly(A) addition site into plasmid pYSVHdB (40), downstream of the SV40 Hind IIIB fragment. Pre-RNA synthesis from this template was directed by a synthetic E. coli promoter (48). A 1300 nt capped pYSVHd3-A pre RNA synthesized by run-off transcription contrains the following sequences: 5'7mGppApU / 5171 to 4001 nt (SV40) / 20 nt (PstI and SalI sites) / 2644 to 2533 nt (SV40) / 6 nt (Eco RI site) -3'. S1 nuclease analysis was used to analyze the 3' ends of the RNA after processing; protected fragments are 130 nt for pre RNA and 71 nt for pre RNA accurately processed at the poly(A) site. B, Splicing and 3' end processing of pYSVHd3-A pre RNA. S1 nuclease analysis of the RNA 3' ends: lane 1, pYSVH3-A pre-RNA;

lane 2, pre RNA processed in nuclear extract (NE). Size analysis of processed pre-RNAs: lane 3, pYSVHd3-A pre RNA; lane 4, pre-RNA processed in NE; lane 5, pre RNA processed in S100 lysate; lane 6, RNA size markers: 1200 nt, 650 nt, 415 nt and 250 nt; lane 7, pre-RNA generated from pYSVHd3-A that was cleaved at the SalI site, upstream of the poly(A) site sequences; lane 8, pre RNA from lane 7 processed in NE. The arrow points to the position of the SV40 large T antigen intron lariat.

Methods: Precursor RNA was synthesized in a 50 µl reaction containing 2 pmol E. coli RNA polymerase (gift from S. Beychok) and 0.8 pmol template in a buffer comprised of 20 mM Tris-HCl (pH 7.9), 0.4 M NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 2.5 mM DTT, 1 mM 7mGpppA, 100 µM ATP, 500 µM CTP, UTP and GTP, and 40 µCi α-^{32}P UTP. After an hour at 30 °C transcription was terminated by adding rifampicin at 500 µg/ml and 10 min later DNase I at 20 µg/ml and incubated for an additional 10 min at 30 °C. RNA was extracted twice with phenol:chloroform (1:1), twice with chloroform and then precipitated twice with ethanol. Transcriptions performed using this protocol yielded 7-10 pmol of precursor RNA. The nuclear extract was prepared from HeLa cells essentially by the method of Dignam et al. (49) except the dialysis buffer contained 0.042 M $(NH_4)_2SO_4$ instead of 0.1 M KCl and PMSF was not used in any of the buffers. S100 lysate was also prepared by the method of Dignam et al. except after centrifugation at 100,000g the supernatant was ammonium sulfate precipitated by adding 0.35 gm $(NH_4)_2SO_4$ for each ml of S100 supernatant and 1 µl of 1 M NaOH for each gm of $(NH_4)_2SO_4$. The precipitate was centrifuged at 20,000g, resuspended in 0.1 volumes buffer D, and dialyzed against 2 x 100 volumes of Dignam's buffer D without PMSF. A standard 25 µl processing reaction, unless otherwise indicated, contained 0.1 pmol (45 ng) precursor RNA, 10 µl nuclear extract, 8 mM HEPES (pH 7.9), 8% (v/v) glycerol, 17 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$, 0.08 mM EDTA, 0.2 mM DTT, 4 mM creatine phosphate, 500 µM ATP and 2.5% polyvinyl alcohol. After 2.5 hours at 30 °C reactions were stopped by addition of 40 µg of proteinase K in 225 µl of 20 mM Tris pH 7.9, 0.1 M NaCl 10 mM EDTA and 1% SDS, followed by a 30 min incubation at 30 °C. 1/5 of the RNA extracted from a processing reaction was used for size analysis and 1/5 for S1 nuclease analysis. Size analysis of the RNA was performed by glyoxalation and 1.2% agarose gel electrophoresis (50). S1 nuclease analysis of the 3' ends of the processed RNA was carried out by the method of Berk and Sharp (51). A 524 bp SalI to StyI DNA fragment, which contains the poly(A) addition site fragment and 400 bp of downstream vector sequences, was 3' end labeled by filling recessed 3' ends with ^{32}P-labeled dNTPs and Klenow fragment of E. coli DNA polymerase I (New England BioLabs), and then strand separated (52). Hybridization was carried out for 12 hours at 37 °C in 50% formamide and S1 nuclease digestion was performed using 600 units/ml S1 nuclease (Sigma) for 30 min at 37 °C. Products were analyzed on 8% sequencing gels (53).

This may be due at least in part to the relative instability of the SV40 early pre-mRNA. In addition, we have observed different ionic requirements for 3' end processing of this precursor than have been observed for others (41). For these reasons, we have been unable to detect reproducibly the downstream cleavage product that is predicted if the cleavage reaction is endonucleolytic. However, we believe that endonucleolytic cleavage is involved in processing SV40 early pre-mRNA, as it is in others, and we will assume this is to be the case throughout this chapter.

From recent in vitro RNA processing data it has been suggested that the cap structure at the 5' end of the pre-mRNA molecule may be required for efficient splicing (27, 28) and 3' end formation (42). To address this question, we analyzed capped and uncapped pYSVHd3-A pre mRNA after processing in a nuclear extract. In addition, we tested the effects of cap analogs, which could compete for factors that recognize the 5' cap structure, on these reactions. Agarose gel electrophoresis of the processed RNA (Fig. 2A) shows that the uncapped RNA was significantly degraded, most likely by a 5' to 3' exonuclease present in the extract (4). Nonetheless, S1 analysis of the RNA shows that both uncapped and capped transcripts were efficiently cleaved at the poly(A) site, and that high levels of the cap analogs 7mGpppA or GpppA had no detectable effect on 3' end formation (Fig. 2B). From these results we conclude that a 5' cap structure is not required for 3' end formation.

The ATP analog cordycepin triphosphate (3'dATP) has been shown to inhibit polyadenylation in vivo and in vitro, most likely by being incorporated into the growing poly(A) tract, resulting in premature chain termination (43, 44). In order to test the effects of 3'dATP on cleavage and polyadenylation, we added it to processing reactions in the presence of ATP and creatine phosphate. Under these conditions polyadenylation was inhibited, as demonstrated by agarose gel electrophoresis of glyoxolated RNA purified from reactions incubated in the presence and absence of 3'dATP (Fig. 3A). S1 analysis of these RNA samples demonstrated that the cleavage reaction was inhibited

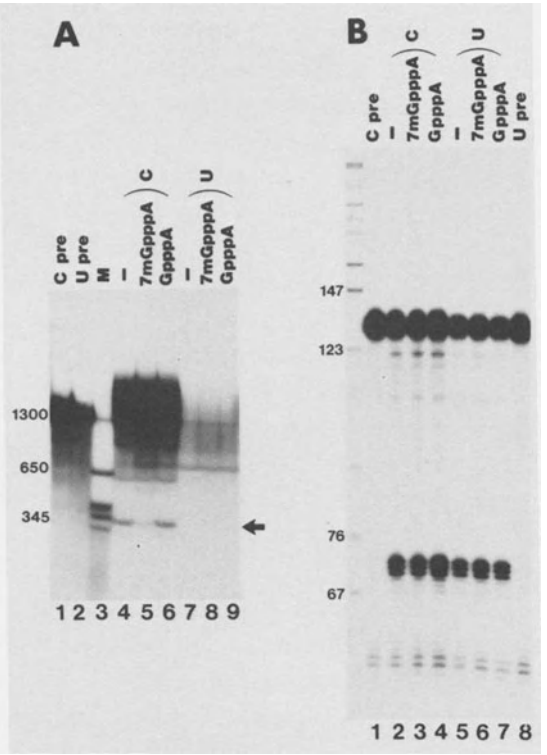


Figure 2. The effect of the 5' end cap structure on cleavage and polyadenylation. A, Size analysis: Lane 1, capped precursor RNA; lane 2, uncapped precursor RNA; lane 3, RNA size markers. Lanes 4-6, capped pre-RNA processed in the presence of: Lane 4, no analogs; lane5, 100 µM 7mGpppA; lane 6, 100 µM GpppA. Lanes 7-9, uncapped pre RNA processed in the presence of: Lane 7, no analogs; lane 8, 100 µM 7mGpppA; lane 9, 100 µM GpppA. The arrow points to the position of the large T antigen mRNA intron. B,S1 analysis of RNA from A. Uncapped RNA was prepared by the method in figure 1, except no cap structure was added to the reaction.

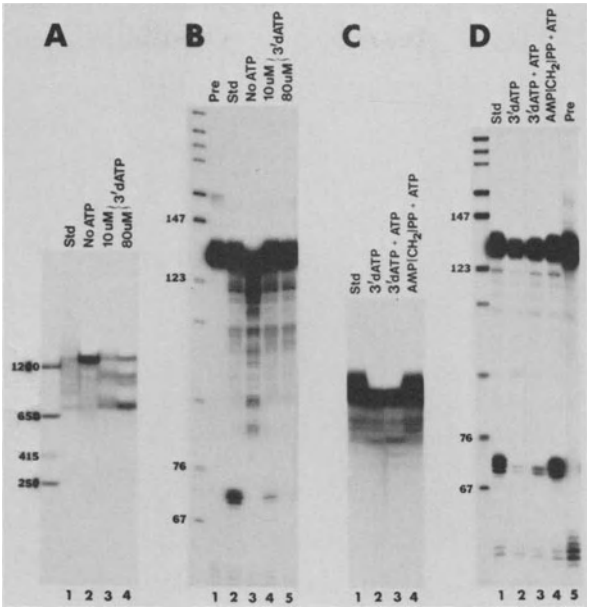


Figure 3. Inhibition of cleavage and polyadenylation by cordycepin triphosphate (3'dATP). A and B, ATP dependence of cleavage and polyadenylation in the presence of ATP. A, Polyadenylation obscures the presence of some splicing intermediates and products due to the heterogeneity in length of poly(A) tracts added to their 3' ends. Bands corresponding to these products become apparent when polyadenylation is inhibited. Size analysis: Lane 1, standard processing reaction; lane 2, pre RNA processed in the absence of ATP; lane 3, processed in presence of 10 μM 3' dATP and 500 μM ATP; lane 4, in the presence of 80 μM 3'dATP and 500 μM ATP. B,S1 analysis of the RNA from A. C and D, The effect of AMP(CH_2)PP in the presence of ATP on cleavage and polyadenylation as compared with 3'dATP, and the effect of 3'dATP in the absence of ATP. No creatine phosphate was added to these reactions. C, Size analysis: Lane 1, standard processing reaction; lane 2, 500 μM 3'dATP and no ATP; lane 3, 100 μM 3'dATP and 500 μM ATP; lane 4, 100 μM AMP(CH_2)PP and 500 μM ATP. D,S1 analysis of the RNA from C and precursor RNA in lane 5.

as well (Fig. 3B). When 3'dATP was substituted for ATP, rather than added in addition to ATP, low levels of cleavage were detected (Fig. 3D). In contrast, when another ATP analog, AMP(CH_2)PP, was added to processing reactions in the presence of ATP and creatine phosphate, no effect on cleavage or polyadenylation was observed (Fig. 3C and 3D).

The utility of micrococcal nuclease in RNA processing studies stems from its strict dependence on Ca^{2+} for activity. Incubation of an extract with micrococcal nuclease and Ca^{2+} is followed by addition of the chelating agent EGTA, which completely inactivates the nuclease. The nuclease treated extract can then be used to process precursor RNA. We have used this approach to determine whether a nuclease-sensitive component is required for mRNA cleavage and polyadenylation. A 30 minute micrococcal nuclease treatment of the nuclear extract was sufficient to degrade most of the abundant small RNAs, while treatment in the presence of EGTA had no detectable effect on these RNAs (Fig. 4). When pretreated extract was used to process pYSVHd3-A pre-mRNA, both polyadenylation (Fig. 5A) and 3' end cleavage (Fig. 5B and 5C) were inhibited. In addition, splicing of this transcript was completely abolished, as demonstrated by the absence of the large T-antigen mRNA intron (Fig. 5A) and by S1 analysis of the 3' splice site (data not shown).

To determine whether inactivation of the nuclear extract was caused by the degradation of a specific RNA-containing component, several fractions were added back to a nuclease-treated extract and RNA processing assays were performed. A cytoplasmic lysate from HeLa cells (S100) does not have detectable splicing activity (Fig. 5A; Ref. 1, 2) nor does it polyadenylate or cleave the pre-mRNA at its 3' end, even though it is contaminated with nuclear components such as snRNPs (data not shown). Addition of the S100 lysate to the nuclease-treated nuclear extract restored splicing and end polyadenylation (Fig. 5A), but did not restore cleavage activity, even when more than five times the amount of S100 required to restore splicing was added to the reaction (Fig. 5B). Purified *E. coli* RNA added to the micrococcal nuclease pretreated extract also

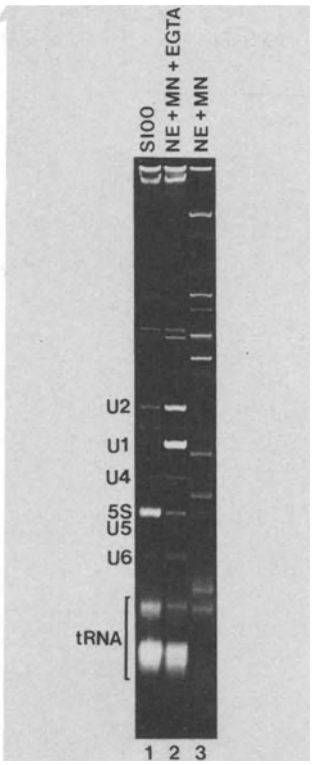


Figure 4. Micrococcal nuclease digestion of snRNAs. 7 M urea, 10% acrylamide gel electrophoresis of: lane 1, RNA extracted from 50 µl of a HeLa cell S100 lysate; lane 2, RNA extracted from 50 µl of a HeLa cell nuclear extract that was pretreated with micrococcal nuclease (MN) in the presence of EGTA; lane 3, the same as lane 2 except no EGTA was present in the incubation.

Methods: MN digestion was performed by mixing 5 µl of 5 mM $CaCl_2$ and 10µl of 15 unit/µl MN (Boehringer Mannheim) with 50 µl of HeLa cell nuclear extract and incubating at 30 °C for 30 min. The nuclease treatment was stopped by adding 2.5 µl of 20 mM EGTA. RNA was extracted by addition of 100 µg of proteinase K in 300 µl of 20 mM Tris pH 7.9, 0.1 M NaCl, 10 mM EDTA and 1% SDS, followed by a 30 min incubation at 30 °C. RNA was extracted with phenol:chloroform (1:1) and then ethanol precipitated. RNA was analyzed on a 7 M urea, 10% acrylamide:bis (27:1), 45 mM Tris borate pH 8.3, 1.25 mM EDTA gel.

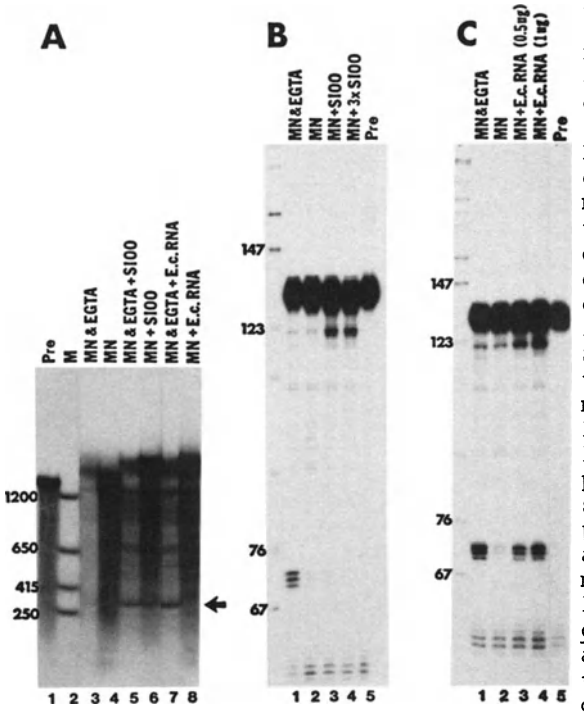


Figure 5. Micrococcal nuclease sensitivity of the cleavage and polyadenylation reaction. Size analysis: Lane 1, precursor RNA; lane 2, RNA size markers; lane 3, pre RNA processed in nuclear extract (NE) that was micrococcal nuclease (MN) treated in the presence of EGTA; lane 4, RNA processed in MN pretreated extract; lane 5, same as lane 3 except 2.5 µl of S100 lysate was added to the NE after MN pretreatment; lane 6, same as lane 4 except 7.5 µl S100 lysate was added after MN pretreatment; lane 7, same as lane 3 except 1 µg of E. coli RNA was added after MN pretreatment; lane 8, same as lane 7 except 1µg of E. coli RNA was added. The arrow points to the position of the large T antigen mRNA intron. B, S1 analysis: Lane 1, pre RNA processed in 7.5 µl NE nuclease pretreated in the presence of EGTA; lane 2, pre RNA processed in 7.5 µl pretreated NE; lane 3, same as lane 2 except 2.5 µl S100 lysate was added to the extract after MN pretreatment; lane 4, same as lane 3 except 7.5 µl S100 lysate was added; lane 5, precursor RNA. C,S1 analysis: Lane 1, pre RNA processed in NE pretreated with MN in the presence of EDTA; lane 2, pre-RNA processed in pretreated NE; lane 3, same as lane 2 except 0.5 µg E. coli RNA was added after pretreatment; lane 4, same as lane 2 except 1 µg E. coli RNA was added after MN pretreatment; lane 5, precursor RNA.

Methods: MN pretreatment was performed by mixing 1 µl of 5 mM $CaCl_2$ and 2 µl of 15 unit/µl MN (Boehringer Mannheim) with 10 or 7.5 µl NE and incubating at 30 °C for 30 min. The nuclease treatment was stopped by adding 2 µl of 5 mM EGTA. E. coli RNA was purified by the method of Salser et al. (54).

restored end polyadenylation, but had no effect on the lost splicing activity (Fig. 5A). Somewhat surprisingly, S1 analysis demonstrated that RNA processed in pretreated extract plus E. coli RNA was efficiently cleaved at the correct site (Fig. 5C). This finding suggests that the inhibition of cleavage and polyadenylation caused by micrococcal nuclease pretreatment was due to a general requirement for mass of RNA rather than to destruction of a required component.

DISCUSSION

The results presented here show that efficient 3' end cleavage, polyadenylation and splicing of an SV40 early pre-mRNA can occur under the same conditions in a HeLa cell nuclear extract. 3' end processing and splicing do not appear to be coupled in any way, since 3' end processing can be inhibited (with 3'dATP) without affecting the efficiency of splicing, and splicing can be inhibited (by the degradation of snRNPs) without affecting the efficiency of 3' end processing. Additional evidence, not presented here, also supports this conclusion. Different extract preparations have had different intrinsic abilities to splice or process 3' ends. For example, an extract preparation that had a high 3' end processing efficiency did not necessarily have a high splicing efficiency, and vice versa. In addition, a pre-RNA that could not be cleaved or polyadenylated, made by run-off transcription of pYSVHd3-A cleaved upstream of the sequences required for poly(A) addition, had the same splicing efficiency as a pre-RNA which was cleaved and polyadenylated (unpublished data). Taken together these results imply that 3'end processing and splicing are not interdependent in an in vitro processing reaction.

We have demonstrated that the ATP analog cordycepin triphosphate (3'dATP) inhibits both cleavage and polyadenylation in the presence of ATP, whereas AMP(CH_2)PP does not. Inhibition of polyadenylation is most likely due to incorporation of 3'dATP into the growing poly(A) tract. Inhibition of cleavage, however, may be interpreted in several ways. One is that cleavage and polyadenylation are obligatorily coupled such that

if polyadenylation is inhibited, cleavage is concomitantly inhibited. This is probably not the case because Moore and Sharp have demonstrated that the cleavage and polyadenylation reactions can be uncoupled by substituting AMP(CH_2)PP for ATP, which inhibits polyadenylation but not 3' end cleavage (10). Another interpretation is that the cleavage reaction requires ATP as an allosteric activator. 3'dATP, in competition with ATP, does not mediate this activation, whereas AMP(CH_2)PP does. An interesting possibility is that this hypothetical allosteric activation may be mediated through the poly(A) polymerase.

A 5' cap is apparently not required for mRNA 3' end processing *in vitro*, because the presence of a 5' cap on the pre-RNA had no effect on the efficiency of 3' end processing, other than to stabilize the pre-RNA. This lack of a 5' cap requirement *in vitro* is consistent with results obtained *in vivo*. A transcript transcribed by RNA polymerase III, which should not contain a 5' cap, was found to be accurately and efficiently cleaved and polyadenylated *in vivo* (29). Recently, Hart *et al.* (42) reported the results of *in vitro* experiments in which they found the presence of a 5' cap enhances 3' end processing and that addition of 7mGpppG cap analog to reactions greatly reduced 3' end processing. This discrepency with the results presented here may arise from difficulties in the interpretation of the data resulting from the prime extension assay used by Hart *et al.*. This assay detects only correctly cleaved and polyadenylated pre mRNA; the uncleaved precursor that remained at the end of the reaction was not detected. Therefore, a direct measurement of the efficiency of cleavage/polyadenylation could not be made. Furthermore, the efficiency of the reaction was low, making the assay very sensitive to small fluctuations in the stability of the pre-mRNA.

Micrococcal nuclease (MN) pretreatment of the nuclear extract inhibited splicing, 3' cleavage and polyadenylation. However, restoration of cleavage and polyadenylation (but not splicing) was achieved by addition of purified *E. coli* RNA, suggesting that inhibition caused by MN pretreatment was due to a general requirement for mass of RNA, rather than to the

destruction of required components(s). This idea is consistent with our previous results, which showed that end-polyadenylation in a HeLa whole-cell extract was much less efficient when the concentration of added RNA was reduced (3). Similarly, the efficiency of DNA-directed transcription systems is strikingly dependent on the DNA template concentration (45). This is thought to reflect a nonspecific and inhibitory binding of proteins (probably histones) to the DNA template such that transcription only occurs above a certain threshold DNA concentration. We speculate that the MN pretreatment inhibits cleavage and polyadenylation by degrading endogenous nucleic acids, which can function as sinks for nonspecific binding proteins. Additional nucleic acid binding proteins, perhaps including snRNP proteins (46), may also be released by MN degradation. When a small amount of precursor RNA is added to the pretreated extract these proteins may bind to the RNA, probably nonspecifically, thereby preventing access to processing enzymes and inhibiting cleavage and polyadenylation. Splicing is apparently not inhibited in this way, possibly due to the rapid and stable interaction of the abundant snRNPs, U1 and U2, with the pre RNA (30, 31).

In the last few years evidence has accumulated which suggests that a snRNP may be involved in mRNA 3' end processing (4, 10, 36, 47). It was therefore somewhat surprising to find that the efficiency and accuracy of 3' end processing was not reduced in a nuclear extract in which all of the detectable snRNAs were at least partially degraded. Several explanations for this result can be made: 1) a snRNP could have retained activity even if its RNA moiety had been partially degraded. A possible precedent for this idea exists in pre-mRNA splicing, when a snRNP (perhaps U5) was shown to retain the ablity to bind to a 3' splice site after its RNA component had been partially degraded by MN (32). It is not clear, however, whether this "crippled" snRNP was still functional. 2) An unidentified snRNP that was resistant to MN may be required. 3) A small quantity of one of the abundant snRNPs remained undigested after MN treatment, and this amount was sufficient

for efficient mRNA 3' end formation. 4) A snRNP is not required in the formation of polyadenylated mRNA 3' ends. Resolution of this issue will require identification and characterization of the required factors.

Another unresolved question is whether separate enzymes catalyze the cleavage and polyadenylation reactions. One clue may come from observations that both reactions require and intact AAUAAA consensus sequence in the pre mRNA. Thus polyadenylation that occurs in the absence of cleavage, at artificially created RNA 3' ends within 20-400 nt downstream of the authentic cleavage site (3, 4), or at the exact point of in vivo polyadenylation (18), requires an AAUAAA signal sequence. Conversely, 3' end cleavage that occurs when polyadenylation is blocked also requires this sequence (18). While several interpretations of these findings are possible, one is that the poly(A) polymerizing and cleaving activities are present in one enzyme, or enzyme complex, and thereby recognize and interact with the pre-mRNA simultaneously. The findings presented here that cordycepin triphosphate, long known to be an inhibitor of poly(A) polymerases, can in addition block cleavage of an SV40 early pre-mRNA, is also suggestive of a close association between these activities.

ACKNOWLEDGMENTS

We thank J. Noble for advice and S1 analysis of the 3' splice site, H. Ge and W. Erhman for expert technical assistance, and K. Berg for E. coli RNA. This work was supported by U.S. Public Health Service grant GM 28983 from the National Institutes of Health.

REFERENCES

1. Hernandez, N. and W. Keller, Cell 35: 88-99, 1983.
2. Krainer, A.R., T. Maniatis, B. Ruskin and M.R. Green, Cell 36: 993-1005, 1984.
3. Manley, J.L., Cell 33: 595-605, 1984.
4. Moore, C.L. and P.A. Sharp, Cell 36: 581-591, 1984.
5. Padgett, R.A., S.F. Hardy and P.A. Sharp., Proc. Natl. Acad. Sci. U.S.A. 80: 5230-4234, 1983.
6. Price, D.H. and C.S. Parker, Cell 38: 423-429, 1984.
7. Ford, J.P. and M. -T. Hsu. J. Virol. 28: 795-801, 1978.
8. Frayne, E.G., E.J. Leys, G.F. Crouse, A.G. Hook and R.E. Kellems, Mol. Cell. Biol. 4: 2921-2924, 1984.
9. Manley, J.L., P.A. Sharp and M.L. Gefter, J. Mol. Biol. 159: 581-600, 1982.
10. Moore, C.L. and P.A. Sharp, Cell 41: 845-855, 1985.
11. Nevins, J.R. and J.E. Darnell Jr. Cell 15: 1477-1487.
12. Proudfoot, N.J. and G.G. Brownlee, Nature 263: 211-214, 1976.
13. Fitzgerald, M. and T. Shenk, Cell 24: 251-260, 1981.
14. Higgs, D.R., S.E. Goodbourn, J. Lamb and N.C. Proudfoot, Nature 306: 398-400, 1983.
15. Manley, J.L., H. Yu and L. Ryner, Mol. Cell. Biol. 5: 373-379, 1985.
16. Montell, C., E.F. Fisher, M.H. Caruthers and A.J. Berk, Nature 305: 600-605, 1983.
17. Orkin, S.H., T.-C Chen, S.E. Antonarkis and H.H. Kazarian Jr, EMBO J. 4: 453-456, 1985.
18. Wickens, M. and P. Stephenson, Science 226: 1045-1051, 1984.
19. Zarkower, D., R. Stephenson, M. Sheets and M. Wickens, Mol. Cell. Biol. 6: 2317-2323, in press, 1986.
20. Conway, L. and M.P. Wickens, Proc. Natl. Acad. Sci. U.S.A. 829: 3949-3953, 1985.
21. Gil, A. and N.J. Proudfoot, Nature 312: 473-475, 1984.
22. McDevitt, M.A., M.J. Imperiale, H. Ali and J.R. Nevins, Cell 37: 993-999, 1984.
23. Mclaughlan , J., D. Gaffney, J.L. Whitton and J.B. Clements, Nucl. Acids Res. 13: 1347-1368, 1985.
24. Sadofsky, M., S. Connelly, J.L. Manley and J.C. Alwine, Mol. Cell. Biol. 5: 2613-2519, 1985.
25. Simonsen, C.C. and Levinson, A.D., Mol. Cell. Biol. 3: 2250-2258, 1983.
26. Woychik, R.P., R.H. Lyons, L. Post and F.M. Rottman, Proc. Natl. Acad. Sci. U.S.A. 81: 3944-3948, 1984.
27. Konarska, M.M., R.A. Padgett and P.A. Sharp. Cell 38: 731-736, 1984.
28. Edery, I. and N. Sonenberg, Proc. Natl. Acad. Sci. U.S.A. 82: 7590-7594, 1985.
29. Lewis, E.D. and J.L. Manley, Proc. Natl. Acad. Sci. U.S.A., in press, 1986.
30. Krainer, A.R. and T. Maniatis, Cell 42: 725-736, 1985.

31. Black, D.L., B. Chabot and J.A. Steitz, Cell 42: 737-750 1985.
32. Chabot, B., D.L. Black, D.M. Lemaster and J.A. Steitz. Science 230:1344-1349, 1985.
33. Berget, S.M. and B.L. Robberson, Cell 46: 691-696, 1986.
34. Black, D.L. and J.A. Steitz, Cell 46: 697-704, 1986.
35. Strub, K., G. Galli, M. Busslinger and M.L. Birnstiel, EMBO J. 3: 2801-2807, 1984.
36. Hashimoto, C. and J.A. Steitz, Cell 45: 581-591, 1986.
37. Tooze, J. (ed.), DNA Tumor Viruses, 2nd edition, Cold Spring Harbor Laboratory, Cold Spring harbor, N.Y., 1981.
38. Keller, W., Cell 39: 423-425, 1984.
39. Birnstiel, M.L., M. Busslinger, and K. Strub. Cell 41: 349-359, 1985.
40. Noble, J.C.S., C. Prives, and J.L. Manley, Nucl. Acids Res. 14: 1219-1236, 1986.
41. Ryner, L.C. and J.L. Manley, Mol. Cell. Biol., in press.
42. Hart, R.P., M.A. McDevitt, and J.R. Nevins, Cell 43: 677-683, 1985.
43. Edmonds, M., The Enzymes Vol. XV, 217-244, 1982.
44. Jacob, S.T. and K.M. Rose, In: Enzymes of Nucleic Acid Synthesis and Modification, Vol. 2, (S.T. Ed. Jacob, Boca Raton, Florida: CRC Press), 1985, pp. 135-157.
45. Manley, J.L., A. Fire, A. Cano, P.A. Sharp and M. Gefter, Proc. Natl. Acad. Sci. U.S.A. 77: 3855-3859, 1980.
46. Mount, S.M., I. Pettersson, M. Hinterberger, A. Karmas, and J.A. Steitz, Cell 33: 509-518, 1983.
47. Berget, S.M., Nature 309: 179-182, 1984.
48. Rossi, J.J., X. Soberon, Y. Marumoto, Y. McMahon and K. Itakura, Proc. Natl. Acad. Sci. U.S.A. 80: 3203-3207, 1983.
49. Dignam, J.D., R.N. Lebovitz and R.G. Roeder, Nucl. Acids Res. 11: 1475-1489, 1983.
50. McMasters, G. and G. Carmichael, Proc. Natl. Acad. Sci. U.S.A. 79: 4835-4838, 1977.
51. Berk, A.J. and P.A. Sharp, Proc. Natl. Acad. Sci. U.S.A. 75: 1274-1278, 1978.
52. Maniatis, T., E.F. Fritsch and J. Sambrook, In: Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982, pp. 113-114.
53. Maxam, A.M. and W. Gilbert, Meth. Enzymol. 65: 499-560, 1980.
54. Salser, W., R.F. Gesteland, and A. Bolle, Nature 215: 588-591, 1967.

6

REGULATION OF VIRAL TRANSCRIPTION UNITS BY SV40 T-ANTIGEN

J. BRADY, M. LOEKEN, M.A. THOMPSON, J. DUVALL and G. KHOURY

Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

SUMMARY

We have investigated the ability of SV40 T-antigen to trans-activate the SV40 late and the Adenovirus E2 promoters. Transcriptional control signals required for T-antigen trans-activation of the SV40 late promoter include T-antigen binding site II and the SV40 72-bp repeats. In vivo competition with recombinant plasmids containing the entire SV40 late regulatory region and promoter sequences (mp 5171-272) results in quantitative removal of limiting trans-acting factors required for late gene expression in COS-1 cells. Insertion of increasing lengths of DNA sequences between the T-antigen binding sites and the 72-bp repeats dramatically reduces the competition efficiency, suggesting a physical interaction between proteins binding to the separate regulatory domains. Transfection experiments have been performed in ts2 COS cells, which express the ts 1609 SV40 T-antigen. Transfection at the non-permissive temperature (40°C) resulted in a 5- to 10-fold reduction in SV40 late promoter activity compared to the permissive temperature (33°C), suggesting that trans-activation of the SV40 late promoter requires continued expression of T-antigen.

SV40 T-antigen trans-activates the Ad E2 promoter as effectively as does the Ad E1A protein. While 79 bp of upstream sequences are required for basal, E1A or T-antigen stimulated E2 promoter function, our experiments indicate that cellular factors which mediate stimulation by T-antigen and E1A are different. The sequences between -75 and -30 contain two imperfect 14 bp repeats separated by 16 bp. Using chemically synthesized DNA fragments containing the inverted repeat, we demonstrate that E1A efficiently induces transcriptional activity

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

when these sequences are inserted in either orientation upstream of a heterologous promoter. Similar results were obtained using SV40 T-antigen. This suggests that target sequences for both E1A and T-antigen trans-activation of the E2 promoter are located between -85 and -29 and function in an orientation-independent fashion.

INTRODUCTION

Simian Virus 40 (SV40) and the other small DNA viruses have served as model systems for the identification of cis-dependent control sequences as well as the mechanisms by which trans-acting regulatory proteins modulate expression of eukaryotic transcriptional units. We have been particularly interested in SV40 late transcription control because, unlike the SV40 early regulatory unit, the SV40 late transcription unit functions inefficiently until late in the lytic cycle of the virus in monkey kidney cells (1). In nonpermissive cells, virus infection leads to a significant level of early gene expression, but transcription from the late promoter is weak or absent. These observations are of particular interest since the SV40 late transcription unit lacks some of the more common transcriptional control sequences of a typical polymerase II transcription unit. For instance, there is no well-defined TATA sequence upstream of the major late transcriptional initiation site. In addition, even though the SV40 early enhancer (72-bp repeat) works with either the early promoter or heterologous promoters in a manner which is relatively independent of orientation and position (2,3,4), this sequence is unable to induce late transcription immediately after introduction of the gene into permissive or nonpermissive cells. In view of these unusual transcriptional properties, we have been interested in identifying mechanisms by which the late transcription unit is activated or positively regulated.

Like the SV40 late promoter, the Adenovirus E2 promoter is distinguished from other eukaryotic promoters by the absence of a classical TATA sequence (5). The E2 promoter is not transcribed efficiently in the absence of specific viral proteins. Several studies have shown that the Ad E2 promoter is positively regulated by the product of the 13S Ad E1A mRNA (6,7,8). The SV40 early gene product, T-antigen, and the pseudorabies virus immediate-early gene also trans-

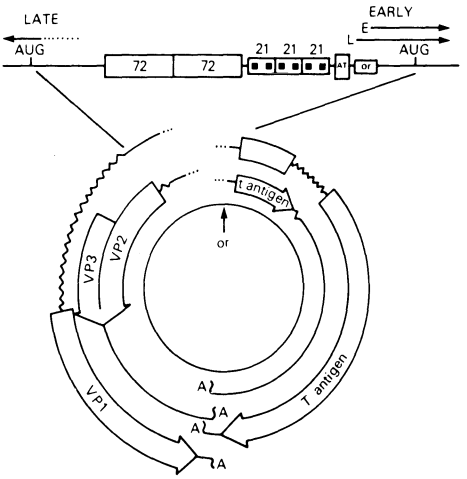


Figure 1. The genomic map and control region of SV40. The diagram presents the control region for expression of the SV40 early genes (large and small T-antigens) and late genes (VP1, 2 and 3). The origin for viral DNA replication (or) is flanked by the early transcriptional control sequences, including the Goldberg-Hogness box (AT), the three 21-bp repeats (each containing two copies of a GC-rich hexanucleotide), and the tandem 72-bp enhancer element. The early transcripts are initiated predominantly at position E early in infection and shift to position L after DNA replication. The late viral transcripts have heterogeneous 5′ ends, indicated by dots. (Reprinted, with permission, from Hamer and Khoury, ref. 26).

activates the E2 promoter (9,10,11). There is no evidence for direct binding of E1A, SV40 T-antigen or the immediate early proteins to the E2 promoter, suggesting that activation is mediated in part by activation of cellular factors which bind to the E2 promoter or enhancer-like region (12). It is not known whether the trans-acting proteins from adenovirus, SV40 or pseudorabies virus interact with the same transcription factor.

RESULTS

SV40 Late Gene Expression After Transfection of CV-1 and COS-1 Cells with SV40 DNA. The efficiency of SV40 late gene expression (Fig. 1) was determined after transfection of CV-1 and COS-1 cells with purified form I SV40 DNA (13). To exclude template amplification,

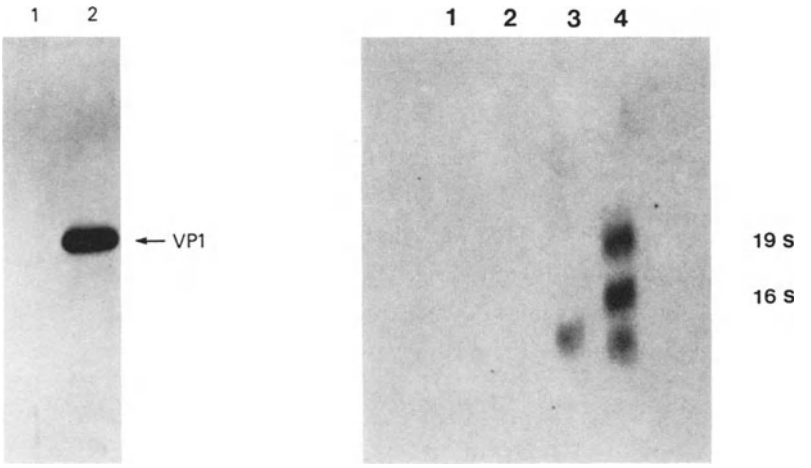


Figure 2. SV40 late gene expression in CV-1 (lane 1) and COS-1 (lane 2) cells after transfection with SV40 DNA. Parallel cultures of CV-1 and COS-1 cells (10-cm plate) (≈60-70% confluent) were transfected with SV40 DNA (2 μg) by the calcium phosphate precipitation method. After transfection, cells were maintained in Dulbecco's minimal essential medium with fetal calf serum (10%) and cytosine arabinoside (25 μg/ml). A) At 30 hr after transfection, whole-cell protein extracts were prepared, and 40 μl was analyzed by immunoblot analysis. The arrow indicates the position of migration of control SV40 VP-1 protein. B) At 30 hr post-transfection, total cellular RNA was isolated by hot acid phenol extraction. Twenty μg of RNA was analyzed by Northern blot analysis using a probe specific for the late region of SV40.

cytosine arabinoside (25 μg/ml) was added to the culture media. After transfection of CV-1 and COS-1 cells, whole-cell extracts were prepared and analyzed by immunoblot analysis with anti-SV40 VP-1 antisera. The level of SV40 late gene expression 30 hr after DNA transfection was dramatically enhanced in COS-1 cells (Fig. 2A, lane 2) compared to control CV-1 cells (Fig. 2A, lane 1). We estimate that VP-1 synthesis was increased by at least 20- to 50-fold in COS-1 cells. No SV40 DNA replication could be detected in the presence of cytosine arabinoside, indicating that enhanced late gene expression is not due to template amplification.

In addition to analysis of the late gene product, VP-1, we analyzed the level of SV40 late mRNA synthesized in CV-1 and COS-1 cells after transfection with SV40 DNA in the presence of cytosine arabinoside. Northern blot hybridization analysis showed that the level of SV40 late mRNA synthesis is dramatically increased in COS-1 cells (Fig. 2B, lane 4). Therefore, we conclude that SV40 T-antigen regulates late gene expression at the transcriptional level.

Studies in our laboratory using deletion and point mutants have defined two important domains for the SV40 T-antigen induced late gene expression (13, 14). One of the regions important for trans-activation includes T-antigen binding sites I and/or II. Deletion of T-antigen binding site I and one-half of site II as in mutant pSVsL18, results in a late trans-activation induction level which is decreased to approximately 5-10% of that observed with the wild-type template. Similarly, deletion of 4 or 6 bp in T-antigen binding site II dramatically reduces the efficiency of SV40 late gene expression. The other regulatory sequence required for efficient late gene activation is located in the SV40 72-bp repeats. Using a set of clustered point mutants which significantly affect early enhancer function (15) and are located in a single retained copy of the 72-bp repeats, we have found a parallel decrease in late gene expression, i.e., the sequences within the 72-bp repeats that serve as the SV40 early enhancer core element are apparently also important for late gene induction by T-antigen in the absence of DNA replication. In contrast to the effect of mutations within the SV40 T-antigen binding sites and the 72-bp repeats, mutations approximately 25-bp upstream of the major late transcriptional initiation site which increase or decrease homology to the conserved eukaryotic TATA sequence, did not affect the T-antigen mediated trans-activation of the late transcription unit (16).

Binding of Trans-Acting Factors is Dependent on Spacing Between T-Antigen Binding Sites I and II and the 72-bp Enhancer Element. We next wanted to determine if the transcriptional control sequences represent binding sites for trans-acting factors. To address this point and to determine how the two upstream control elements interact to facilitate T-antigen induced trans-activation, we have used a template competition analysis (14, 16). In the presence of a fixed

amount of template and increasing levels of cloned competitor DNA fragments, which included promoter regions found to be important in cis for trans-activation of the SV40 late promoter and thus are capable of binding limiting trans-acting factors in the COS-1 cell, a decrease in SV40 late gene expression was observed. Efficient competition for trans-acting factors occurs only when the regions representing the SV40 T-antigen binding sites and the 72-bp repeats are linked on the same competitor molecule. These findings suggest that efficient binding of the trans-acting factors requires interaction between one or more protein(s) and the two transcriptional domains.

To test this hypothesis, we have analyzed the effect of increasing the distance between the two transcriptional domains on competition efficiency. A series of mutants were recently generated by Innis and Scott (17) by the insertion of DNA sequences at the SV40 NcoI site (mp 37). Using standard recombinant DNA techniques, we transferred a BglI/SphI DNA fragment (mp 0-128) from each mutant to a plasmid containing SV40 sequences from the HindIII site (mp 5171) to the PvuII site (mp 272), thus reconstituting the spacer mutants in plasmids containing the entire control region (Fig. 3). Competition with plasmid pJI-1, which contained no insertion of DNA sequences between the two transcriptional domains, resulted in a quantitative decrease in SV40 late gene expression (lanes 2 and 3). Insertion of 4 bp of DNA at the NcoI site had a minimal effect on the competition efficiency, reducing it approximately 20% (lanes 4 and 5). In contrast, pJI-42 (lanes 6 and 7), pJI-90 (lanes 8 and 9), or pJI-260 (lanes 10 and 11) were ineffective competitors, producing little or no decrease in the level of late gene expression. These results suggest that efficient binding of the trans-acting transcription factors requires a precise physical relationship between both sets of DNA sequences and the putative transcriptional factors with which they interact.

Continued Expression of SV40 T-Antigen is Required for Activation of the Late Transcription Unit. The experiments presented above do not determine whether the continued expression of SV40 T-antigen is required for stimulation of SV40 late gene expression. For example, it is conceivable that SV40 T-antigen either modifies existing transcription factors or induces the synthesis of a particular transcription factor

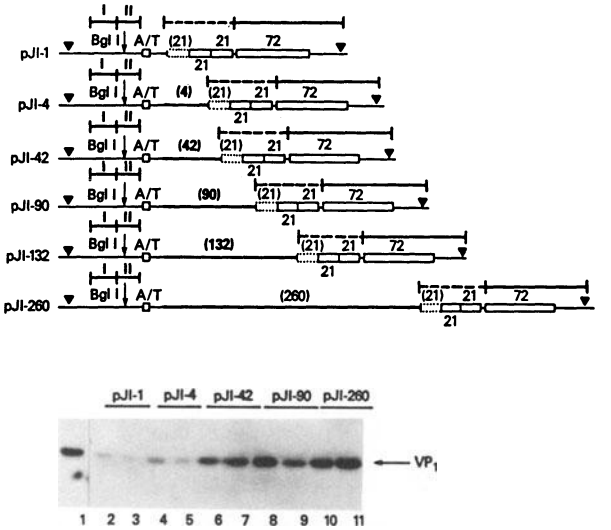


Figure 3. In vivo competition for COS-1 trans-acting factor with plasmids containing insertions between T-antigen binding sites I and II and the 72-bp repeats. COS-1 cells were transfected with template SV40 DNA (0.1 μg) and competition plasmids (1.0 μg). Transfected cultures were maintained and late gene expression was assayed as described in Fig. 2. Lanes: 1. 0.1 μg of SV40 DNA template; 2 and 3, 0.1 μg of SV40 template DNA and 1.0 μg of pJI-1 competition plasmid; 4 and 5, 0.1 μg of SV40 template DNA and 1.0 μg of pJI-4 competition plasmid; 6 and 7, 0.1 μg of SV40 template DNA and 1.0 μg of pJI-42 competition plasmid; 8 and 9, 0.1 μg of SV40 template DNA and 1.0 μg of pJI-90 competition plasmid; 10 and 11, 0.1 μg of SV40 template DNA and 1.0 μg of competition plasmid pJI-260.

required for late gene expression. Subsequently, T-antigen might not be required if the activation were through an indirect mechanism. Alternatively, if T-antigen is involved directly in SV40 late gene activation, the continuous expression of an active T-antigen would be required. To test these alternatives, we have utilized a transformed CV-1 cell line which contains a temperature sensitive SV40 T-antigen (ts 1609) (18).

Temperature-sensitive COS-1 cells were maintained at the permissive temperature of 33°C prior to transfection. A plasmid containing the SV40 late promoter (mp 5171-346) upstream of the coding sequences for the bacterial enzyme chloramphenicol acetyltransferase (CAT) was transfected by the calcium phosphate precipitation technique. Following a

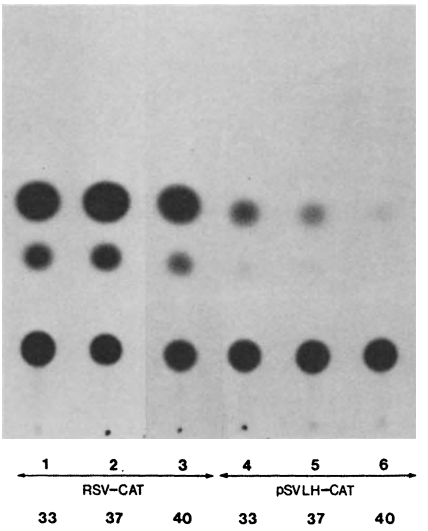


Figure 4. CAT assay of trans-activation of the SV40 late promoter in ts COS cells. The SV40 late construct, pSVLH-CAT contains the HindIII-HpaII (5171-346) SV40 sequence inserted in the late direction in front of the chloramphenicol acetyltransferase gene coding sequence. ts COS cells (35 mm plate) were transfected by the calcium phosphate precipitation method. Total DNA transfected was maintained constant at 10 μg by addition of the carrier plasmid p3MdlCAT. Cells were maintained at 33°C for 12 hr and then washed. The medium was replaced with medium containing cytosine arabinoside (25 μg/ml), and plates were transferred to the indicated temperature. Cells were harvested 48 hr after transfection. Cell extracts and assays for chloramphenicol acetyltransferase activity were performed as described previously (11).

14 hr incubation at 33°C, the cell monolayer was washed and incubation continued for 34 hr at either 33°C, 37°C or 40°C. As a metabolic control, RNA synthesis from a T-antigen independent promoter was assayed following transfection of duplicate monolayers with the plasmid pRSV-CAT which contains the Rous sarcoma virus LTR upstream of the CAT coding sequences. The results of a typical transfection assay are shown in Fig. 4. The level of SV40 late gene expression decreased dramatically as the ts COS-1 cell line was shifted to the non-permissive temperature. At the permissive temperature of 33°C, a 10.3% conversion of the ^{14}C-chloramphenicol to the acetylated derivative was observed. In contrast, only 1.3% ^{14}C-chloramphenicol conversion was observed at 40°C. Expression from the control pRSV-CAT plasmid was not significantly

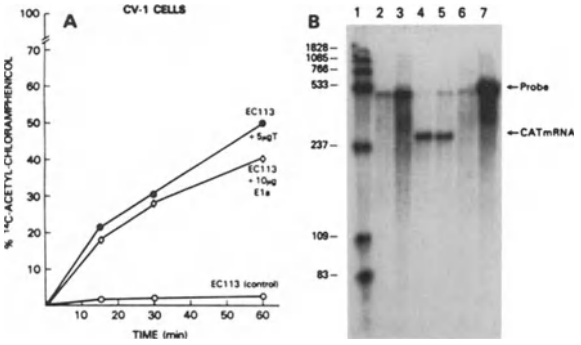


Figure 5. SV40 T-antigen and Ad E1A stimulate transcription from the E2 promotor in CV-1 cells. A) CV-1 cells were transfected with 5 μg of pEC113 alone (o) or with 10 μg of pE1A (◇) or 5 μg of pRSV-T (●). At 48 hr, extracts were prepared, and CAT enzyme activity was determined. B) CAT mRNA levels produced in response to T-antigen or E1A. Samples (20 μg) of whole cell RNA were probed in S1 nuclease protection analysis for CAT mRNA as described in Materials and Methods. Lanes: 1, molecular weight markers; 2, mock transfection; 3, transfection with 15 μg of pEC113 alone; 4, transfection with 15 μg of pEC113 plus 10 μg of pE1A; 5, transfection with 15 μg pEC113 plus 5 μg of pRSV-T; 6, probe alone digested with S1 nuclease; 7, intact probe. The positions of the intact probe and the DNA fragment protected by CAT mRNA are shown.

altered by the change in incubation temperature. These results indicate that the continued expression of SV40 T-antigen is required for stimulation of SV40 late gene expression. Thus, either T-antigen serves as a transcription factor and is directly involved in the SV40 late gene activation, or it continually induces or modifies a labile transcription factor.

SV40 T-Antigen Trans-Activates the E2 Promoter. In a separate set of experiments, we have analyzed the ability of SV40 T-antigen to stimulate transcription from the Adenovirus E2 promoter. We first compared the activity of the Ad E2 promoter (pEC113) in the presence or absence of trans-acting viral proteins (11). The pEC113 plasmid (19), in which CAT gene expression is under the control of the E2 promoter containing 285 bp upstream and 40 bp downstream (-285 to +40 bp) of the

mRNA cap site, was transfected into CV-1 cells alone or with plasmids encoding SV40 T-antigen (pRSV-T) or Ad E1A (pE1A). Extracts examined at 48 hr after transfection with pEC113 alone contained a low, but detectable, level of CAT enzyme activity during a 60 min reaction (Fig. 5A). In contrast, cotransfection of pEC113 with plasmids encoding either SV40 T-antigen or E1A resulted in a 14- to 17-fold enhancement of CAT enzyme levels. An even greater enhancement of basal CAT levels was observed when optimum concentrations of T-antigen or E1A were present (data not shown). Similar effects of T-antigen and E1A on pEC113 CAT activity were observed in HeLa cells (data not shown).

To establish that the increase in E2 CAT activity was due to elevated levels of CAT mRNA, we examined steady-state E2 CAT mRNA by quantitative S1 nuclease analysis (Fig. 5B). The probe used in the S1 analysis was synthesized from a single-stranded M13 recombinant plasmid. The size of the undigested probe was 458 nucleotides, whereas the size of a fragment protected from S1 nuclease digestion by CAT mRNA would be 256 nucleotides. The probe did not include the 37 bases homologous to the E2 sequence between the E2 mRNA cap site and the start of the CAT-coding sequence. Thus, this analysis measures the level of total CAT mRNA, but not the 5′ end. CAT mRNA was not detected in mock-transfected cultures or in cultures transfected with pEC113 alone (Fig. 5B, lane 2 or 3, respectively). An S1-protected DNA fragment of 256 bases was detected in cultures transfected with pEC113 plus either pE1A (Fig. 5B, lane 4) or pRSV-T (Fig. 5B, lane 5). The probe alone treated with S1 nuclease is shown in Fig. 5B, lane 6. These results demonstrated that SV40 T-antigen and Adenovirus E1A stimulate transcription from the E2 promoter in CV-1 cells.

Additive Effect of T-Antigen and E1A Suggests Different Mechanisms of Stimulation of the E2 Promoter. When 10 μg of pEC113 and optimum amounts of pRSV-T (5 μg) and pE1A (10 μg) were transfected together, the CAT enzyme levels observed were approximately the sum of those obtained at saturating concentrations of either *trans*-acting protein alone (Fig. 6). Based on earlier titration studies, at these plasmid concentrations the limiting factor for *trans*-activation by either T-antigen or E1A is likely to be a cellular transcription factor. The fact that T-antigen and E1A stimulate the E2 promoter in an additive

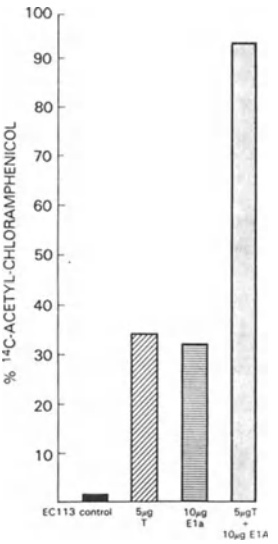


Figure 6. Effect of SV40 T-antigen and Ad E1A together on the E2 promoter in CV-1 cells. A limiting amount of pEC113 (10 μg) was transfected alone; with 5 μg of pRSV-T; 10 μg of pE1A; or 5 μg of pRSV-T plus 10 μg of pE1A. In all cases, the final amount of DNA was brought to 35 μg with pML2 DNA. Extracts were prepared 48 hr post-transfection and analyzed for CAT enzyme activity during a 60 min reaction.

manner suggests that the trans-acting proteins interact with different cellular transcription factors to activate the E2 promoter. If the cellular factors which mediated the effects of T-antigen and E1A were identical, cotransfection of E1A and T-antigen would not have increased the E2 promoter activity above that observed with saturating amounts of either T-antigen or E1A independently.

The additive effect of T-antigen and E1A on E2 transcription is not likely to be due to elevated concentrations of T-antigen or E1A as a result of stimulation by one viral protein on the production of the other. When plasmids were cotransfected the levels of T-antigen and E1A mRNA, as determined by Northern blot analysis, were within 2-fold of the levels obtained upon transfection of either plasmid alone. In addition, mRNA levels obtained following cotransfection were not as high as mRNA levels obtained upon transfection of higher concentrations of T-antigen or E1A alone, which did not further elevate E2 transcription.

Analysis of E2 Trans-Activation Using Synthetic Oligomers. We have found that sequences downstream of -79 in the E2 promoter are required for efficient trans-activation by T-antigen (11) which is in agreement with sequences required for efficient basal or E1A trans-activated promoter function (20,21). Within this minimal promoter region is an almost perfect 14 bp inverted repeat between -75 to -60 and -43 to -30. In addition, a transcriptional regulatory sequence (-82 to -66) has been identified by linker-scanning mutational analysis. Unlike many eukaryotic RNA polymerase II transcription units, no canonical "TATA"-box is present. A "pseudo-'TATA'"-box, which appears to be important for the major transcription start site is contained between -28 to -21 (21). We were interested in determining if T-antigen and E1A responsive regulatory sequences were located exclusively in the upstream regulatory sequences or whether the activation also required downstream promoter sequences. We constructed a chemically synthesized E2 promoter (-85 to -29) and inserted it in either orientation into a vector containing a heterologous polymerase II promoter upstream of the assayable CAT gene.

When CAT activity from the plasmid containing E2 promoter sequences -85 to -29, in either the sense or antisense orientation, and a heterologous promoter was compared to that from a plasmid containing -97 to +40 of the E2 promoter (pEC-97), the fold trans-activation by T-antigen and E1A was similar (Table 1). This suggests that both T-antigen and E1A trans-activation of the E2 promoter is primarily dependent upon sequences between -85 to -29, but not specifically the E2 downstream promoter element. Furthermore, the E2 regulatory sequences between -85 to -29 function in an orientation-independent manner.

Table 1: E2 Trans-Activation Using Synthetic Oligomers.

Promoter Sequences		CAT Activity (%)		
Upstream	Downstream	Basal	+E1A	+SV40 T
E2 (pEC-97) (-97 to -29)	Homologous (-29 to +40)	9	73	85
E2 (-85 to -29) (S)[1]	Heterologous	7	64	54
E2 (-29 to -89) (AS)[2]	Heterologous	1	18	17

1. Sense orientation.
2. Antisense orientation.

DISCUSSION

We have shown that at least two upstream control regions are required for efficient trans-activation of the SV40 late promoter by T-antigen (13,14,16). The first sequence consists of T-antigen binding sites I and II; the second sequence is located in the SV40 72-bp tandem repeats. Deletion of either transcriptional control sequence from the template decreases late gene expression by an order of magnitude. The in vivo competition assays support these data and contribute additional information towards an understanding of the potential interaction of the upstream sequences with putative regulatory molecules. The ability to compete for transcriptional factors suggests the direct interaction of the factors with the control sequences. Efficient binding of the limiting transcriptional factor requires the presence of the two transcriptional control sequences in cis and at a critical distance from one another. These results suggest that binding of the limiting transcription factor(s) requires a cooperative interaction between the protein(s) and DNA sequences located within these domains.

The mechanism by which the upstream transcriptional control sequences activate SV40 late transcription is still unclear. An obvious possibility is the direct interaction of T-antigen with T-antigen binding sites I and II. While this explanation is consistent with our previous studies, which showed a correlation between the ability of mutant SV40 T-antigens to bind the origin and the efficiency of late gene activation (13), a direct role for T-antigen in the induction of VP-1 expression has yet to be rigorously demonstrated. The second

region of the SV40 template, which is required for trans-activation of the late transcription unit, is the 72-bp tandem repeats. This regulatory element serves as a transcriptional enhancer for SV40 early gene expression (2,3,4). We have previously demonstrated that base substitution mutants that quantitatively affect enhancer function in the early orientation similarly affect the level of T-antigen-mediated late gene expression (14). Thus, it seems likely that DNA sequences that control early gene enhancement also are important for late gene activation.

Specific in vivo competition for the SV40 early enhancer function apparently involves only the 72-bp tandem repeats (16). This result suggests that the enhancer binding factors required for early gene expression are able to bind independently to the 72-bp element. Our in vivo competition studies demonstrate that the putative DNA binding proteins responsible for late gene expression do not efficiently associate with DNA fragments containing either the 21- and 72-bp repeats or the 72-bp repeat element alone. These results raise the possibility that different proteins may interact with the SV40 72-bp repeat region in the induction of early gene expression and T-antigen-mediated late gene expression. Further experiments will be required to demonstrate this point conclusively.

Results obtained upon transfection of an SV40 late promoter CAT plasmid into the ts Cos cell demonstrated that expression of the SV40 late promoter was significantly reduced at the nonpermissive temperature. This observation is consistent with a model for the trans-activation of the SV40 late promoter which involves either direct interaction of T-antigen with the template or induction/modification of a cellular transcription factor.

Our present studies are directed toward the development of in vitro transcription systems in which the trans-activation of the late promoter can be studied. Preliminary results suggest that activation of the SV40 late promoter may be possible in vitro using COS-1 cell extracts (M.A. Thompson, unpublished results). Such systems will help to characterize, at the molecular level, the interaction of T-antigen and other transcriptional regulatory proteins with the template DNA.

We have shown that SV40 T-antigen stimulates CAT enzyme production under the control of the E2 promoter and that CAT enzyme levels are

correlated with steady-state CAT mRNA levels (11). Our studies suggest that the mechanism of promoter activation by SV40 T-antigen and Ad E1A is different, and that a specific promoter might be activated by more than one mechanism. This conclusion is drawn from two observations. First, the effect of saturating levels of T-antigen and E1A on the E2 promoter are additive in CV-1 cells. Second, E1A, but not T-antigen, can activate the E2 promoter in COS-1 cells (11).

We have demonstrated that in CV-1 cells the synthesis of CAT enzyme from 10 μg of pEC113 reached a plateau with either 5 μg of pRSV-T or 10 μg of pE1A. This was not due to a limitation in the ability to produce T-antigen or E1A. We found that T-antigen and E1A mRNA levels, as determined by Northern blot analysis, continued to increase when 2.5 to 20 μg of T-antigen- or E1A-encoding plasmids was transfected into CV-1 cells (data not shown). It seems more likely that the plateau in E2 expression results from the titration of a limiting cellular transcription factor. If SV40 T-antigen and Ad E1A activation of the E2 promoter were mediated by the same transcription factor, cotransfection of pE1A and pRSV-T should not increase the level of E2 expression over that observed in the presence of either T-antigen or E1A alone. In contrast, if different cellular factors are required for trans-activation by T-antigen and E1A, cotransfection of pE1A plus pRSV-T might lead to E2 expression, which is the sum of the two independent reactions. Clearly, our results are consistent with the latter possibility, suggesting the presence of distinct cellular factors which mediate the indirect effects of T-antigen and E1A. Along these lines, it has been shown that it is possible to activate a transfected human β-globin gene by more than one mechanism. Activation may be accomplished by introduction of the SV40 enhancer and transfection into HeLa cells or by transfection of the β-globin plasmid without an enhancer into 293 cells (22). It is reasonable to assume that the factors which trans-activate the transfected β-globin gene in 293 cells are different from those which regulate β-globin expression in reticulocytes.

The 5′ deletion analysis of the E2 promoter indicates that similar distal transcriptional control sequences are required for E1A and T-antigen trans-activation as well as basal E2 promoter activity. It is

possible, however, that T-antigen and E1A require separate sequences downstream from -79 bp. Sequences between -79 and -70 bp may represent the minimal promoter limit. Our transcription studies using a chemically synthesized Ad E2 promoter sequence from -85 to -29, suggest that E1A and T-antigen *trans*-activation is independent of the downstream promoter sequences between -29 and +40. Further analysis is required to determine if E1A and T-antigen recognize distinct target sequences within the -85 to -29 region.

It is not clear whether direct binding of T-antigen or E1A to the E2 promoter is involved in promoter activation. There is little apparent similarity between promoters which are activated by T-antigen, which include SV40 late (13,14,16,23,24), Ad E2 (11), Ad E3 (25), Rous sarcoma virus long terminal repeat (25), and Ad major late promoters (J. Manley, personal communication). Thus, for some promoters, it is likely that indirect mechanisms are involved. In support of this concept, *in vitro* assays of T-antigen binding to the E2 promoter have been negative (M. Brown and D. Livingston, personal communication). In the case of E1A, which does not bind directly to Ad DNA (12), little sequence similarity is observed in the promoters which it activates. It seems likely, therefore, that E1A may also activate specific promoters by an indirect mechanism. The evidence presented here suggests that the transcription factors which mediate T-antigen and E1A *trans*-activation of the E2 promoter are different.

REFERENCES

1. Tooze, J. (ed.) Molec. Biology of Tumor Viruses, 2nd ed., Cold Spring Harbor Laboratory, 1980.
2. Banerji, J., Rusconi, S. and Schaffner, W. Cell 27:299-308, 1981.
3. Gruss, P., Dhar, R. and Khoury, G. Proc. Natl. Acad. Sci. USA 78:943-947, 1981.
4. Moreau, P., Hen, R., Wasylyk, B., Everett, R., Gaub, M.P. and Chambon, P. Nucleic Acids Res. 9:6047-6068, 1981.
5. Baker, C.C. and Ziff, E.B. J. Mol. Biol. 149:189-221, 1981.
6. Berk, A.J., Lee, F., Harrison, J., Williams, J. and Sharp, P. Cell 17:935-944, 1979.
7. Guilfoyle, R.A., Osheroff, W.P. and Rossini, M. Embo J. 4:707-713, 1985.
8. Nevins, J.R. Cell 26:213-220, 1981.

9. Feldman, L.T., Imperiale, M.J. and Nevins, J.R. Proc. Natl. Acad. Sci, USA, 79:4952-4956, 1982.
10. Imperiale, M.J., Feldman, L.T. and Nevins, J.R. Cell 35:127-136, 1983.
11. Loeken, M.R., Khoury, G. and Brady, J. Mol. Cell. Biol. 6:2020-2026, 1986.
12. Ferguson, B., Krippl, B., Andrisani, O., Jones, N., Westphal, H. and Rosenberg, M. Mol. Cell. Biol. 5:2653-2661, 1985.
13. Brady, J., Bolen, J.B., Radonovich, M., Salzman, N. and Khoury, G. Proc. Natl. Acad. Sci. USA 81:2040-2044, 1984.
14. Brady, J. and Khoury, G. Mol. Cell. Biol. 5:1391-1399, 1985.
15. Weiher, H., König, M. and Gruss, P. Science 219:626-631, 1983.
16. Brady, J., Loeken, M.R. and Khoury, G. Proc. Natl. Acad. Sci. USA 82:7299-7303, 1985.
17. Innis, J. and Scott, W. Mol. Cell. Biol. 4:1499-1507, 1984.
18. Rio, D.C., Clark, S.G. and Tjian, R. Science 227:23-28, 1985.
19. Imperiale, M.J., Hart, R.P. and Nevins, J.R. Proc. Natl. Acad. Sci. USA 82:381-385, 1985.
20. Imperiale, M.J. and Nevins, J.R. Mol. Cell. Biol. 4:875-882, 1984.
21. Murthy, S.C.S., Bhat, G.P. and Thimmappaya, B. Proc. Natl. Acad. Sci. USA 82:2230-2234, 1985.
22. Treisman, R., Green, M.R. and Maniatis, T. Proc. Natl. Acad. Sci. USA 80:7428-7432, 1983.
23. Keller, J.M. and Alwine, J. Cell 36:381-389, 1984.
24. Keller, J.M. and Alwine, J. Mol. Cell. Biol. 5:1859-1869, 1985.
25. Alwine, J.C. Mol. Cell. Biol. 5:1034-1042, 1985.
26. Hamer, D.H. and Khoury, G. In Gluzman, Y. and Shenk, T. (Eds.): Enhancers and Eukaryotic Gene Expression, Cold Spring Harbor Laboratory, 1-15, 1983.

7

REGULATION OF GENE EXPRESSION FROM THE POLYOMA LATE PROMOTER.

F.G. KERN, P. DELLI-BOVI, S. PELLEGRINI, AND C. BASILICO

Department of Pathology, New York University School of Medicine, New York, New York 10016.

ABSTRACT

An experimental system has been developed in which foreign coding sequences for easily assayed genes or selectable markers are placed under the control of the polyoma late promoter. The use of this system has allowed the elucidation of the cis-acting elements that control expression from this promoter and the demonstration of transactivation of this promoter by the viral early proteins. In addition to these forms of transcriptional control, the results also suggest that late transcription, both in polyoma transformed cells and early in the course of a lytic infection, may also be regulated at a posttranscriptional level.

INTRODUCTION

The polyoma virus (Py) genome, consisting of approximately 5.3 kilobases (kb) of double-stranded DNA, is divided into two transcriptional units that, during productive infection of permissive mouse cells, appear to be temporally regulated. Transcription of the early strand, which produces the mRNAs coding for the three viral early proteins, the large T (LT), middle T (MT) and small T (ST) antigens via alternative splicing patterns and utilization of overlapping reading frames, proceeds rapidly after the introduction of the viral genome into the infected cells (1). While the function of the MT and ST remains obscure, the LT, being the viral replication protein, is essential for productive infection (2). LT binds to defined regions located near the origin (ori) of viral DNA replication (3, 4, 5). This binding is involved in the initiation of viral DNA

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

synthesis in cells possessing the additional necessary "permissivity factors" as well as in autoregulation of early region transcription (6, 7).

It is only after the onset of DNA replication that mRNA's derived from the opposite late strand are clearly detectable in the cytoplasm. These mRNA's, again produced by differential splicing patterns and utilization of overlapping reading frames, code for the three viral structural proteins VP1, VP2, and VP3 (1). An unusual feature of late transcription is the inefficient use of termination and polyadenylation signals (8). This leads to the formation of giant multimeric transcripts and the splicing out of a genome length intron to produce mature mRNAs that possess a reiterated leader sequence 57 bp in length (9). Although this leader possesses a 40S ribosomal subunit binding site (10) and a 10 bp sequence with complementarity to the 3' end of 18S rRNA (11), recent evidence indicates that while a leader sequence is necessary, these functions of the leader, per se, are not essential for viability of the virus. Rather, the leader appears to serve an essential spacer function and other non-viral sequences can substitute for the ribosomal binding site and 10 bp sequence in serving this function (12).

Situated between the early and late transcriptional units is a non-coding segment of approximately 460 bp in length that contains the ori (13, 14), the LT binding sites (15), and a region of 244 bp that contains the viral enhancer sequences (16). The necessity of these enhancer sequences in cis for early region transcription was readily established via deletion analysis (17). The independent requirement of these sequences for late transcription however was a question that could not be easily approached owing to the apparent necessity of DNA replication and/or LT for late coding sequences to be detected as mRNA.

Although it is known that enhancer sequences can operate bidirectionally (18), the early and late promoters appear to represent two different classes of RNA polymerase II promoters. The early control region possesses both the canonical TATA and CAACT boxes and transcripts initiate from a limited number of sites (19). In

contrast, the late promoter lacks both of these upstream control elements, and the initiation of late transcription is extremely heterogeneous (9, 20). Thus, it is conceivable that these promoters might have different additional upstream control elements. Supplying LT in trans through the use of helper virus or introducing the genomes into mouse cells consitutively producing LT is not able to circumvent this problem since the enhancer sequences are also required for DNA replication, independent of the enhancer requirement in cis for early region expression (21).

If viral DNA replication is in some manner inhibited, either through the use of mutants defective in the ori or in the LT coding sequences, or through the infection or transfection of cells such as rat or hamster cells that lack the necessary permissitivity factors, then the virus can undergo an alternative interaction with the cell resulting from the integration of the viral genome into the host cell genome (22). As a result of the expression of the integrated early region, in particular the MT coding sequences (23), the cell obtains a transformed phenotype. In transformed cells, the enhancer sequences are required in cis for efficient early region expression and the same 5' control sequences, mRNA initiation sites, splicing signals, as well as 3' polyadenylation signals that operate during the lytic cycle are also used for transcripts initiating from integrated templates (22).

In a situation that may be analogous to that found early in the lytic cycle, gene expression from integrated templates also appears to be limited to the early region. Late transcripts are not easily detected in the population of steady-state mRNA extracted from transformed cells in which the viral DNA exists only in the integrated state, even when it can be demonstrated that the late coding sequences remain intact. In a minor population of transformed rat cells, the integrated sequences can excise by a process of in situ replication and homologous recombination. This process requires a functional ori LT and is facilitated by regions of homology generated by the head to tail tandem insertions of viral DNA typically found in transformed cells (22, 24). In such a situation where extrachromosomal copies of

viral DNA exist in the cell, late mRNAs can be readily detected (25), thus ruling out the possibility that the transformed cells lack any possible cellular factors required for late region expression. Again, however, in this situation late expression is dependent on viral DNA replication and early gene expression.

While the reason for the lack of expression of late m-RNAs in Py transformed cells is still not clear, our laboratory has taken advantage of the observation that transfection of cells with plasmids that contain foreign coding sequences linked to the Py late promoter results in the efficient expression of these sequences (25). The use of these "indicator" genes has enabled us to determine the cis-acting elements and trans-acting factors that regulate expression from the Py late promoter, as well as the sequences within the Py control region that interact with these trans-acting factors. As a second approach to examining the control of late gene expression, we have molecularly cloned the viral insertion present in a novel cell line that produces appreciable steady-state levels of late mRNA. Restriction enzyme analysis reveals extensive rearrangements and leads to testable models concerning the possible role of posttranscriptional processing events in late gene expression.

MATERIALS AND METHODS

Details of the procedures used for cell culture, DNA transfection, selection in G418 containing medium, RNA extraction and Northern blot analysis, mapping of 5' termini by S1 nuclease analysis, plasmid isolation and construction, and assay for chloramphenicol acetyl transferase activity (CAT) are described in other publications from this laboratory (25, 26, 27, 28). For the genomic cloning of the Py insertion present in the SS1A cell line (29), DNA was digested with EcoRI and size fractionated on an agarose gel. The 16-21 kb region containing the right end of the Py insertion was electroeluted and ligated to dephosphorylated EcoRI arms of the EMBL4 lambda phage cloning vector (30) and packaged. Plaques were screened using a 2.4 kb HhaI fragment that spans the late region of Py.

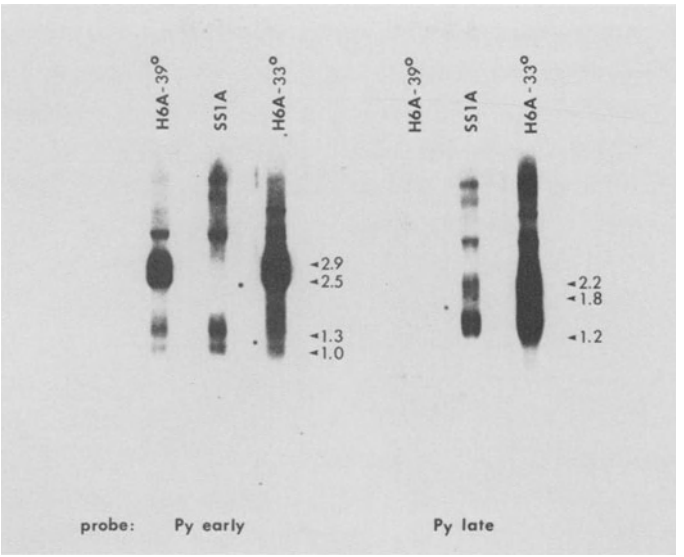


Fig. 1. Northern analysis of early and late mRNAs present in a tsA Py transformed cell line H6A at permissive (33°C) and nonpermissive (39°C) temperatures for LT function and in a cured revertant cell line SS1A. Lanes contain 5.0 µg of polyA$^+$ RNA from H6A at 39°C and SS1A at 37°C and 2.0 µg from H6A at 33°C. Blots were hybridized to a ^{32}P-labeled 1.8 kb Py PstI-3 early region fragment (left) or a 2.4 kb Py HhaI late region fragment (right).

RESULTS

Late transcription is not detectable in Py transformed cell lines in the absence of LT function.

As stated above, the Py genome will integrate into a proportion of infected nonpermissive rat cells, and subsequent expression of the early region results in the cells obtaining a transformed phenotype. Three early mRNA's with the same size as found early in a lytic infection are easily detectable in transformed cells by Northern analysis. This is illustrated in Figure 1 using polyA$^+$ RNA extracted from the cell line H6A grown at 39°. This cell line was obtained by infection of rat fibroblasts with the tsA strain of Py (29). Consequently, at 39° the LT produced by these cells is

non-functional and excision of integrated Py sequences does not occur. Figure 1 also shows that at 39° late transcripts are virtually non-detectable. At 33° the LT is functional and the ensuing excision and replication of the integrated Py sequences results in the presence of extrachromosomal viral DNA copies, that presumably can now function as templates for late transcription. As expected at 33°, late transcripts are abundant in the population of H6A polyA$^+$ mRNA. Figure 1 also illustrates that a novel cell line, designated SS1A, that was isolated as a cured revertant of the H6A cell line (29), produces truncated versions of the three species of early mRNA's and therefore a LT lacking the carboxy terminal portion and no longer capable of supporting viral DNA replication. Unexpectedly, it was observed that this cell line also produces detectable levels of late mRNAs (31). Southern blotting indicated extensive rearrangement of portions of the integrated sequences. This suggested that one approach to examining the control of late expression would be to determine more precisely the nature of these rearrangements via molecular cloning of the insert, and a restriction map of a portion of the viral insertion cloned into the EMBL4 phage vector is described below.

Replacement of late coding sequences with foreign genes results in detectable steady-state levels of mRNA.

A new alternative approach to the question of late gene regulation was suggested by the behavior of rat cells transfected with plasmid pPyNeo (Fig. 2), that contained a complete tsA Py early region and linked to the Py late promoter at the BclI site at nucleotide (n) 5022, the coding sequences for the Tn5 bacterial transposon gene for neomycin-resistance (neo) with a herpes simplex virus (HSV) thymidine kinase (tk) gene polyadenylation signal 3' to the neo sequences. When placed under the control of an active eukaryotic promoter and transferred into cells in culture, these sequences are able to confer resistance to the drug G418 (32). Agar colonies of cells morphologically transformed were isolated at 39°C and expanded into cell lines. Given the lack of detectable authentic late transcription in the absence of a functional LT observed in Py transformed cells, it was somewhat of a surprise to observe that 5 of 6 cell lines derived

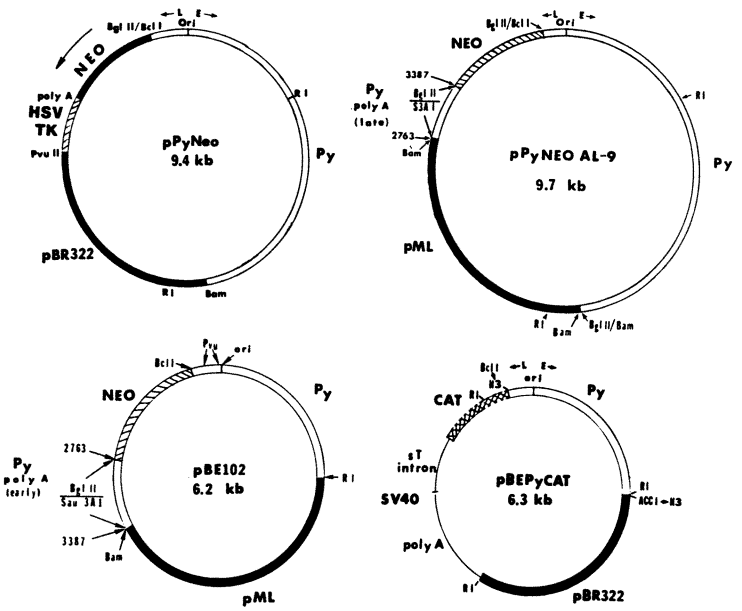


Fig. 2. Schematic representation of plasmids used to study gene expression from the Py late promoter. Plasmids pBE102 and pBEPyCAT were used for subsequent deletion analysis of the sequences required for late promoter function and transactivation by the Py early proteins.

from the pPyNeo transfection had high plating efficiencies in medium containing G418 (25). Northern analysis using polyA$^+$ mRNA isolated from one of these cell lines indicated that steady-state levels of neo specific mRNA approximated the levels of early mRNA present in the same cell line (Fig. 3, left panel). The size of the major neo transcript was 1.2 kb and if one assumed that the message was unspliced, this placed the approximate 5' initiation sites of the transcript at the junction of the neo and Py sequences. S1 analysis confirmed that the transcripts were initiating from within the Py late promoter region (Fig. 4). Moreover, longer autoradiographic exposures of the same gel indicate that the same initiation sites utilized

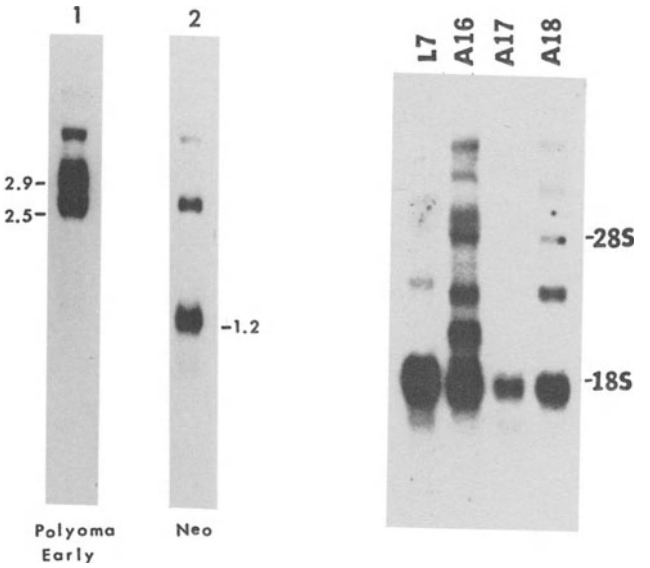


Fig. 3. Northern analysis of neo-specific mRNAs in rat cell lines derived from agar colonies following transfection with pPyNeo (left panel) or pPyNeoAL-9 (right panel). Lanes contain 5μg of polyA$^+$ mRNA from cells grown at 39°C hybridized to a ^{32}P-labeled Py 1.8 kb PstI-3 early region fragment (left panel, lane 1) or to a ^{32}P-labeled 1.1 kb BglII-SalI fragment containing neo coding sequences (left panel, lane 2 and right panel). L7 is a cell line selected directly in G418-containing medium. Cell lines A16, A17, A18 are derived from agar colonies that were subsequently found to be G418^R.

during a lytic infection of mouse 3T6 cells or during transcription of extrachromosomal templates in rat cells are also used by cells transformed by the chimeric Neo plasmids.

Since the pPyNeo plasmid contained an HSV tk polyadenylation signal, the possibility existed that the ability to detect neo transcripts in transformed cells and the inability to detect authentic late transcripts was due to inefficient utilization of the Py 3' processing signals rather than an inactive late promoter. Further support for this hypothesis comes from the above mentioned findings that these signals are inefficiently utilized during a lytic infection. Additional neo-containing constructs that were tested

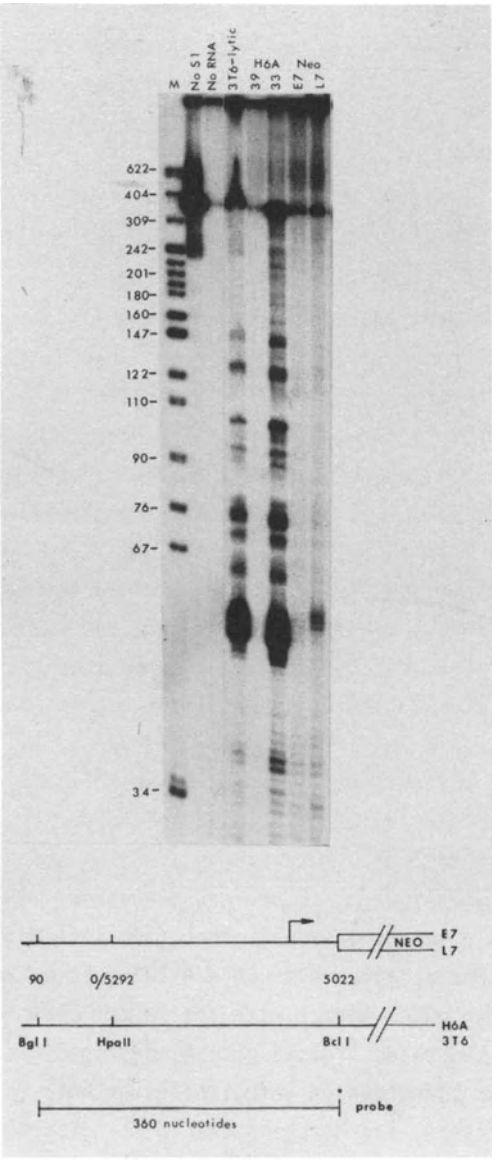


Fig. 4. S1 mapping of the 5' termini of neo-specific mRNAs and Py late mRNAs in transformed rat and lytically infected mouse cells. The probe is the 360 nucleotide L strand of the Py BclI-BglI fragment end-labeled at the BclI site. Lanes show the size of protected fragments found using 20μg of polyA$^+$ RNA from the tsA Py transformed rat cell line H6A grown at 33° and 39° or from two G418^R cell lines E7 and L7 grown at 39° or with 1μg of poly A$^+$ RNA from 3T6 cells extracted 48 hours after infection with tsA Py virus at 33°. Reprinted with permission from reference 25.

ruled out this possibility, however. Rat cells were transfected with a plasmid pPyNeoAL-9 which is similar to pPyNeo with the exception that in this case, the Py late polyadenylation signal was linked 3' to the neo sequences (Fig. 2). PolyA$^+$ RNA was extracted from cell lines derived from agar colonies that were subsequently found to be G418-resistant (G418^R). Northern analysis indicated the presence of a 1.6 kb neo transcript in all of the cell lines tested, which was the size expected if the late polyadenylation signal were being used (Fig. 3, right panel). Thus, even in the absence of direct selection for G418-resistance, the late polyadenylation signal was being effectively utilized.

These results therefore indicate that both the 5' and 3' processing signals for late transcription are operative when integrated into the host cell genome and therefore strongly imply that the absence of detectable late transcripts in Py transformed cells is due to posttranscriptional processing events.

One possibility that we have recently investigated is that the splicing of the primary late transcript to form the mature spliced mRNA for VP1 is an inefficient process. neo sequences were attached at the unique viral EcoRV site at n 4106 site just downstream of the splice acceptor site for the VP1 message at n 4127 (10). The efficiency of G418^R colony formation following transfection of rat cells was then compared to a plasmid in which the neo sequences were attached to the BclI site at n 5022. Both plasmids contained a Py early polyadenylation signal 3' to the neo sequences. In the case of the former plasmid, a splicing event between the late splice donor site at n 5023 and the acceptor site at n 4127 would be required to form a functional neo mRNA whereas with the latter plasmid, a functional mRNA is generated from an unspliced transcript. Both plasmids had similar transfection efficiencies on both rat F2408 and mouse NIH3T3 fibroblasts, and Northern analysis indicated the presence of a transcript consistent with the size expected if the usual donor to acceptor splice had occurred (data not shown). These results rule out a splicing defect as being the cause for the absence of late transcription in transformed cells. In addition, they suggest that an attenuation mechanism prematurely terminating transcription within the

stretch of coding sequences between the BclI site and EcoRV sites is also not operative.

Another possibility that is presently being explored is that late transcripts are produced in transformed cells but are then rapidly degraded. If true, this would imply that the late RNAs transcribed from integrated DNA molecules are different from those produced in lytic infection and contain destablizing structures or lack sequences conferring stability to the RNAs. To investigate this possibility, we have studied the arrangement of integrated Py DNA sequences in the cell line SS1A, which, as described above, is the only Py transformed cell line in which we could detect the presence of stable late m-RNAs. A 19.0kb EcoRI fragment that contained the late coding sequences was molecularly cloned into the EMBL4 phage vector and analyzed by restriction mapping (Fig. 5).

Such an analysis failed to reveal any gross rearrangement of late promoter or enhancer sequences within the region immediately adjacent to the late coding sequences. However, this analysis did reveal that the region upstream of these sequences had undergone extensive rearrangements involving amplification of an approximately 560 bp stretch of DNA that contains the late promoter, enhancer and ori. In addition, an excision event has removed most of the early coding sequences that were originally present in the parental H6A cell line, with the exception of a truncated early coding region lacking the carboxy terminus of LT that is immediately adjacent to the flanking cellular sequences. Also removed along with these early coding sequences were the regions that contain the polyadenylation signals.

To verify that the ability to detect stable late transcripts was indeed due to these rearrangements and not to a peculiarity of the cell line, the 19.0 kb EcoRI insert was subcloned into a plasmid vector and cotransfected onto rat F2408 fibroblasts with a plasmid conferring G418-resistance. $G418^R$ colonies were picked, mRNA was extracted and analyzed by Northern analysis for the presence of late mRNAs. As expected, colonies containing the 19.0 kb SS1A EcoRI fragment produced late mRNAs in appreciable quantities.

The analysis of the structure of this insert reveals that it is now possible that a primary late transcript could initiate at the far

right end of the insertion or within any of the amplified late promoter regions and proceed through the remaining amplified origin, enhancer and late promoter sequences until it reaches the sole remaining copy of the late coding region and late polyadenylation signal. Since this transcript could contain up to nine copies of the late leader sequence as well as the leader splice donor and acceptor

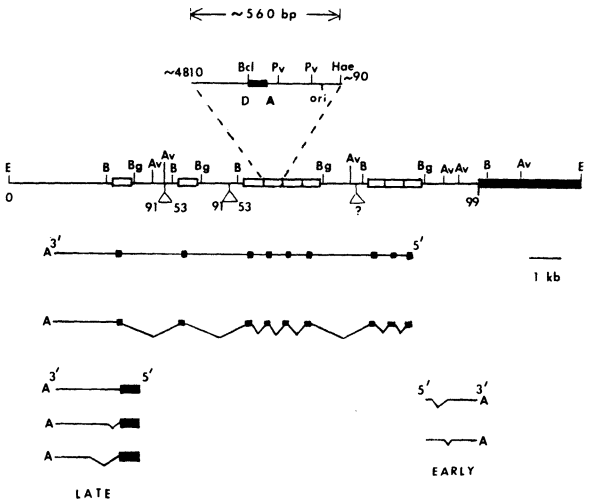


Fig. 5. Restriction enzyme mapping of the Py insertion present in the SS1A cell line. The 19.0 kb EcoRI fragment containing the right hand portion of the complete Py insertion and flanking cellular rat sequences (heavy black line) was molecularly cloned using the EMBL4 phage vector. Shown are the locations of the Bam HI (B), BglI (Bg), and AvaI (Av) restriction enzyme sites. Open box represents the area of ~ 560 bp from ~ n 4810 to n 90 that is amplified within the insertion. Within this 560 bp region is contained the Py origin of replication (ori), the enhancer region located within the region spanning the BclI (Bcl) to the second PvuII site (Pv), the Py late promoter, the sequence that forms the reiterated leader (blackened area) bordered by splice donor (D) and splice acceptor (A) sites and a Hae II (Hae) restriction enzyme site. Triangles below the map indicate the deleted areas of Py genome with numbers representing the map units deleted. The line immediately below the map represents the hypothetical primary late transcript that due to the deletions could extend from the first copy of the late promoter to the sole remaining copy of the late polyadenylation and mRNA cleavage signals. This primary transcript could then be spliced to generate a reiterated leader sequence (black box) linked to the late coding sequences.

sites, the possibility exists that mature spliced late mRNAs could be generated that contain a reiterated leader sequence identical to that observed during a lytic infection. Normally in transformed cells containing an intact early region, the presence of the late polyadenylation addition and mRNA cleavage signals that overlap the early polyadenylation signal would preclude the possibility of generating such a transcript.

We are currently in the process of determining whether such a transcript containing a reiterated late leader sequence does indeed exist in the SS1A cells and whether there is an actual increase in the rate of late transcription compared to the parental H6A cell line. It is possible that the ability to detect late transcription in the SS1A cell line is not so much due to an increase in the transcription rate, but rather is due to stabilization of late mRNAs by the presence of the reiterated leader sequence. An alternative explanation is that the increased number of enhancer sequences present due to the amplification event facilitates an increased rate of late transcription (33).

Cis acting factors affecting expression from the late promoter.

Regardless of the reason why authentic late transcripts are not found in transformed cells, our finding that the Py late promoter efficiently directs the expression of foreign coding sequences provided an experimental system with which to study through deletion analysis the regulatory elements of this promoter. In order to demonstrate that these elements also acted on the late promoter independent of their effect on early transcription, it was first necessary to demonstrate that a plasmid devoid of early coding sequences could still direct the expression of the neo gene when it was attached to the late promoter. A plasmid was constructed that again contained the neo gene and linked 3' Py early polyadenylation signal attached to the Py late promoter at the BclI site at n 5022. Upstream Py sequences extended through the ori and ended at n 153, before the start of the early coding sequences. When transfected into rat F2408 fibroblasts, this plasmid was found to have the same relative efficiency of transformation to G418-resistance at 39°C as the control plasmid containing the tsA early region (26). Thus,

expression from the late promoter did not require the presence of any of the three viral early proteins, and the effects of deletions in the non-coding regulatory region on late promoter function could be directly assessed.

To accomplish this goal, we used plasmids that contained the deletions within the Py control region that are shown schematically in Fig. 6. All plasmids contained neo sequences and a 3' Py early

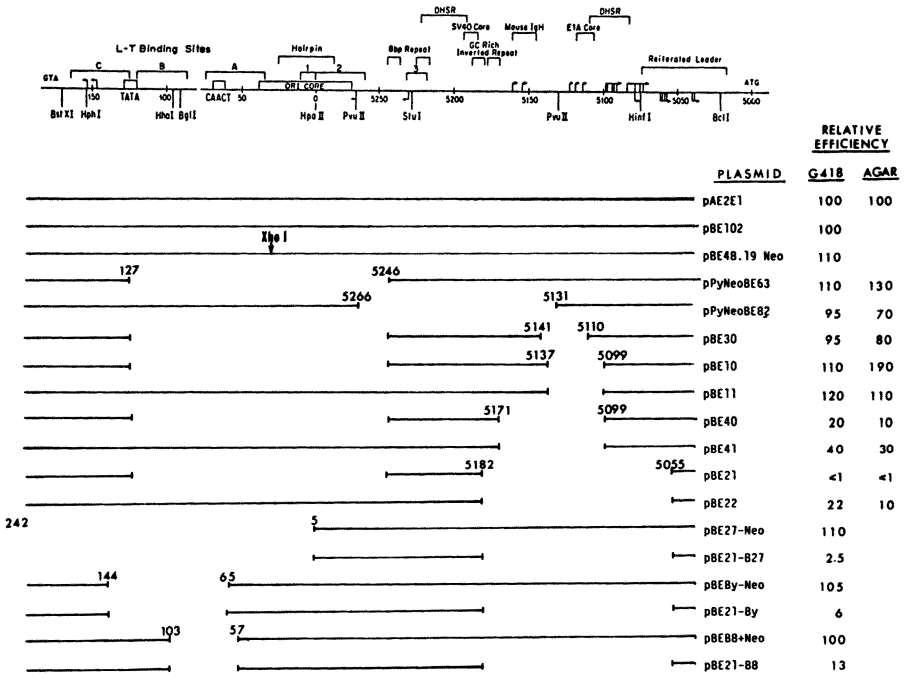


Fig. 6. Deletion-mapping of cis-acting elements affecting Py early and late promoter function. Deleted regions are indicated by the open spaces. Relative efficiencies of colony formation in G418-containing medium or in soft agar are based on comparison to the frequency observed with the wild-type pBE102 plasmid shown in Fig. 2 in transfection experiments using rat F2408 fibroblasts as recipient cells. Pertinent features of the Py control region are referenced in the text and in references 26 and 27. Arrows pointing left indicate the direction and location of the 5' termini of early (43) and late-early (44) transcripts. Arrows above the line pointing right indicate the direction and location of 5' termini of late lytic transcripts. Arrows below the line pointing right indicate additional 5' termini of transcripts in the late direction observed after transfection with chimeric plasmids (25).

polyadenylation signal linked to the late promoter at the BclI site and, in addition, contained a truncated early coding region ending at the Py EcoRI site at n 1560 (Fig. 2). Since these plasmids can encode a complete MT and ST, but only a truncated LT, we could also assess the effects of these deletions on early gene expression independent of their effects on replicating ability by scoring for the number colonies formed in soft agar.

The results of a number of independent experiments are summarized in Fig. 6. The finding of major interest is that the levels of both early and late transcription are regulated by the same set of regulatory sequences. A deletion on n 5246-127 that spans the ori had no effect on either early or late gene expression, as did a deletion of the PvuII-4 fragment that contains the so-called B enhancer sequences (34). Present within this fragment are the omega sequences present in all mutants viable in neuroblastoma or embryonal carcinoma cells (35), the SV-40 core enhancer sequences (36), and various regions with homology to the mouse Ig heavy chain (IgH), bovine papilloma virus, Rous sarcoma virus, and Adenovirus 5 E1A enhancers (37). Bal 31 nuclease generated deletions around the PvuII site at n 5130 that spanned n 5099 to 5141 also had no effect on both early and late transcription. This A enhancer region (34) contains the Ad 5 E1A core enhancer sequences (38), the P-motif sequences found in the SV40 enhancer (33), as well as regions with homology to the Moloney murine sarcoma virus long terminal repeat and IgH enhancer (37). Extension of this deletion to n 5099 had the effect of diminishing activity in both directions, and S1 analysis of 5' initiation sites indicated that for late region expression, this decrease was due to removal of regulatory rather than structural elements of the late promoter (26). A deletion spanning n 5055-5182 had the effect of even further decreasing and again to similar extents, both early and late expression.

A surprising observation was made that when a second ori deletion of n 5246-127, which by itself in these assay systems had no effect on gene expression in either direction, was placed on this n 5055-5182 deletion mutant background, the effect was that of essentially abolishing transcription in both directions. We attempted to define

whether a particular set of sequences was involved by substituting three other ori deletions into this pBE22 enhancer region mutant. Again, these ori deletions by themselves had no measurable effect on the ability to form colonies in G418. However, all three deletions had the effect of potentiating the repressive effect of the pBE22 deletion on late gene expression. No particular set of sequences could be localized that was responsible; rather it appeared that the larger the size of the deletion of the noncoding sequences in this area, the greater the effect. It therefore appears that the non-coding regulatory region for late transcription extends upstream beyond the borders of the Py enhancer region defined by the BclI site at n 5021 and PvuII site at n 5267, and like the enhancer sequences, these upstream regulatory sequences may also consist of multiple complementing elements.

Transactivation of the Py late promoter by the viral early proteins.

Although the results presented above indicated that a functional early region was not necessary for late transcription, this did not preclude the possibility that the viral early proteins could affect expression from the late promoter. To approach this question, we have used a transient assay system in which the bacterial chloramphenicol acetyl transferase (cat) gene was linked to the late promoter at the BclI site by attaching HindIII linkers to a Py BclI-EcoRI fragment and inserting this fragment into HindIII digested pSVOCAT vector DNA (39). The prototype plasmid pBEPyCAT therefore contains a truncated early coding region capable of encoding functional ST and MT, cat linked to the late promoter with 3' processing and polyadenylation signals from the SV40 early region (Fig. 2). The plasmid also lacks the carboxy terminal portion of LT. The truncated LT is incapable of supporting viral DNA replication.

To rule out effects due to template amplification that could occur when the above described plasmid was cotransfected with a plasmid encoding a complete functional LT, an additional plasmid was constructed using the BclI-EcoRI fragment from plasmid p48.19 that was generously provided by G. Veldmann and R. Kamen. This fragment contains a XhoI linker inserted between n 35 and n 37 within the ori core sequences that essentially destroys the ability of the plasmid to

replicate. Additional plasmids were constructed that contained various deletions in the noncoding regulatory region that removed either LT binding sites or enhancer sequences. All target plasmids were replication-defective due either to the presence of the XhoI linker insertion or to removal of the ori sequences. A final non-replicating target plasmid that totally lacked any early coding sequences was also constructed using the origin spanning BclI-BstXI fragment (n 5021-n 173) from p48.19.

These target plasmids were transfected onto mouse NIH 3T3 or rat F2408 cells or cotransfected with a plasmid that encodes only for LT. Two different cotransfecting plasmids were utilized. One contained the LT sequences present in a Py ori-deleted vector while the other plasmid had LT sequences inserted into the 91023 vector (40) under the control of the SV40 early promoter (28). Thus, both cotransfected plasmids were also incapable of replication. Protein extracts were prepared 60 hr post transfection and assayed for CAT activity. The results of this analysis are summarized in Fig. 7. While for reasons of simplicity only the results obtained in mouse NIH3T3 cells are presented here, suffice it to say that transfection into rat cells gave essentially concordant results (27), the only significant difference being that in these cells all cat-containing plasmids exhibited a significantly higher base line activity (5-10 fold higher than in mouse cells). The degree of stimulation by LT was therefore somewhat lower, even if the final levels of expression were higher than in mouse cells.

In mouse cells target plasmids that contained the Py enhancer sequences and the truncated early coding region directed a low but measurable level of cat expression. When the p48.19 BECAT plasmid, which contains the complete non-coding regulatory region, was cotransfected with LT-coding plasmids, CAT activity was increased 20-70 fold. Northern analysis indicated that this increase in CAT activity was reflected in a similar increase in the level of cat mRNA (28). Deletion of minor LT binding site 3, the site most proximal to the late promoter, had no effect on LT-mediated stimulation of CAT activity, and a deletion of major LT binding sites A and B and minor sites 1 and 2 also had little effect. A deletion of all three major

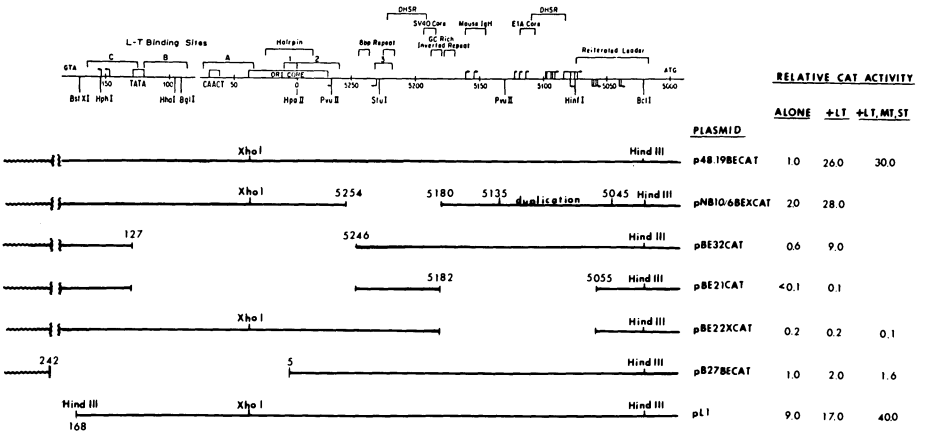


Fig. 7. Deletion mapping of the sequences within the Py control region required for transactivation of the late promoter by the viral early proteins. Deleted sequences are indicated by the open spaces. Numbers to the right of the plasmid name indicate normalized values based on data presented in reference 28 and unpublished data. The numbers indicate the relative level of CAT activity observed after transfection of NIH3T3 cells with the plasmid alone or after cotransfection with the replication-defective plasmids pBLTwt12 which encodes only the LT or p48.19 which encodes all three of the early proteins (27).

LT binding sites as well as most of minor site 1 in plasmid pB27BECAT had the effect of reducing the degree of LT mediated stimulation to only two to four fold over baseline levels. Also, a large deletion in the enhancer region in plasmid pBE22XCAT eliminated the LT effect. These results therefore suggested two sets of target sequences for LT transactivation. First, there appeared to be a requirement for at least one major LT binding site. While our results would suggest that binding site C is the most important, we have yet to rule out the possibility that either A or B can substitute. Secondly, there also appeared to be a requirement for enhancer sequences.

An alternative explanation for these results however was raised by the results obtained when the target plasmid pL1 lacking the truncated early coding sequences was utilized. This plasmid had all the LT binding sites as well as the complete Py enhancer. Baseline activities with this plasmid were typically five to ten fold higher than the baseline activity of the p48.19BECAT plasmid containing the truncated early region. However, cotransfection of pL1 with LT typically resulted in only a two to three-fold stimulation of CAT activity. Moreover, the final values obtained were always lower than that observed when p48.19BECAT was cotransfected with a LT encoding plasmid. This result therefore pointed to a possible involvement of the ST and or MT in potentiating the LT effect on late promoter activity. It also raised the possibility that the reason for the diminished response of the pB27BECAT plasmid lacking the LT binding sites was not due to the absence of these sites, but rather to the deletion of the amino terminal sequences of the ST and MT. Likewise, the reduced response of the enhancer-deleted pBE22XCAT plasmid could be due to a reduction in the level of early region expression caused by the lack of cis acting elements.

To address these questions the pB27BECAT plasmid lacking LT binding sites and the pBE22XCAT plasmid with the enhancer region deletion, as well as the pL1 plasmid, which contains both of these sets of sequences but lacks the truncated early coding region, were all cotransfected with an ori-defective non-replicating plasmid that encoded all three of the viral early proteins. The results indicated that cotransfection with the complete early region was not able to cause any further stimulation of pB27BECAT above that observed with LT only and had no effect on pBE22XCAT activity (Fig. 7). This, therefore, confirmed that both LT binding sites and enhancer sequences were both required for a maximal response to LT.

Cotransfection with the complete early region did however have a marked effect on the level of CAT activity directed by the pL1 plasmid, with the final values now being two to four fold higher than that observed with cotransfection with LT only and higher than that obtained with similarly cotransfected p48.19BECAT (Fig. 7 and reference 28). This result therefore argues for a role of either the

ST and/or MT in transactivation of the Py late promoter by LT. Since we also observe a slight stimulation when pL1 is cotransfected with a plasmid coding only for MT and since others have reported transactivation of a variety of heterologous promoters by MT (A. Pannuti, G. La Mantia and L. Lania, manuscript submitted), we currently favor the hypothesis of an involvement of MT in this phenomenon. Experiments are currently in progress that are designed to clarify this point and to determine whether MT and LT have additive or cooperative and synergistic effects on late gene expression.

DISCUSSION

Our finding that the Py late promoter can efficiently direct the expression of linked foreign coding sequences has enabled us to develop an experimental system that has permitted the study of the *cis* and *trans*-acting elements that control gene expression from this promoter. Our results indicate that the regulation of late gene expression occurs at both transcriptional and posttranscriptional levels. Analysis of deletion mutants revealed that the early and late promoters share a common set of regulatory elements that have been previously defined by others as enhancer sequences, as well as another set of sequences around the viral ori that affect late expression. Since this region contains the TATA and CAACT boxes for the early promoter as well as early mRNA initiation sites, it is difficult to determine whether this region also contains additional common regulatory elements that act bidirectionally.

An important finding is that the late promoter can function independent of the presence of the viral early proteins and also in the absence of extrachromosomal DNA replication. In fact, in transient CAT assays, we observe that a plasmid completely deleted of early coding sequences typically exhibits a five to tenfold higher level of activity directed by the late promoter than a plasmid that contains a truncated early region. In addition, the level of late CAT activity is approximately the same as that seen with a plasmid having the same polyoma sequences (n 5021-168) in the opposite orientation, thus placing the *cat* sequences under the control of the Py early promoter (unpublished results). These observations are difficult to

reconcile with the well documented absence of significant amounts of stable late transcripts early in the viral infectious cycle, as well as with the absence of late transcripts in cells containing exclusively integrated Py genomes.

One possibility suggested by the difference in baseline CAT activities between the p48.19BECAT and pL1 plasmids is that the early coding sequences in some manner acts in a cis fashion as a negative regulatory element to repress late transcription, perhaps by placing the DNA in a conformation that either allows the binding of a repressor molecule or inhibits the binding of a late promoter specific transcription factor. Subsequent production of LT and binding of this protein to the LT binding sites could then alter this conformation and reverse this process. This hypothesis is difficult to reconcile with the fact that neo transcripts initiating from within the late promoter are readily detected when constructs that also contain tsA or truncated early coding regions are utilized. One would have to postulate that when integrated, the DNA also undergoes a conformational change that would either allow the putative late transcription factor to bind or preclude the possibility of repressor binding. Therefore, this would also imply that the lack of authentic late expression in transformed cells and the inability to detect late transcripts early in a lytic infection are the result of two different molecular mechanisms. Whereas in lytic infections the control could be at the level of mRNA initiation, in transformed cells the control would be at a posttranscriptional level.

A second possibility is that in the initial phases of the infection the early promoter outcompetes the late promoter for necessary transcription factors. Deletion of the early coding sequences in the late promoter-cat constructs or production of LT and the resulting autoregulation of early transcription during the course of a viral infection would no longer permit this transcriptional unit to compete effectively, and transcription from the late promoter could now proceed at an increased rate. However, we have previously reported that a TK^- cell line that contained a single partial insertion of plasmid having the HSV tk coding sequences under the control of the Py late promoter as well as the tsA early region was

able to grow both in selective HAT medium as well as in semi-solid medium at 39°C (27). It therefore appears that at least when the plasmid is integrated, both the early and late promoters can be active on the same molecule. Thus, it seems unlikely that this model of promoter competition could fully account for the lack of late transcription in transformed cells.

A third possibility is that both early and late promoters are at least to some degree transcriptionally active early in infection as well as in transformed cells, but late precursor mRNAs are either inefficiently or improperly processed and are then rapidly degraded. Later in the course of infection, the increase in the number of late transcripts due to template amplification, autoregulation of the early transcriptional unit and transactivation of the late promoter by LT overcomes this fact that processing is inefficient, or, alternatively, there is also a change in the processing or type of late precursor mRNA molecule that is produced at this time.

Recent data from another laboratory suggests that late in the course of infection, there is a depletion of cellular factors required for termination of transcription and polyadenylation and mRNA cleavage (41), with the result that large multimeric transcripts consisting of the sense strand for late mRNA and antisense strand for early mRNA are produced. These multimeric transcripts are then spliced to generate the late mRNAs that contain the reiterated leader sequence. Normally in transformed cells the structure of the integrated Py sequences and the fact that the late polyadenylation signal functions efficiently in an integrated state, prevents the formation of such multimeric transcripts. In addition, in transformed cells there would most likely be a selection against the formation of an anti-sense early mRNA (42). However, as mentioned above, the structure of the integrated sequences present in the SS1A cell line isolated as a cured revertant raised the possibility that formation of such a reiterated leader structure may be important in stabilizing late mRNA. Thus, an additional factor regulating the level of late transcripts in infected cells may be a change in the stability of the mRNA that would act in concert with the template amplification due to DNA replication and the changes in promoter utilization due to LT mediated transactivation and autoregulation.

This concept of late mRNAs being inherently unstable is supported by our findings that both 5' and 3' late region control sequences are functional when they are used to direct the expression of integrated neo sequences. Thus, it is possible that the same molecular mechanism is responsible for the inability to detect authentic late transcripts both in transformed cells and during the early phase of a viral infection. This model would predict that late transcripts are initiated in transformed cells, and also that either portions of the late coding sequences would be able to confer instability to a normally stable transcript or the neo sequences would be able to confer stability to an unstable transcript. Experiments testing these predictions are currently in progress.

While our results with the late promoter-neo constructs argues for posttranscriptional events being involved in the control of late transcription, the results of cotransfection experiments of late promoter-cat constructs with plasmids encoding Py early sequences indicates that the late promoter is also regulated at a transcriptional level by the viral early proteins. Since the LT binding sites are necessary for the maximal transactivation effect to be observed and since these target sequences are located upstream of the late mRNA initiation sites, it is unlikely that the major effect of the early proteins is to increase the transport, stability, or translation of late mRNAs. In contrast to the usual twenty to sixty fold stimulation by the viral early proteins observed with targets having LT binding sites, a consistent two to four fold stimulation by LT was still observed with a target plasmid that lacked these binding sites (28). It is therefore possible that there is a secondary effect of LT that is indeed operating at a posttranscriptional level. However, since the enhancer region is also necessary for trans-activation, we favor the hypothesis that the major effect is due to the binding of LT to the viral regulatory region which in some manner facilitates the formation of a complex of transcriptional factors that interact with the enhancer region. The binding of LT is necessary but not sufficient for maximal transactivation and our data suggests that the MT and/or ST may also be involved in the formation of this complex. Investigations into which of these proteins is involved and the possible mechanism of interaction are currently being pursued.

ACKNOWLEDGMENTS

We thank R. Kamen, G. Veldman, A. Cowie, L. Dailey, and P. Amati for gift of plasmids; V. Levytska and E. Deutsch for excellent technical assistance; and D. Nazario for assistance in the preparation of the manuscript. This investigation was supported by PHS grants CA16239, CA42568, and CA41367 from the National Cancer Institute.

REFERENCES

1. Tooze, J. DNA tumor virus: Molecular Biology of Tumor Viruses, part 2. Cold Spring Harbor Laboratory, Cold Spring Harbor N.Y., 1980.
2. Franke, B. and Eckhart, W. Virology 55:127-135, 1973.
3. Gaudray, P., Tyndall, C., Kamen, R. and Cuzin, F. Nucleic Acids Res. 9:5697-5710, 1981.
4. Dilworth, S.M., Cowie, A., Kamen, R. and Griffin, B.E. Proc. Natl. Acad. Sci. U.S.A. 81:1941-1945, 1984.
5. Pomerantz, B.J., Mueller, C.R. and Hassell, J.A. J. Virol 47:600-610, 1983.
6. Fenton, R.G. and Basilico, C. Virology 121:384-392, 1982.
7. Farmerie, W.G. and Folk, W.R. Proc. Natl. Acad. Sci. U.S.A. 81:6919-6923, 1984.
8. Acheson, N.H. Proc. Natl. Acad. Sci. U.S.A. 75:4754-4758, 1978.
9. Treisman, R. Nucleic Acids Res. 8:4867-4888, 1980.
10. Soeda, E., Arrand, A., Smolar, N., Walsh, J. and Griffin, B., 1980. Nature 283:445-453.
11. Legon, S. J. Mol. Biol. 134:219-240, 1979.
12. Adami, G.R. and Carmichael, G.G. J. Virol. 58:417-425, 1986.
13. Luthman, H., Nilsson, M.G. and Magnusson, G. J. Mol. Biol. 161:533-550, 1982.
14. Katinka, M. and Yaniv, M. J. Virol. 47:244-248, 1983.
15. Cowie, A. and Kamen, R. J. Virol. 52:750-760, 1984.
16. de Villiers, J. and Schaffner, W. Nucleic Acids Res. 9:6251-6264, 1981.
17. Tyndall, C., La Mantia, G., Thacker, C.M., Favaloro, J. and Kamen, R. Nucleic Acids Res. 9:6231-6250, 1981.
18. Banerji, J., Rusconi, S. and Schaffner, W. Cell 27:299-308, 1981.
19. Jat, P., Novak, U., Cowie, A., Tyndall, C. and Kamen, R. Mol. Cell. Biol 2:737-751, 1982.
20. Cowie, A., Tyndall, C. and Kamen, R. Nucleic Acids Res. 9:6305-6322, 1981.
21. de Villiers, J., Schaffner, W., Tyndall, C., Lupton, S. and Kamen, R. Nature 312:242-246, 1984.
22. Basilico, C. Pharmac. Ther. 26:235-272, 1985.
23. Treisman, R., Novak, V., Favaloro, J. and Kamen, R. Nature 292:595-600, 1981.
24. Pellegrini, S., Dailey, L. and Basilico, C. Cell 3:943-949, 1984.
25. Kern, F.G. and Basilico, C. Mol. Cell. Biol. 5:797-807, 1985.
26. Kern, F.G., Dailey, L. and Basilico, C. Mol. Cell. Biol. 5:2070-2079, 1985.
27. Kern, F.G., Pellegrini, S. and Basilico, C. In: Cancer Cells, vol. 4, DNA Tumor Viruses: Control of Gene Expression and Replication. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., in press.

28. Kern, F.G., Pellegrini, S., Cowie, A. and Basilico, C. J. Virol., in press.
29. Colantuoni, V., Dailey, L. and Basilico, C. Proc. Natl. Acad. Sci. U.S.A. 77:3850-3854.
30. Frischauf, A.M., Lehrach, H., Poutska, A. and Murray, N. J. Mol. Biol. 170:824-842.
31. Basilico, C., Fenton, R.G. and Della Valle, G. In: Expression of Differentiated Functions in Cancer Cells (Ed. R.F. Revoltella), Raven Press, New York, 1982, pp. 323-335.
32. Colberre-Garapin, F., Horodniceanu, F., Kourilsky, P. and Garapin, A.C. J. Mol. Biol. 150:1-14, 1981.
33. Zenke, M., Grundstrom, T., Matthes, H., Winzerith, M., Schatz, C., Wildeman, A. and Chambon, P. EMBO J. 5:387-397, 1986.
34. Herbomel, P., Bourachot, B. and Yaniv, M. Cell 39:653-662, 1984.
35. Maione, R., Passananti, C., De Simone, V., Delli-Bovi, P., Augusti-Tocco, G. and Amati, P. EMBO J. 4:3215-3221, 1985.
36. Weiher, H., Konig, M. and Gruss, P. Science 219:626-631, 1983.
37. Veldman, G.M., Lupton, S. and Kamen, R. Mol. Cell. Biol. 5:649-658, 1985.
38. Hearing, P. and Shenk, T. Cell 33:695-703, 1983.
39. Gorman, C.M., Moffat, L.F. and Howard, B.H. Mol. Cell. Biol. 2:1044-1051, 1982.
40. Wong, G.G., Witch, J.S., Temple, P.A., Wilkins, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown. E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A. and Clark, S.C. Science 228:810-815, 1985.
41. Lanoix, J., Tseng, R.W. and Acheson, N.H. J. Virol. 58:733-742, 1986.
42. Izant, J.G. and Weintraub, H. Cell 36:1007-1015, 1984.
43. Kamen, R., Jat, P., Treisman, R. and Favaloro, J. J. Mol. Biol. 159:189-224, 1982.
44. Fenton, R.G. and Basilico, C. Proc. Natl. Acad. Sci. U.S.A. 79:7142-7146, 1982.

8

CHARACTERIZATION OF AN IMMUNOLOGICALLY DISTINCT POPULATION OF SIMIAN VIRUS 40 LARGE TUMOR ANTIGEN

Lori Covey, Perry Kwok, and Carol Prives*

Department of Biological Sciences, Columbia University, New York, New York 10027

ABSTRACT

We have previously described experiments suggesting that an immunologically distinct form of simian virus 40 (SV40) large tumor (T) antigen is involved in binding specifically to viral DNA. A unique monoclonal antibody, PAb 100, was shown to immunoprecipitate 10% of T antigen from infected cells, but greater than 60% of the T antigen DNA binding activity using an in vitro immunobinding assay. Moreover, in contrast to other T-antigen monoclonal antibodies, PAb 100 failed to recognize a mutant T antigen, C6, that cannot bind specifically to viral DNA. Experiments were performed to further understand the nature of the PAb 100 determinant on T antigen. We compared PAb 100 to another monoclonal antibody, PAb 416, which recognizes the vast majority of the immunoreactive T antigen in extracts of infected or transformed cells. Experiments were carried out to examine whether PAb 100 has a relatively weaker affinity for T antigen which may be stabilized either by binding of the protein to viral DNA or by the formation of T-antigen oligomers. It was found that different oligomeric forms of T antigen were recognized to similar extents by both PAb 100 and PAb 416 antibodies. Additionally, the relative quantities of T antigen, bound or not bound to DNA, that were immunoprecipitated by PAb 100 and PAb 416, were found not to differ significantly for either antibody under the conditions tested. These experiments suggest that the PAb 100 determinant is not necessarily stabilized by either oligomerization or by DNA binding. To further characterize

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

the PAb 100 epitope various treatments of T antigen were compared for their effects upon recognition by both monoclonal antibodies. Concentrations of sodium dodecyl sulfate that did not significantly affect the binding of PAb 416 to T antigen completely abolished the recognition of T antigen by PAb 100. However heating cell extracts to 30°C diminished recognition of T antigen by PAb 416 but not by PAb 100. Thus the PAb 100 class of T antigen, while denaturation sensitive, is more heat stable than the majority of T antigen, and therefore represents a unique subpopulation of this protein.

INTRODUCTION

Large T antigen, a product of the SV40 A gene, is a multifunctional protein with distinct roles in viral growth and transformation (for review, see 1). T antigen is required for the initiation and maintenance of transformation in nonpermissive cells (2-6), facilitating the growth of human adenovirus in monkey kidney cells (7-9), and the efficient synthesis of late capsid proteins (10). T antigen is also required for the initiation of each round of viral DNA replication (11-13), and the regulation of early transcription (14-16), processes which are directly linked to the protein's ability to bind to discrete sites at the viral origin of replication (ori) (17-20). Additional biochemical and genetic analyses have revealed that T antigen has an intrinsic ATPase activity (21,22) and that it binds to (23, 24) and stabilizes (25,26) the host-encoded 53K transformation-related protein. An active goal has been to establish how the various functions of this protein are determined at the level of domains within the protein, and in the context of the subpopulation of T-antigen molecules. It has been shown that T antigen exists as discrete oligomeric forms distinguished on the basis of distinct sedimentation properties (23,24,27-31). These forms most likely consist of monomers, dimers, tetramers, and higher multimetric forms. The possibility that these

forms may represent functional classes is reinforced by observed differences among them of degree of phosphorylation (28,32), extent of association with the p53 protein (29), and levels of ori-specific DNA binding (30,31).

The introduction and development of methods for producing monoclonal antibodies against specific proteins has facilitated the study of proteins and their role in regulation and development (33). The isolation of monoclonal antibodies specific for the SV40 tumor antigens has led to the identification of subclasses of T antigen based on conformational information (34-36). Monoclonal antibodies against the host p53 protein and subclasses of T antigen have been used successfully to probe the maturation of the T antigen-p53 complex (37). Antigenic determinants usually lie on the outside of a protein and may comprise either continuous stretches of amino acids or non-contiguous regions which are brought into close proximity by the three-dimensional folding of the molecule (for review see 38). Determinants which are dependent upon the native conformation of the protein are necessarily sensitive to denaturation while antibodies directed against small continuous segments of the whole protein are generally insensitive to denaturation. Two monoclonal antibodies directed against sequences encoded between 0.33 and 0.28 map units on the SV40 genome have been found to inhibit the ATPase activity of T antigen, suggesting that amino acids coded by these sequences may be directly involved in ATP hydrolysis (21). In contrast, antibodies have been synthesized against overlapping synthetic peptides which correspond to a region of the molecule containing the putative DNA binding domain. When T antigen was immunoprecipitated with these antibodies in the presence of SV40 DNA fragments it was found that the protein retained the ability to bind the SV40 ori region specifically (39). This suggests that the ori binding domain may be formed from non-contiguous segments of the molecule and is dependent on the three-dimensional structure of the

molecule.

We have previously reported the identification of an immunologically distinct subclass of T antigen which displayed a high ori binding activity. This subclass was bound specifically by PAb 100, a monoclonal antibody which failed to recognize a replication-defective mutant large T antigen, (C6) (40,41). We have also reported that the lack of recognition by PAb 100 antibody is tightly correlated to a single internal mutation in the T antigen molecule which renders the C6 non-functional for ori binding (42,43). In this paper we report the further characterization of the PAb 100 determinant on large T antigen in terms of denaturation and thermal stability, as well as the ability of PAb 100 to immunoprecipitate different forms of T antigen. In this set of experiments we have used as a control a monoclonal antibody, PAb 416, which precipitates most of the immunoreactive T antigen in the cell (36). Our results suggest that the types of T antigen determinants recognized by these monoclonal antibodies are intrinsically different and that the PAb 100-specific site may correlate with a T antigen conformation which is involved in origin binding.

MATERIALS AND METHODS

Cells and antibodies.

Cos 7 (44) and C6 (41) cells were grown as monolayers in Dulbecco's modified medium (DME) supplemented with 10% fetal calf serum (FCS). PAb 416 cells which were originally described by Harlow et al (36) were obtained from Ed Harlow. PAb 100 hybridoma cells which were isolated and described by Gurney et al (35) were purchased from American Type Tissue Culture. Both cell lines were grown in DME plus 20% FCS.

Preparation of nuclear extracts.

Extraction of Cos 7 and C6 cells was carried out essentially as described previously (40). Cell monolayers were washed three times with phosphate-buffered saline,

then the nuclei were isolated by the addition of 4 ml/100mm^2 culture plate of Nuclear Retention Buffer (NR) (10 mM NaCl, 20 mM 2[N-morpholino]ethanesulfonic acid [MES], and 1 mM $MgCl_2$[pH 6.0]). After incubation for 10 min the cells were scraped and homogenized in a Dounce homogenizer with 15 strokes of an A pestle. The cell extract was centrifuged for 5 min. at 2000 rpm and the pellet resuspended in 0.6 ml HIP buffer (0.15 M NaCl, 20 mM HEPES, pH 8.5, 1 mM $MgCl_2$, 0.5% Nonidet-P40) per 100 mm^2 dish. After 10 min incubation, the extract was centrifuged for 10 min at 3000 rpm. In some experiments, extraction of nuclei was carried out using 0.45 M NaCl in HIP buffer. If this was the case, the extract was diluted 3-fold with HIP buffer lacking NaCl prior to immunoprecipitation. Phenylmethysulfonyl flouride (PMSF) and 1-1-tosylamide-2-phenylethychloromethyl ketone (TPCK) were added to a concentration of 0.25 mg/ml to all buffers prior to use.

Immunoprecipitation of T antigen by monoclonal antibodies.

Cos 7 and C6 cells were labelled with [^{32}P]-orthophosphate (200 µCi/ml) or [^{35}S]-methionine (100-200 µLi/ml) in phosphate-free or methionine-free media respectively for 2 hr prior to extraction using the procedure described above. After centrifugation of the nuclear extracts for 45 min at 38,000 rpm, 100 µl aliquots of the supernatants (approximately 10^6 cells) were preabsorbed for 1 hr with FCS (10 µl), followed by addition of 50 µl formaldehyde-fixed Staphylococcus A bacteria (Staph A). The preabsorbed supernatant extracts were centrifuged (1 min at 8000 rpm), and the supernatant incubated with PAb 100 or PAb 416 hybridoma supernatant at an amount predetermined to be saturating (200 µl PAb 100/100 µl extract, 100 µl PAb 416/100 µl extract) for 1 hour at 0°C. Immune complexes were precipitated by the addition of 50 ul Staph A bacteria and incubated 30 min on ice. If the extract was immunoprecipitated with PAb 100, the Staph A was preincubated with rabbit anti-mouse antisera for 1 hr (PAb 100 is a class IgG_1 antibody (35)) followed by extensive washing with NET

buffer (0.15 M NaCl, 0.01 M Tris, 0.01 M EDTA, 0.05% NP40, [pH 7.5]) before complexing with the antibody-antigen complex. Bound complexes were collected by centrifugation of 3000 rpm and washed 3 times in NET buffer. T antigen was released from the complex in electrophoresis sample buffer (45), and analyzed by SDS-polyacrylamide gel electrophoresis (46) on a 12.5% separating gel.

RESULTS

T antigen exists as multiple oligomeric forms defined by sedimentation values in a linear sucrose gradient. It was reported that the slowly sedimenting form of T antigen (5-7S) is composed of newly synthesized molecules (24,28) which are underphosphorylated compared to the total T antigen population (24,49), and possess DNA binding activity (30-32). Since PAb 100 recognizes a putative functional subclass of T antigen which has a high DNA binding activity, it would be of interest to test whether this recognition is limited to a distinct protomeric form of the protein. [^{32}P]-labelled Cos 7 extracts were centrifuged through a 5-20% sucrose gradient, individual samples collected, and immunoprecipitated with either PAb 100 or PAb 416 hybridoma supernatant. Fig. 1A shows the sedimenting forms of T antigen recognized by PAb 100 (upper panel) and by PAb 416 (lower panel). Fractions 13-15 correspond to the 5-7S form and fractions 6-9 correspond to the 16S form. Both monoclonal antibodies immunoprecipitated both forms of T antigen. Although the 16S form is less apparent in the PAb 100 immunoprecipitates, it is proportionately similar in quantity to the 16S form recognized by PAb 416. Thus the PAb 100 determinant is most likely not altered by the complexing of T antigen with itself or with other cellular proteins. In addition, the presentation of multiple copies of the PAb 100 determinants (as would be predicted with oligomers of T antigen) does not increase the stability of the T antigen-PAb 100 complex.

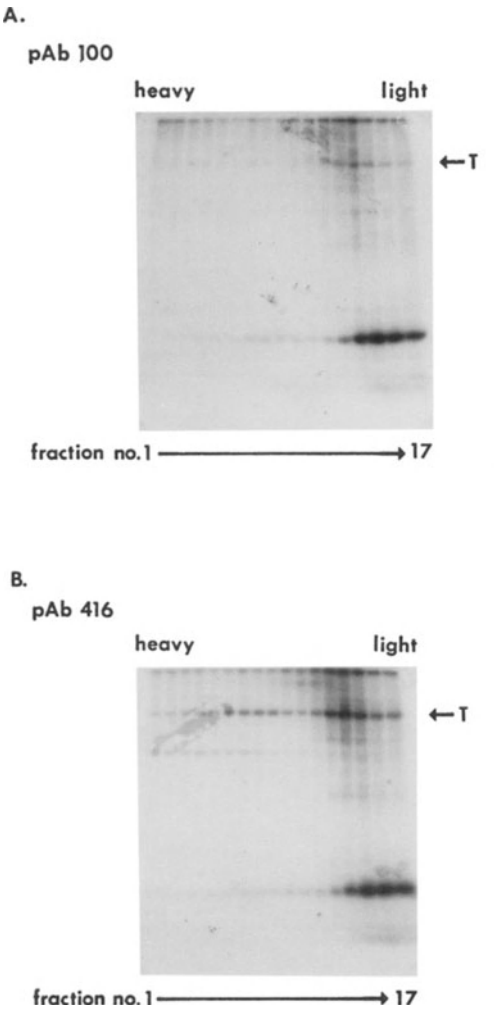


Fig. 1; Detection of various sedimenting forms of T antigen by PAb 100 and PAb 416.

Cos 7 were labeled with [^{32}P] and extracted as described in Materials and Methods. Extracts of approximately 4 x 10^6 cells were sedimented through 5-20% sucrose gradients (30), eighteen fractions collected, and one half of each fraction immunoprecipitated with either PAb 100 (A) or PAb 416 (B).

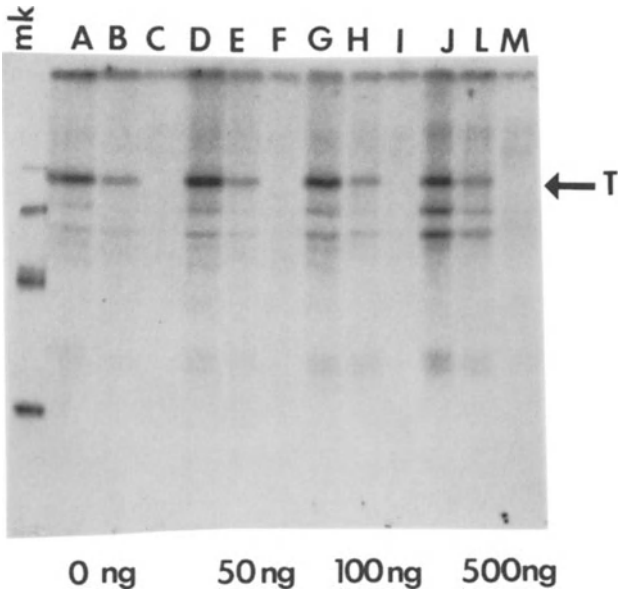


Fig. 2. Immunoprecipitation of T antigen in the presence of SV40 DNA fragments.

[^{32}P]-labeled extracts of Cos 7 cells were immunoprecipitated with either PAb 416 (lanes A,D,G,J), PAb 100 (lanes B,E,H,L), or DME+10% FCS (lanes C,F,I,M) in the presence of 0 (lanes A-C), 50 (lanes D-F), 100 (G-I), or 500 (J-M) nanograms of BstN-1 digested pSVR1 DNA under conditions previously established as optimal for ori-binding by T antigen (40). Polypeptides were separated on a 12.5% agarose gel and visualized by autoradiography.

PAb 100 recognizes approximately 10% of the cellular population of T antigen, yet it recognizes greater than 60% of the DNA binding activity (40). One explanation for this observation is that PAb 100 binds to T antigen with a much reduced affinity which is greatly increased once the T antigen has bound specifically to the ori region. To test this possibility [^{32}P]-labelled extracts of Cos 7 cells were incubated with various amounts of SV40 fragments digested with restriction enzyme BstNI (Fig. 2) under conditions previously shown to be optimal for in vitro binding, (Fig. 2). Over a 100-fold concentration range of exogenously added fragments, no substantial increase in the relative proportion of T antigen recognized by PAb 100 or PAb 416 was observed. Therefore, the stability of the PAb 100 T antigen complex did not appear to be influenced by specific binding to SV40 DNA.

In an effort to discern whether the PAb 100 determinant is denaturation stable we exposed T antigen-containing extracts from C6 or Cos 7 cells to 0.1% SDS for 1 hour, diluted the extract 10-fold, and after 30 min immunoprecipitated the extracts with either PAb 416 or PAb 100. Under conditions where the PAb 416 determinant remains stable (Fig. 3 lanes a' and b') the PAb 100-specific determinant is completely lost (compare Fig. 3 c' and d'). This result supports the contention that the PAb 100 determinant is effected by conformational changes in the molecule in comparison to the PAb 416 determinant which may be dictated strictly by the primary amino acid sequence. We have also observed that similar treatment of T antigen-containing extracts with 0.1% SDS abolished virus-specific DNA binding (not shown). In addition, our results show that less mutant C6 protein is recognized by PAb 416 after denaturation (Fig. 3 a and b).

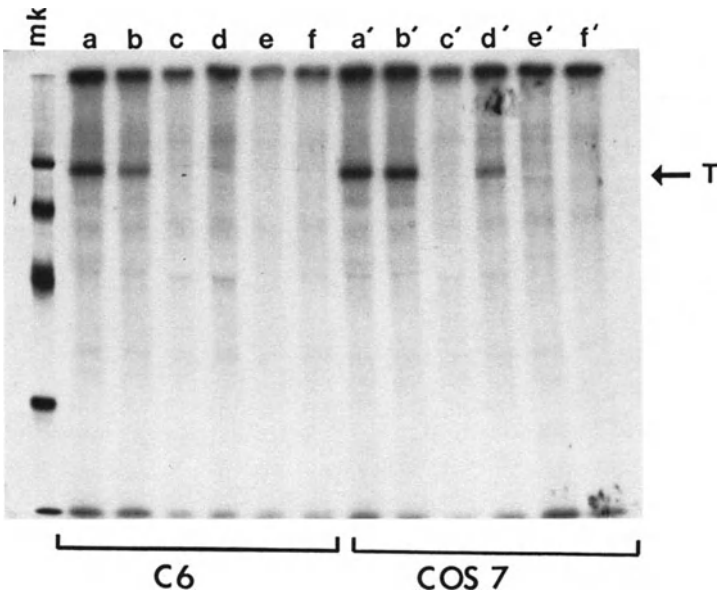


Fig 3. Analysis of the effects of ionic detergent on two T-antigen determinants.

Extracts of 10^6 [^{32}P]-labeled C6 (a-f) or Cos 7 (a'-f') cells were incubated for 1 hour at 0°C without further treatment (lanes a,c,e,b',d',f') or after addition of 10% sodium dodecyl sulfate (SDS) to a final concentration of 0.1% for 1 hr immediately followed by dilution of 0.01% SDS with Hip buffer (lanes b,d,f,a',c',e'). Samples were incubated with either PAb 416 (lanes a-b,a'-b'), PAb 100 (lanes c-d, c'-d') or DME+1% FCS (lanes e-f, e'-f') and immunoprecipitates were collected after incubation with rabbit anti-mouse-conjugated to protein A-Sepharose. Polypeptides were analyzed by SDS-polyacrylamide gel electrophoresis. The gels were dried and autoradiographed on Kodak SB-5 film.

T antigen has previously been shown to be a heat-labile protein (50). Its DNA binding property has been shown to be thermolabile as well (17,51). However, after limited exposure to increased temperature (30°C for up to 30 minutes) the *in vitro* binding of purified T antigen to the SV40 *ori* was shown to increase (52). This result suggested that if PAb 100 recognizes a DNA binding subclass of T antigen, an increase in the amount of PAb 100-recognizable T antigen might be obtained under similar conditions of elevated temperature. We incubated [^{32}P]-labelled extracts of Cos 7 cells at 30°C for varying times between 0 and 30 minutes. Immunoprecipitation of samples with either PAb 100 or PAb 416 was carried out at 4°C and proteins analyzed by gel electrophoresis (Fig. 4A). Over the indicated time period there was a distinct decrease in the amount of PAb 416-recognizable T antigen, while the absolute amount of T antigen recognized by PAb 100 remained relatively constant. Consequently, the ratio $T_{PAb\ 100}/T_{PAb\ 416}$ dramatically increased between 0 and 30 minutes (compare lanes a and g, to f and 1) suggesting a differential sensitivity of the two determinants to heat. However, because the experiments was performed with [^{32}P]-labelled proteins it is possible that the absolute amount of T antigen recognized by either monoclonal antibody was constant and that certain phosphorylated moities present on the PAb 416-specific population but absent from the PAb 100-specific population of T antigen were heat sensitive. Therefore an equivalent experiment using [^{35}S]-methionine-labelled protein was carried out. Results from this experiment were similar to those obtained with [^{32}P]-labelled T antigen (Fig. 4B). An increasing reduction in the amount of PAb 416-recognizable T antigen was obtained when extracts were heated to 30°C. In contrast, the amount of T antigen recognized by PAb 100 increased substantially over the same time period. This shows that the PAb 100 and PAb 416 determinants on the T Ag molecule differ in their response to conditions of elevated temperature.

The heat stability of the PAb 100 determinant could be

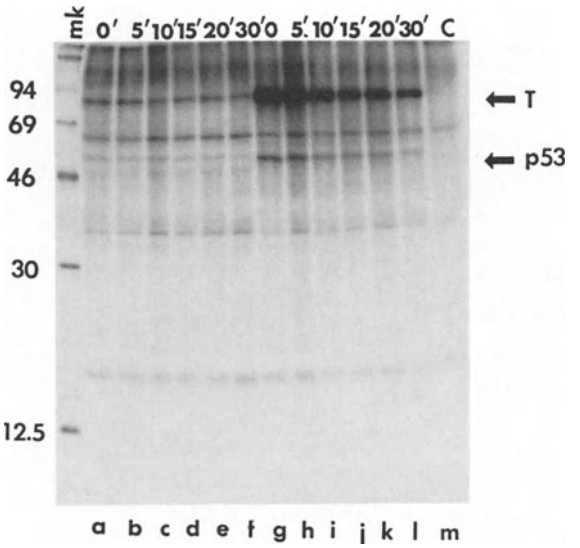


Figure 4A. Immunoprecipitation of [^{32}P]-labeled large T antigen after heating at 30°C.

[^{32}P]-labeled HIP extracts of Cos 7 cells were incubated at 30°C for increasing time intervals (minutes designated at top of gel) followed by immunoprecipitation with either PAb 100 (a-f) or PAb 416 (g-l). Immunoprecipitates were collected and polypeptides analyzed by SDS polyacrylamide gel electrophoresis. mk refer to molecular weight standards designated by numbers on left; C=control immunoprecipitation with DME+10% FCS; T=large T antigen; p53=cellular transformation-related protein.

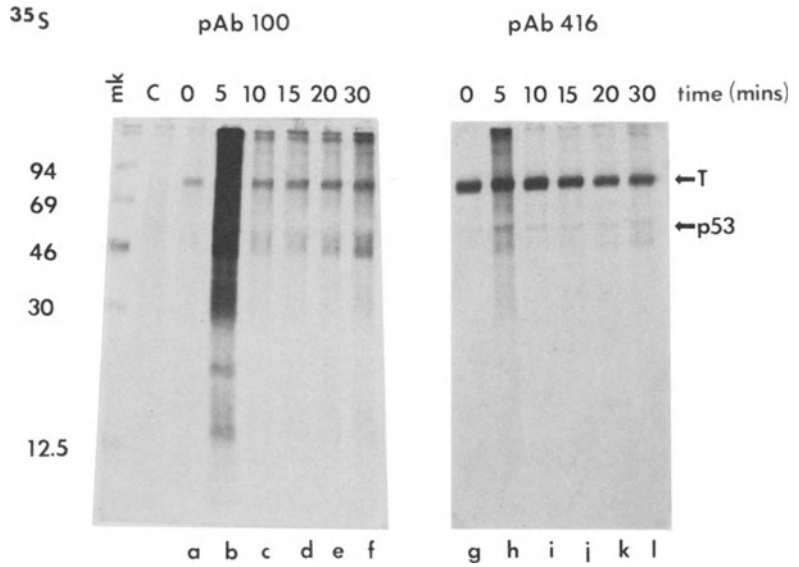


Figure 4B. Immunoprecipitation of [^{35}S]-methionine labeled large T antigen after incubation at 30°C.

[^{35}S]-methionine labeled Cos 7 extracts were incubated for increasing time intervals at 30°C (minutes at 30°C designated at top), then immunoprecipitated with either PAb 100 (lanes a-f) or PAb 416 (lanes g-l). Analysis of the proteins was determined by separation on a 12.5% polyacrylamide gel.

a result of an association of T antigen with other cellular proteins. T antigen is known to form tight complexes with and to stabilize the transformation-related protein, p53 (25,26). The possibility that this interaction is somehow responsible for the relative stability of the PAb 100 determinant to elevated temperature appears unlikely since the proportion of p53 immunoprecipitated with PAb 100 was always lower than that observed with other monoclonal antibodies (unpublished data). However, it is possible that other cellular proteins are involved in the stabilization of this determinant. To determine whether this was the case, T antigen was purified by immunoaffinity binding to a PAb 419-Sepharose column (47,48) and then heated at 30°C for different times followed by immunoprecipitation at 4°C with either PAb 416 or PAb 100 (Fig. 5). The immunoaffinity purified material contained three protein species as detected by [^{35}S]-methionine labelling (Fig. 5A) or by silver staining (not shown). These species were the large T antigen (seen as a doublet), the cellular p53 protein, and the small t antigen. As shown previously, PAb 416 and PAb 100 failed to immunoprecipitate the small tumor antigen suggesting that sequences required for recognition are contained in residues beyond amino acid 82 (35,36). Immunoprecipitating this relatively purified preparation of T antigen with PAb 100, after heating the extract did not affect the amount of T antigen recognized by this antibody (Fig. 5B). Surprisingly, the amount of T antigen recognized by PAb 416 actually increased slightly with heating. The results of this experiment suggested both that the heat stability of the PAb 100-determinant is not dependent on additional cellular proteins and that the lability of the PAb 416 determinant at 30°C in crude nuclear extracts results from factors which are purified away from T antigen by affinity chromatography. Furthermore, upon extended heating of immunoaffinity purified material small t antigen was immunoprecipitated by monoclonal antibody PAb 416. This observation may suggest either that sequences within the PAb 416 determinant are

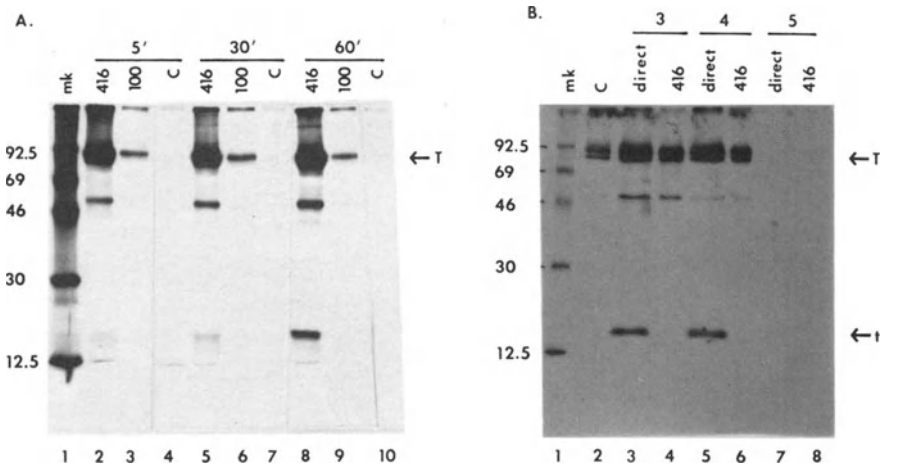


Figure 5. Immunoprecipitation of affinity-purified T antigen by PAb 416 and PAb 100 after exposure to elevated temperatures for varying times.

A. [^{35}S)-methionine-labelled T antigen was purified from labeled extracts of Cos 7 cells by immunoadsorption with PAb 419-conjugated protein A Sepharose as previously described. Aliquots from fraction 3 (shown in B) were heated for time periods designated above gel then immunoprecipitated with either PAb 416 (lanes 2,5,8), PAb 100 (lanes 3,6,9), or DME+10% FCS (lanes 4,7,10). Lane 1=mw standards indicated by size on left; T=Large T antigen. B. Elution profile of SV40 tumor antigens and the p53 cellular protein from 419-conjugated protein A-Sepharose column. Numbers at top refer to column fractions. Samples were immunoprecipitated with either PAb 416 (lanes 4,6,8) or DME+10% FCS (lane 2). Lanes 3,5,7 represent direct aliquots from each fraction.

found in the small t antigen but are available for PAb 416 binding only when the molecule is heated at 30°C or that small t antigen can associate with large T antigen under these conditions.

DISCUSSION

Our experiments indicate that two determinants on large T antigen which are recognized by specific monoclonal antibodies are fundamentally different. The PAb 100-specific determinant is easily denatured with low concentrations of ionic detergent suggesting that it is dependent on the native conformation of the protein. This determinant may therefore be assembled from residues far apart in the amino acid sequence that are brought together on the T antigen surface by the folding of the protein in its native conformation. This idea is reinforced by results of Deppert et al (53) in which PAb 100 was found to exhibit differential preferences for T antigen-related polypeptides coded by the Adenovirus 2-SV40 hybrid viruses, Ad2+ND1, ND2, and ND4, suggesting a role for conformation in binding by this antibody. In contrast, PAb 416, which has been previously shown to deplete the majority of the immunoreactive T antigen from extracts of SV40 transformed cells (36), recognizes a determinant which is resistant to detergent denaturation. The structure of this determinant, in relation to the topography of the T antigen molecule, may be determined by a short continuous amino acid sequence (but not necessarily involving contiguous residues in the segment). It was originally thought that sequences downstream from amino acid 82 (ie. unique to large T antigen) formed the PAb 416 determinant because PAb 416 did not immunoprecipitate small t Ag but did immunoprecipitate a T antigen variant produced by the mutant virus SVGT14BG that includes amino acid residues 1 through 272 (36). Surprisingly, PAb 416 also did not recognize the 107K Ad2+D2 protein which contains amino acid residues between 82 and 708 but lacks sequences common to small t antigen (36). Thus it is possible that amino acids upstream from residue 82 contribute to the binding of PAb 416 to T antigen. This idea is reinforced by

our observation that small t antigen can be immunoprecipitated by PAb 416 when heated at 30°C (Fig. 5A), suggesting that sequences in the region common to both large T and small t antigens may be normally exposed in the former but masked in the latter. Our results support the previous observation of others (54,55) that the NH_2-terminal regions are conformationally different in the two molecules.

The PAb 100 and PAb 416 determinants were found to be differentially sensitive to incubation at 30°C over varying lengths of time. This difference was not observed if the experiments were carried out using affinity purified T antigen. The PAb 100-determinant was found to be heat stable regardless of the sources of protein while PAb 416 recognized less T antigen heated to 30°C in crude extracts in its purified form. The relative heat stability of the PAb 100 determinant is consistent with previous observations that the DNA binding activity of T antigen is stimulated at 30°C (52). Other heat stable determinants that have been described previously for T antigen form a class referred to as U antigens (56-58). That the SV40-related 28K antigen coded for by Ad2+-ND1 is heat stable suggests that U antigen(s) reside(s) in the C-terminal 20% of the molecule. However, it seems unlikely that this is the determinant recognized by PAb 100, since it has been shown previously that the 28K hybrid protein is not recognized by PAb 100 (53). It is possible that common sequences which confer the thermostable phonotype are present in both the U and PAb 100 determinants or alternatively, that additional determinants in the molecule, at a distance from the COOH-terminus, are also heat stable.

PAb 100 was shown to interact with both the light (5-7S) and the heavy form (16-18S) of T antigen suggesting that the PAb 100 determinant was not formed by a T antigen oligomeric structure. However, caution must be used when interpreting these results since this gradient does not readily resolve the 5S (putative monomers) and 7S (putative dimers) forms and therefore it remains possible that PAb 100 requires a protein-protein interaction in order to bind T Ag (smallest

recognition unit would therefore be a dimer). What is clear is that recognition of large T antigen by PAb 100 is not limited to the more rapidly sedimenting forms and the pattern of immunoprecipitation across a gradient is similar to that found when PAb 416 is used as the antibody. This result is somewhat contradictory to two observations made previously. First, it has been shown that the ori binding activity of T antigen resides in the 5-7S form and more specifically in the 7S form (30). Second, PAb 100-specific T antigen is more highly phosphorylated than the majority of T antigen. It has been shown that newly synthesized T antigen, which is much more active in DNA binding, is underphosphorylated in comparison to the total T population (32). One possible explanation could be that under the conditions used for extraction and immunoprecipitation (pH 8.5) the T Ag which is recognized by PAb 100 is easily modified by phosphorylation. However, if immunoprecipitation is carried out at a lower pH (pH 6.8 is the optimal pH for in vitro (binding) then this modification does not occur. We have previously examined the quantity of T antigen precipitated by PAb 100 under conditions where 60% of the DNA binding activity was precipitated by this monoclonal. At the lower pH, while we did not observe an increase in the total amount of PAb 100-specific T antigen, the number of other non-specifically bound proteins was greatly increased (data not shown). Whether these proteins somehow facilitate the binding of the PAb 100-recognizable T antigen to DNA will be determined in future experiments using affinity purified T antigen in a DNA solution binding assay with PAb 100 as antibody or alternatively by testing PAb 100 immunopurified T antigen directly for ori-specific binding.

The PAb 100-subclass of T antigen remains intriguing because of its possible functional significance although the precise nature of the subclass remains elusive. PAb 100 does not recognize the C6-2 mutant T antigen that cannot bind to the SV40 origin of replication and which is also non-functional in viral replication (41,42). This lesion

replaces a Thr for an Asn and thus could potentially introduce a new phosphorylation site into a region previously shown to be unphosphorylated. If this region is important in establishing a three-dimensional conformation which can be recognized by PAb 100 then an additional phosphorylation site in this region could dramatically alter the existing determinant. Studies of the structure and function of T antigen purified using PAb 100 affinity columns may help to establish more completely the nature of the PAb 100 determinant and the subclass of T antigen recognized by this antibody.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service Grants CA26905 and CA33620 from the U.S. National Institutes of Health.

REFERENCES

1. Tooze, J., J. Molecular Biology of Tumor Viruses. Part 2, 2nd Ed. Cold Spring Harbor Laboratory, New York, 1981.
2. Tegtmeyer, P. J. Virol. 15: 613-618, 1975.
3. Brugge, J.S. and Butel, J.S. J. Virol. 15: 619-635. 1975.
4. Martin, R.G. and Chou, J.Y. J. Virol. 15: 599-612, 1975.
5. Osborn, M. and Weber, K. J. Virol. 15: 636-644, 1975.
6. Kimura, G., and Itagaki, A. Proc. Natl. Acad. Sci. USA 72: 6730677, 1975.
7. Friedman, M.P., Lyons, M.J. and Ginsberg, H.S. J. Virol. 5: 589-597, 1970.
8. Kimura, G. Nature (London) 248: 590-592, 1974.
9. Grodzicker, T., Anderson, C., Sambrook, J., and Mathews, M.B. J. Virol. 19: 559-571, 1977.
10. Tornow, J., Polvino-Bodnar, M., Santangelo, G. and Cole, C.N. J. Virol. 53: 415-424, 1985.
11. Tegtmeyer, P. J. Virol. 10: 591-598, 1972.
12. Chou, J.Y. and Martin, R.G. J. Virol. 13: 1101-1109, 1974.
13. Tegtmeyer, P., Schwartz, M., Collins, J.K. and Rundell, K. Virol. 16: 168-178, 1975.
14. Reed, S.I., Start, G.R. and Alwine, J.C. Proc. Natl. Acad. Sci. USA. 73: 3083-3087, 1976.
15. Alwine, J.C., Reed, S.I. and Stark, G.R. J. Virol. 24: 22-27, 1977.
16. Reed, S.I., Ferguson, R., Davis, R.W. and Stark, G.R. Proc. Natl. Acad. Sci. U.S.A. 72: 1605-1609, 1975.
17. Jessel, D., Landau, T., Hudson, J., Lalor, T., Tenen, D. and Livingston, D.M. Cell 8: 535-545, 1976.
18. Tjian, R. Cell 13: 165-179, 1978.

19. Shortle, D.R., Margolskee, R.F. and Nathans, D. Proc. Natl. Acad. Sci. USA 76: 6128-6131, 1979.
20. Rio, D., Robbins, A., Myers, R. and Tjian, R. Proc. Natl. Acad. Sci. USA 77: 5706-5710, 1980.
21. Clark, R., Lane, D. and Tjian, R. J. Biol. Chem. 256: 11854-11858, 1981.
22. Giacherio, D. and Hager, L.P. J. Biol. Chem. 254: 8113-8116, 1979.
23. McCormick, F. and Harlow, E. J. Virol. 34: 213-224, 1980.
24. Greenspan, D.S. and Carroll, R.B. Proc. Natl. Acad. Sci. USA 78: 105-109, 1981.
25. Linzer, D.I.H. and Levine, A.J. Cell 17: 43-52, 1979.
26. Oren, M., Maltzman, N. and Levine, A.J. Mol. Cell. Biol. 1: 101-110, 1981.
27. Prives, C., Beck, Y., Gidoni, D., Oren, M. and Shure, H. Cold Spring Harbor Symp. Quant. Biol. 44: 123-130, 1980.
28. Fanning, E., Nowak, B. and Burger, C. J. Virol. 37: 92-102, 1981.
29. Melero, J.A. Sitt, D.T., Mangel, W.F. and Carrol, R.B. Virology 93: 466-480, 1979.
30. Gidoni, D., Scheller, A., Barnet, B., Hantzopoulos, P., Oren, M. and Prives, C. J. Virol. 42: 456-466, 1982.
31. Bradley, M.K., Griffin, J.D. and Livingston, D.M. Cell 28; 125-134, 1982.
32. Scheidtmann, K.H., Hardung, M., Echle, G. and Walter, G. J. Virol. 50: 1-12, 1984.
33. Kohler, G. and Milstein, C. Nature (London) 256: 495-497, 1975.
34. Martinis, J. and Croce, C.M. Proc. Natl. Acad. Sci. USA 75: 2320-2323, 1978.
35. Gurney, E.G., Harrison, R.O. and Fenno, J. J. Virol. 34: 752-763, 1980.
36. Harlow, E. Crawford, L.V., Pim, D.C. and Williamson, N.M. J. Virol. 39: 861-868, 1981.
37. Carrol, R. and Gurney, E.G. J. Virol. 44: 565-573, 1982.
38. Benjamin, D.C. Berzofsky, J.A., East, I.J. and Gurd, F.R.N., Hannum, C., Leach, S.J., Margoliash, E., Michael, J.G., Miller, A., Prager, E.M. Reichlin, M., Sercarz, E.C., Smith-Gill, S.J., Todd, P.E. and Wilson, A.C. Ann. Rev. Immunol. 2: 67-101, 1984.
39. Paucha, E., Harvey, R. and Smith, A.E. J. Virol. 51: 670-680, 1984.
40. Scheller, A., Covey, L., Barnet, B. and Prives, C. Cell 29: 375-383, 1982.
41. Prives, C., Covey, L., Scheller, A. and Gluzman, Y. Mol. Cell. Biol. 3: 1958-1966, 1983.
42. Gluzman, Y. Davison, J. Oren, J. and Winocour, E. Virol. 22: 256-266, 1977.
43. Gluzman, Y. and Ahrens, B. Virology 123: 78-92, 1982.
44. Gluzman, Y. Cell 23: 175-182, 1981.
45. Prives, C., Beck, Y. and Shure, H. J. Virol. 33: 689-696, 1980.
46. Laemmli, U.K. Nature (London) 227: 680-685, 1970.
47. Simanis, B. and Lane, D.P. Virology 144: 88-105, 1985.

48. Dixon, R.A. and Nathans, D. J. Virol. 53: 1001-1004, 1985.
49. Montenarh, M. and Henning, R. FEBS Lett 114:107-110, 1980.
50. Gilden, R. Carp, R., Taguchi, F. and Defendi, V. Proc. Natl. Acad. Sci. USA 53: 684-692, 1965.
51. Tenen, D.G., Martin, R.G., Anderson, J.L. and Livingston, D.M. J. Virol. 22: 210-218, 1977.
52. Tenen, D.G., Taylor, T.S., Haines, L.L., Bradley, M.K., Martin, R.G. and Livingston, D.M. J. Mol. Biol. 168: 791-808, 1983.
53. Deppert, W., Gurney, E.G. and Harrison, R.O. J. Virol 37: 478-482, 1981.
54. Smith, A.E., Smith, R. and Paucha, E. J. Virol. 28: 140-153, 1978.
55. Walter, G., Scheidtmann, K., Carbone, A., Laudana, A.P. and Doolittle, R. Proc. Natl. Acad. Sci. USA 77: 5191-5200, 1980.
56. Deppert, W. J. Virol. 29: 576-586, 1979.
57. Lewis, A.M. Jr. and Rowe, W.P. J. Virol. 7: 189-197, 1973.
58. Robb, J.A. Proc. Natl. Acad. Sci. USA 74: 447-451, 1977.

9

THE SIMIAN VIRUS 40 AGNOPROTEIN

SUSAN CARSWELL and JAMES C. ALWINE

Department of Microbiology, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6076

ABSTRACT

The 61 amino acid agnoprotein of simian virus 40 (SV40) is encoded between nucleotides 335 and 523 within the leader region of late RNAs. A variety of evidence has suggested that agnoprotein may function in the viral assembly process or the control of late gene expression. We review evidence which suggests that the agnoprotein facilitates the perinuclear/nuclear localization of the major virion structural protein, VP1.

INTRODUCTION

The agnoprotein of SV40 is a 61 amino acid protein encoded between SV40 nucleotides 335 and 523 within the leader region of some late SV40 mRNAs (See Figure 1). Its existence was first postulated after DNA sequence analysis revealed an open reading frame for which no protein had been detected (1; see Fig. 1). Subsequently, Jay et al. (2) detected the agnoprotein in lytically infected African green monkey kidney (AGMK). Using polyacryamide gel electrophoresis of ^{14}C-labeled proteins, a band migrating at the appropriate relative molecular weight (7,900 M_r) was detected in wild type SV40 infected cells but not in uninfected cells. Definitive proof that this was the agnoprotein was provided by deletion mutants of the late leader which showed either no agnoprotein or altered peptides migrating at molecular weights predictable from the sizes of the deletions (2).

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

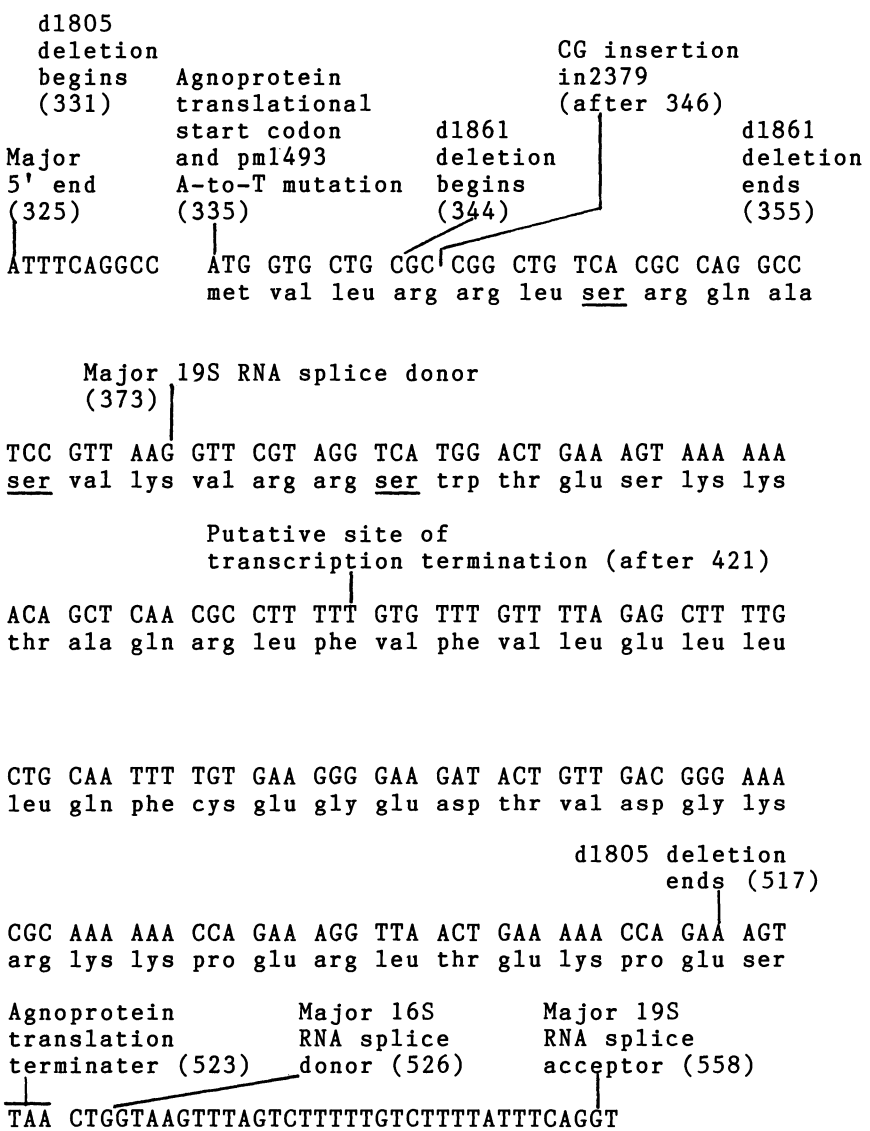


FIGURE 1. SV40 agnogene and late leader sequences. The DNA nucleotide sequences of the SV40 late RNA leader region and agnogene are shown in capital letters. The sites of control elements and mutations discussed in the text are indicated above the nucleotide sequences. Nucleotide numbers (numbering system in ref. 45) are in parentheses. Below the nucleotide sequences are the amino acid sequences of the agnoprotein. The underlined serines are discussed in the text.

The agnoprotein is highly basic, containing approximately 25% arginine plus lysine residues (See Fig. 1), and has the ability to bind double-stranded and single-stranded DNA in vitro (2). The mature agnoprotein contains no methionine residues; apparently the initial methionine is cleaved shortly after translation. Time course studies of labeled proteins from infected cells show that the agnoprotein has a half-life of no more than 2-3 hours (2,3) and is produced at high rates in monkey cells at relatively late times in the viral lytic cycle (2,4,5).

Using anti-agnoprotein serum for immunofluorescence and subcellular fractionation studies, Nomura, Khoury and Jay (6) showed that the intracellular location of the agnoprotein is predominately in the cytoplasmic and perinuclear regions of infected cells. In addition, these workers reported that a small fraction of agnoprotein made in infected cells could be detected in the nucleus. An intranuclear location has also been reported by Jackson and Chalkley (4). Their studies of formaldehyde-fixed protein-DNA complexes indicate that agnoprotein may be associated with virion capsid proteins and replicating DNA. However, agnoprotein has not been detected in mature virions (2,4).

Several groups have attempted to define the function of the agnoprotein by characterizing viral mutants with deletions and insertions in the agnogene (5,7-22). Mutants which produce no agnoprotein form smaller plaques and grow to lower final yields compared to wild type (WT). This suggests that the agnoprotein provides a "non-essential" function which enhances viral growth in tissue culture cells. However, these mutants vary in their degrees of defectiveness in ways that cannot easily be explained by simple abolition of agnoprotein function. For instance, Nomura, Jay and Khoury (18) showed that in2379, a mutant constructed with a 2-base pair insertion at nucleotide 346, creating a frameshift in the agnoprotein coding sequences (See Fig. 1), grows more poorly than a mutant in which the entire agnogene is deleted. When propagated in AGMK cells, the insertion mutant tends to generate second-site mutations, many of which are deletions that remove the agnoprotein translational start codon. Seemingly, then, these findings suggest

that either the agnoprotein is multifunctional or other functions, in addition to the agnoprotein loss, are disrupted when late leader sequences are altered. This latter possibility is quite reasonable when one considers the many regulatory elements within or near the agnogene, including late promoter elements (23-28), sites affecting 5'-end start sites for late mRNAs (10,19,29), late mRNA splice donor and acceptor sequences (30), and possible attenuation signals (3,5,31-37) (See Fig. 1).

The best characterized agnogene deletion mutant is d1805, a spontaneously occurring mutant which deletes SV40 nucleotides 331 to 517, essentially the entire agnogene (See Fig. 1). Despite small plaque size and reduced yield of d1805, examination of various aspects of its gene expression and growth show essentially wild type behavior. For example, the time of initiation and rates of synthesis of viral DNA, early and late RNAs and viral proteins are very similar to WT (7,17; our unpublished observations; J. Resnick and T. Shenk, personal communication). Additionally, virions formed in d1805-infected cells are as stable as WT virus particles (17). However, analyses of intermediate structures in the process of virion assembly revealed that virions seemed to assemble more slowly in infections with d1805 and a similar mutant, d1810, compared to WT SV40 (17). These data suggest that the agnoprotein mutants may have defects in a post-translational aspect of viral development, possibly assembly. This idea was supported by the isolation of spontaneously occurring large plaque revertants of d1805 which were characterized by Barkan and Mertz (Barkan, Ph.D. thesis, Univ. of Wisconsin, 1983). These viruses were found to have second site point mutations that map within the carboxyl terminal region of VP1, the major capsid protein. Conversely, Margolskee and Nathans (38) isolated viruses with point mutations in VP1 which form small plaques and yield reduced virus titers. Spontaneously occurring revertants of these VP1 mutants produce WT-sized plaques and yield WT titers. All such revertants have acquired second-site point mutations in the agnogene. The effect of these agnogene mutations is to change one of the three serine codons at the amino terminal end of the agnoprotein (underlined in

amino acid sequences in Fig. 1) to a hydrophobic amino acid. Taken together, these data suggest that VP1 and agnoprotein interact to enhance virus assembly and production.

Several studies have suggested that agnoprotein may function in late transcriptional control. Studies of mixed infections using WT SV40 and d1861, an agnogene mutant deleting 12 base pairs near the amino terminal end of the agnoprotein (See Fig. 1), suggest a possible role for the agnogene and/or the agnoprotein in late transcriptional regulation (39). Infections using each strain alone resulted in similar steady-state levels of late RNA. Yet, in mixed infections, levels of late RNA were reduced in WT and were nearly undetectable in the mutant. A model proposed consistent with this data is that a trans-acting factor, possibly the agnoprotein, acts on sequences within the agnogene to increase levels of late RNA.

From results of studies assessing transcription products produced in isolated nuclei, Aloni and coworkers (3,5,31-37) also hypothesized a potential role for the agnoprotein in late transcription regulation, although their data suggested that the agnoprotein may be a negative, rather than a positive, effector of late transcription. They observed the production of a small RNA mapping to the late leader region. This RNA seemed to accumulate during times of maximal agnoprotein synthesis (5) and was not detected in nuclei infected with the mutant which has a 2-base pair insertion at nucleotide 346 rendering it unable to make authentic agnoprotein (31) (This mutation is the same one constructed into in2379 shown in Fig. 1 and discussed above). Aloni and co-worker noted that the sequences in the late leader of SV40 share structural similarities with some bacterial operons (31), and, consequently, proposed that a mechanism analogous to bacterial attenuation, which is possibly mediated through the agnoprotein, is operative in SV40 late gene regulation. According to this model, late leader RNA sequences can assume two conformations, each having extensive secondary structure. In the absence of the agnoprotein, the secondary structure which predominates allows translation of agnoprotein. As agnoprotein concentration increases, however,

agnoprotein binds to late leader RNA sequences in cytoplasmic 16S RNA and thereby causes the other secondary structure to predominate. When the RNA is in this alternate conformation, the start codon for agnoprotein is sequestered in a stem-loop structure, such that it is unavailable for translation. Such a strategy would then allow preferential translation of VP1, which is encoded in the distal portion of the 16S RNA molecule. The model also proposes that agnoprotein concomitantly enters the nucleus and binds to the late leader sequences of nascent RNA, thereby favoring the formation of a secondary structure which activates a transcriptional termination signal. This leads to premature termination of 16S RNA and formation of the small RNA mentioned above.

Clearly the data discussed thus far indicates a role for agnoprotein in the lytic cycle. However, this role is shrouded in the complexities of the cis-acting elements within, and surrounding, the late leader (agnogene) region and the possible trans-acting nature of the agnoprotein. In order to study agnoprotein function more directly, we separated the putative trans-activities of the protein from the cis-acting elements within and near the agnogene sequences in the virus. We achieved this by constructing cell lines (40) which constitutively express the agnoprotein (and no other viral protein) and then assessed the effects of these cell lines on WT and agnoprotein-minus mutants of SV40 (40,41). We review this data below.

EXAMINATION OF THE FUNCTION OF THE AGNOPROTEIN

Monkey cell lines which constitutively express the agnoprotein.

We undertook to discriminate the trans activities of the agnoprotein from cis-acting elements which overlap the agnogene by constructing CV-1 cell lines (CV-1 is an established AGMK line and a permissive host for SV40) in which the agnogene is stably integrated and constitutively expressed. This was accomplished by stable transfection of CV-1 cells with a plasmid expressing the agnogene under the control of the Rouse sarcoma virus long terminal repeat (RSV-LTR; 40). This plasmid was constructed by inserting

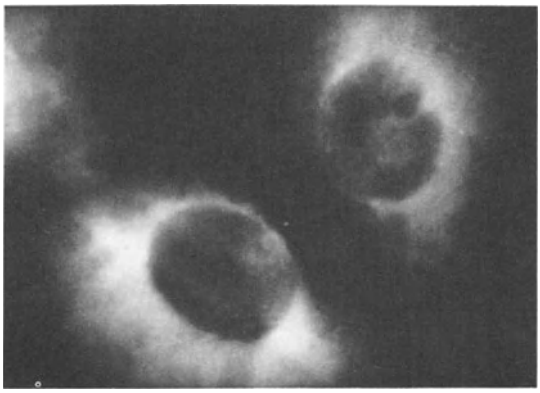


Figure 2. Subcellular localization of agnoprotein in a WT SV40 infection. Indirect immunofluorescence using anti-agnoprotein serum (a gift from G. Jay) of WT SV40 infected CV-1P cells at 48 hpi shows the predominantly cytoplasmic and perinuclear localization of agnoprotein.

	Mean Plaque Size	
Cell line	WT SV40	Agnoprotein-minus mutants
CV1-P	normal	pinpoint
Ag 18	normal	normal
Ag 8	pinpoint	none detected*

Table 1. Effect of agnoprotein-expressing cell lines on plaque sizes of WT SV40 and agnoprotein-minus mutants. Mean plaque sizes measured on day 13 after infection. WT SV40 and agnoprotein-minus mutants d1805 and pm1493 were grown on cell lines indicated. The same results were obtained using either of the two agnoprotein-minus mutants. Normal plaque size is 2.5 - 3 mm on day 13. *Although no plaques were observed by day 13 on agnoprotein-minus mutant infected Ag 8 cells, plaques did form at later times (40).

SV40 nucleotides 270 to 555 into the unique Hind III site of pRSV-0, creating pRSV-agno (40). Following standard cotransfection with a plasmid able to impart neomycin (G-418) resistance to eucaryotic cells (pRSVneo; 42,43), we demonstrated that the agnogene was expressed in several G-418 resistant cell lines (40). In six cell lines, typified by cell line Ag 18, relatively low levels of agnoprotein were expressed, while one line, Ag 8, produced at least ten times more agnoprotein compared to Ag 18 (40). It should be noted, however, that the level of agnoprotein produced in Ag 8 cells, the "high" agnoprotein producing cell line, is significantly lower than the level attained at late times in a productive WT infection. This is not surprising considering that the stable transfectant would have very many fewer gene copies compared to a lytically infected cell at late times in infection. Indirect immunofluorescence assays of Ag 8 and Ag 18 cells using anti-agnoprotein antisera demonstrated that the agnoprotein produced in the cell lines localizes to the cytoplasmic and perinuclear regions of the cell, its site of localization in WT infected cells (6). Figure 2 is a representative micrograph showing this agnoprotein distribution in WT virus infected CV-1 cells.

Wild type and agnoprotein-minus mutant growth in agnoprotein-producing cell lines.

The agnoprotein-expressing cell lines were first tested for their ability to complement the small plaque phenotype of agnoprotein-minus mutants. Two types of mutants were tested; one, d1805 (the gift of J. Mertz; described in the Introduction), deletes nucleotides 331 to 517 which includes almost all of agnogene (See Fig. 1). The other mutant, pm1493, constructed by J. Resnick and T. Shenk, has an A-to-T base substitution at nucleotide 335, which changes the start codon for the agnoprotein from an ATG to a TTG (See Fig. 1) (personal communication). Thus, all putative cis-acting sequences within the agnogene are eliminated in d1805, but remain nearly intact in pm1493. These mutants were specifically selected to enable us to determine the effects, if any, of the agnoprotein from the cell lines on agnogene sequences in the virus. For example, if agnoprotein does execute a function

by interacting with sequences within the agnogene, then we would expect to see an effect on pm1463, where these intragenic sequences are intact, but not in d1805, in which the sequences are deleted. A brief summary of our previously published (40) results are shown in Table 1. We found that the low agnoprotein-expressing cell lines, such as Ag 18, complemented the small plaque phenotype of both agnoprotein-minus mutants. Plaques of the mutants grown on these cells were approximately the size of WT SV40 plaques formed on normal control cells. Unexpectedly, WT virus formed somewhat larger plaques on Ag 18 cells than on control cells. Oddly, however, plaques of both mutant and WT SV40 grown on the high agnoprotein expressing cell line, Ag 8, were smaller than those formed on control cells. Overall these results suggest: 1) that varying the intracellular concentration of agnoprotein present from the beginning of the infectious cycle creates very different effects on the outcome of the infection; 2) that the effect of the agnoprotein is the same regardless of whether or not the infecting virus contains the putative cis-active element(s) within the late leader/agnogene region.

Agnoprotein facilitates perinuclear/nuclear localization of VP1.

To determine the molecular basis for these plaquing phenomena, we performed time course studies of viral DNA, RNA and late protein synthesis in WT SV40 and agnoprotein-minus mutant infections of CV-1, Ag 8 and AG 18 cells. No striking differences in time of initiation of synthesis or steady state levels were observed under any of these conditions. We then examined post-translational events, specifically the fate of the viral structural proteins, VP1, VP2 and VP3, using indirect immunofluorescence time course studies (41). One antiserum used was directed against VP1, the major capsid protein, and the other was against the minor structural proteins, VP2 and VP3, which share amino acid sequences and, thus, antigenic determinants (Both antisera were gifts from H. Kasamatsu). In data to be presented elsewhere (41), we determined that in WT infected CV-1 cells, VP1 partitioned predominantly to the cytoplasm until about 40 hours post-infection (hpi), after which it quite rapidly accumulated first in the perinuclear region

and then in the nucleus. In contrast, VP1 in agnoprotein-minus mutant infected CV-1 cells tended to accumulate initially in the cytoplasm and later in the perinuclear regions until very late times in infection (e.g., 60 hpi), after which it slowly accumulated in the nucleus. These data strongly suggest that the agnoprotein-minus mutants have a difficulty in the efficient localization of VP1 to the nucleus. This is supported by the observation of temporal correlation between time of perinuclear/ nuclear localization of VP1 and time of maximal agnoprotein synthesis (approximately 40 hpi).

Examination of VP1 nuclear localization in infections of agnoprotein producing cell lines with agnoprotein-minus mutants also indicates a role of agnoprotein in the efficient nuclear localization of VP1. In infections of the agnoprotein expressing cell lines with either of the agnoprotein-minus mutants or with WT SV40, VP1 was detected largely in the nucleus at 30 hpi, the earliest time that it could be detected in these assays (41) and continued to be detected only in the nucleus throughout the infectious cycle. This early nuclear entry of VP1 was more pronounced in Ag 8 cells than in Ag 18 cells, probably reflecting the relatively higher concentrations of agnoprotein in the former. Overall it seems likely that the early nuclear localization in the agnoprotein producing cell lines is due to the constitutive expression of agnoprotein and thus its presence at earlier times in the lytic cycle than it would be in a WT infection of normal CV-1 cells.

In analogous studies using the anti-VP2, VP3 serum, it was observed that VP2 and VP3 localized to the nucleus at the earliest times they could be detected in both WT and agnoprotein-minus mutant infected CV-1 cells (41). Seemingly, then, agnoprotein is not involved with the nuclear localization of these minor capsid components. Very similar results and conclusions regarding the role of the agnoprotein in nuclear localization of VP1 have been obtained by J. Resnick and T. Shenk (personal communication).

DISCUSSION

The sequences within the leader region of SV40 late genes encodes not only the agnoprotein, but also several regulatory elements for late gene expression. Studies of this region of the genome are often complicated by the presence of these overlapping domains, since it is difficult to discern whether observed effects are attributable to the agnoprotein, to cis-acting signals within or near the agnogene, or to both. A review of past literature reveals that the only reasonably certain facts known about agnoprotein function is that it enhances plaque size and increases final virus yields. Other data has suggested that the agnoprotein may be involved in the mediation of late gene expression through interaction with sequences within the late leader/agnogene region (3,5,31-37,39). Although these data strongly indicate that elements which modulate late gene expression are located within the late leader/agnogene sequences, it is only speculative that the agnoprotein may interact with these sequences to affect gene expression in <u>trans</u>. On the contrary, the facts that agnoprotein localizes almost exclusively outside of the nucleus (6) and that agnoprotein-minus mutants produce near WT levels of late mRNA and proteins (17, Carswell and Alwine, unpublished observation; Resnick and Shenk, personal communication) suggest that regulation of late gene expression is not its primary activity. Our results further suggest that agnoprotein does not mediate its function through interaction with the leader region/agnogene sequences. This is concluded because the agnoprotein from the agnoprotein-expressing cell lines demonstrated no differential effects between the mutant which had deleted the entire agnogene region (d1805) and the agnoprotein-minus point mutant (pm1493), comparing late RNA levels, plaque size and nuclear localization of VP1 (40,41). If the agnoprotein acted on its own sequences to alter these parameters, only growth of the point mutant could have been affected.

The bulk of evidence favors the view that agnoprotein, through possible interaction with VP1, mediates an event which occurs after the translation of structural proteins. This is shown in studies indicating that viral assembly is defective in agnoprotein-minus

mutants (17); in genetic analyses which show that VP1 point mutants generate spontaneous pseudorevertants with second site mutations in the agnogene (38); and that agnogene pseudorevertants have second site VP1 mutations (A. Barkan, Ph.D. Thesis, Univ. of Wisconsin, 1983). Our findings (41), and those of Resnick and Shenk (personal communication), showing that VP1 localization to the nuclear region is delayed in agnoprotein-minus mutants and that viruses with this defect are complemented by growth in agnoprotein-expressing cell lines strongly suggest a role for the agnoprotein in a post-translational event, namely the perinuclear/nuclear targeting of VP1.

SV40 may use a strategy of agnoprotein-facilitated nuclear entry of VP1 to prevent prematurely early accumulation of capsids in the nucleus. Until the time that agnoprotein synthesis becomes maximal, or possibly until agnoprotein has had sufficient time to exert its effect(s), VP1 remains predominantly cytoplasmic. Only when the virus-infected cell becomes engaged primarily in virion assembly does most VP1 enter the nucleus. It is easy to envision that, without this safeguard, virion assembly would occur early in the late phase of the growth cycle. Under these conditions, viral DNA could be packaged before it had been effectively replicated and before optimal levels of structural proteins had been synthesized. Overall, the expected effect of premature nuclear entry of VP1 would be to reduce virus yields. It is also possible that nuclear accumulation of high levels of VP1, which in vitro is known to form insoluble aggregates (44), impairs nuclear functions. Thus, the virus can achieve optimal macromolecular synthesis and, in turn, higher virus titers by delaying nuclear entry of VP1 until the time that the cell is committed to viral assembly. The curious inhibitory effect that the high agnoprotein expressing cell line, Ag 8, has on plaquing of WT and mutant SV40 may, in fact, be attributable to the abnormally early perinuclear/nuclear localization of high levels of VP1, thus prematurely packaging viral DNA and/or inhibiting viral macromolecular synthesis.

Although the agnoprotein may have additional functions in the lytic cycle, present knowledge strongly suggests that an important

role for it is in facilitating the perinuclear/nuclear localization of VP1 in order to initiate virion assembly at the appropriate time in the lytic cycle.

ACKNOWLEDGEMENTS

This work was supported by Public Health Service Grants CA28379 and CA33656 awarded by the National Cancer Institute to J.C.A. S.C. was supported by a Damon Runyon postdoctoral fellowship. We thank J. Resnick and T. Shenk for communicating data prior to publication.

REFERENCES

1. Dhar, R., Subramanian, K.N., Pan, J. and Weissman, S.M. Proc. Natl. Acad. Sci. USA 74:827-831, 1977.
2. Jay, G., Nomura, S., Anderson, C.W. and Khoury, G. Nature (London) 291:346-349, 1981.
3. Hay, N., Kessler, M. and Aloni, Y. Virol. 137:160-170, 1984.
4. Jackson, V. and Chalkey, R. Proc. Natl. Acad. Sci. USA 78:6081-6085, 1981.
5. Hay, N. and Aloni, Y. Mol. and Cell. Biol. 5:1327-1334, 1985.
6. Nomura, S., Khoury, G. and Jay, G. J. Virol. 45:428-433, 1983.
7. Barkan, A. and Mertz, J.E. J. Virol. 37:730-737, 1981.
8. Brockman, W.W., and Nathans, D. Proc. Natl. Acad. Sci. USA 71:942-946, 1974.
9. Carbon, J., Shenk, T.E. and Berg, P. Proc. Natl. Acad. Sci. USA 72:1392-1396, 1975.
10. Ghosh, P.K., Piatak, M., Mertz, J.E., Weissman, S.M. and Lebowitz, P. J. Virol. 44:610-624, 1982.
11. Ghosh, P.K., Roy, P., Barkan, A., Mertz, J.E., Weissman, S.M. and Lebowitz P. Proc. Natl. Acad. Sci. USA 78:1386-1390, 1981.
12. Haegeman, G., Iserentant, D., Gheysen,D. and Fiers, W. Nucleic Acids Res. 7:1799-1814, 1979.
13. Haegeman, G., Van Heuverswyn, H., Gheysen, D. and Fiers, W. J. Virol. 31:484-493, 1979.
14. Mertz, J. E. and Berg, P. Virology 62:112-124, 1974.
15. Mertz, J.E. and Berg, P. Proc. Natl. Acad. Sci. USA 71:4879-4883, 1974.
16. Mertz, J.E., Murphy, A. and Barkan A. J. Virol. 45:36-46, 1983.
17. Ng, S.-C., Mertz, J.E., Sanden-Will, S. and Bina, M. J. Biol. Chem. 260:1127-1132, 1985.
18. Nomura, S., Jay, G. and Khoury, G. J. Virol 58:165-172, 1986.
19. Piatak, M., Subramanian, K.N., Roy, P. and Weissman, S.M. J. Mol. Biol. 153:589-618, 1981.
20. Shenk, T.E., Carbon, J. and Berg, P. J. Virol. 18:664-671, 1976.

21. Subramanian, K.N. Proc. Natl. Acad. Sci. USA 76:2556-2560, 1979.
22. Villarreal, L.P., White, R.T. and Berg P. J. Virol. 29:209-219, 1979.
23. Alwine, J.C. and Picardi, J. J. Virol. In press, 1986.
24. Brady, J. and Khoury, G. Mol. Cell. Biol. 5:1391-1399, 1985.
25. Ernoult-Lange, M., May, P., Moreau, P. and May E. J. Virol. 50:163-173, 1984.
26. Hartzell, S.W., Byrne, B.J. and Subramanian, K.N. Proc. Natl. Acad. Sci. USA. 81:6335-6339, 1984.
27. Keller, J.M. and Alwine, J.C. Mol. Cell. Biol. 5:1859-1869, 1985.
27A. Keller, J.M. and Alwine, J.C. Cell 36:381-389, 1984.
28. Omilli, F., Ernoult-Lange, M., Borde, J. and May, E. Mol. Cell. Biol. 6:1875-1885, 1986.
29. Piatak, M., Ghosh, P.K., Norkin, L.C. and Weissman, S.M. J. Virol. 48:503-520, 1983.
30. Lai, C.J., Dhar, R. and Khoury, G. Cell 14:971-982, 1978.
31. Hay, N., Skolnik-David, H. and Aloni, Y. Cell 29:183-193, 1982.
32. Skolnik-David, H., Hay, N. and Aloni Y. Proc. Natl. Acad. Sci. USA 79:2743-2747, 1983.
33. Pfeiffer, P., Hay, N., Pruzan, R., Jakobivits, E.B. and Aloni, Y. EMBO J. 2:185-191, 1983.
34. Skolnik-David, H. and Aloni Y. EMBO J. 2:179-184, 1983.
35. Hay, N. and Aloni, Y. Nucleic Acids Res. 12:1401-1414, 1984.
36. Abulafia, R., Ben-Zeev, A., Hay, N. and Aloni, Y. J. Mol. Biol. 172:467-487, 1984.
37. Aloni, Y. and Hay, N. Mol. Biol. Rep. 9:91-100, 1983.
38. Margolskee, R.F. and Nathans, D. J. Virol. 48:405-409, 1983.
39. Alwine, J.C. J. Virol. 42:798-803, 1982.
40. Carswell, S. and Alwine, J.C. J. Virol. In press, 1986.
41. Carswell, S. and Alwine, J.C. J. Virol. Submitted for publication.
42. Gorman, C.M., Padmanabhan, R. and Howard, B.H. Science 221:551-553, 1983.
43. Southern, P.J. and Berg, P. J. Appl. Genet. 1:327-341, 1982.
44. Christensen, M. and Rachmeler, M. Virol. 75:433-437, 1976.
45. Buchman, A.R., Burnett, H.L. and Berg, P. In: DNA Tumor Viruses: Molecular Biology of Tumor Viruses, 2nd Edition (Ed. J. Tooze) Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 779-829.

10

SV40 CHROMATIN STRUCTURE

WALTER A. SCOTT

Department of Biochemistry, The University of Miami School of Medicine, Miami, Florida 33101, USA

ABSTRACT

Some features of the 80S nucleoprotein extracted from the nuclei of SV40-infected cells have been explored. A region of the SV40 genome including the origin of replication and about 400 base pairs of adjacent sequence is contained in an altered chromatin structure, rendering it hypersensitive to endonuclease digestion and accessible to exonuclease digestion once an endonucleolytic cut has been introduced. DNA sequences within this region are responsible for the nuclease-sensitive configuration. At least two genetic elements, each of which can generate a hypersensitive site, have been mapped to the region between nucleotide positions 37 and 287. Insertion of segments of DNA at position 37/38 does not disrupt the hypersensitive site but moves it away from the origin of replication. Insertion of 90 base pairs or more causes the origin to lie outside the hypersensitive site and the resulting genome replicates inefficiently. This result suggests that efficient functioning of the viral replication origin may depend on location within a hypersensitive site in the chromatin. The same may be true for viral promoters although the genetic evidence is incomplete.

INTRODUCTION

It is now well established that most eukaryotic DNA occurs in combination with histones and other nuclear proteins in a quasi-repeat structure termed a nucleosome. In addition, short DNA segments, 200 to 400 base pairs (bp) in length -- often near the 5'-ends of transcription units -- appear to be nucleosome-free and to take on an altered chromatin configuration detected by their hypersensitivity to DNase I and other endonucleases. This phenomenon was initially discovered in SV40 chromatin

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

(1-3) and has since been extended to a wide variety of organisms and genes (4-6). Identification of structural and genetic factors responsible for the nuclease-hypersensitive feature of chromatin is essential to a complete understanding of gene function in eukaryotes.

Papovavirus-infected cells provide unique experimental advantages in the study of chromatin structure. Replication and expression of viral genes occur while the viral genome is separate from cellular chromosomes, allowing the investigator to extract viral chromatin as a physically distinct entity for direct biochemical and immunological analysis. Viral DNA replication, up to a level of 100,000 copies per cell, provides an amplified source of chromatin molecules containing precisely-defined DNA sequence. This article will be concerned with studies carried out in my laboratory on the chromatin structure which forms over SV40 DNA sequences. Papovaviruses can serve as vectors for the introduction and amplification of segments of cellular DNA so that, in principle, the same analysis can be applied to any short segment of a eukaryotic genome.

EXPERIMENTAL

Isolation of SV40 minichromosomes from the nuclei of infected cells.

When nuclei from SV40-infected cells are gently extracted with buffer, two nucleoprotein species containing SV40 DNA are released, sedimenting at approximately 80S and 200S (Fig. 1). The 80S nucleoprotein has a structure which is typical of cellular chromatin -- a nucleosome ladder is observed after staphylococcal nuclease digestion (7) and structures with the dimensions of nucleosomes can be visualized in the electron microscope (8,9). The 200S species corresponds to provirions or intracellular virus particles. (Sedimentation coefficients have been reported between 50 and 80S for SV40 chromatin and up to 250S for provirions. The values depend on the treatment prior to centrifugation and on the sedimentation standards used for comparison.)

Extraction of cells or nuclei with buffer containing Triton X-100 and EDTA (the preferred extraction procedure until about 1979) yields substantial amounts of heterogeneously sedimenting nucleoprotein containing high molecular weight DNA and the SV40 DNA in the 80S peak shows signs of nucleolytic damage (Fig. 1D). Several laboratories (10-13) have developed milder extraction procedures to avoid breakdown of the provirion complexes.

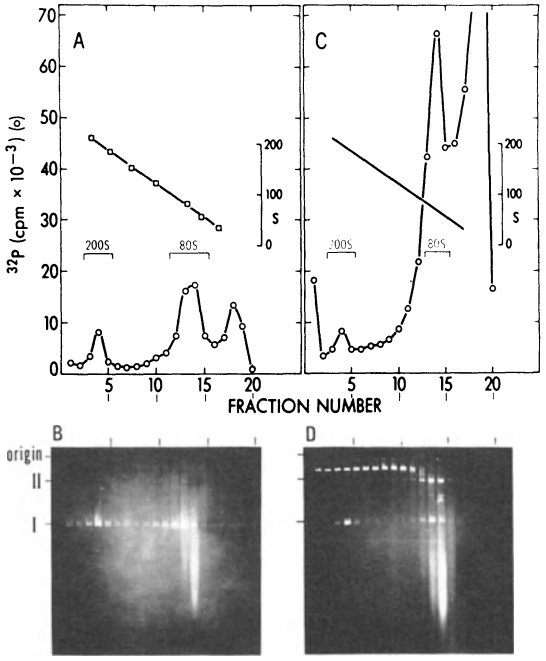


Fig. 1. Sedimentation of nucleoproteins extracted from nuclei of SV40-infected cells. BSC-1 cells were infected with wild type SV40 (strain 776), labeled 24 to 42 hr post-infection with ^{32}P-phosphate, and nuclear extracts were prepared by the isotonic method of Fernandez-Munoz et al. (11) (A,B) or by treatment with Triton X-100 and EDTA (referred to as Triton-EDTA extraction, ref. 1) (C,D). Extracts were fractionated by centrifugation into a 15 to 30% sucrose gradient in isotonic buffer. An aliquot from each fraction was subjected to electrophoresis on 1.4% agarose and stained with ethidium bromide. Sedimentation standards (□) include ribosomal subunits, monosomes and polysomes prepared from uninfected BSC-1 cells. I and II indicate supercoiled-circular and relaxed-circular DNA. (Panel A was reprinted with permission from ref. 14).

Use of one of these procedures is illustrated in Fig. 1A and B. Nucleolytic damage to both cellular and viral DNA is avoided -- presumably because endogenous nuclease known to be present in nuclear extracts is inactive in isotonic salt concentration. Using this procedure 42 hr after infection, each 15 cm culture dish of infected BSC-1 cells yields about 3 µg of SV40 DNA in the 80S peak of viral chromatin.

Detection of the nuclease-hypersensitive site in 80S nucleoprotein.

Weintraub and Groudine (15) initially described the use of endonucleases to probe for differences in structure from one region of chromatin to another. We decided to use a similar approach to ask whether differences in chromatin structure could be detected on the SV40 genome (1). We extracted SV40 chromatin from the nuclei of infected cells and carried out a brief incubation with DNase I or with an endonuclease activity which is released from BSC-1 cells during Triton-EDTA extraction. We isolated full-length linear viral DNA by preparative gel electrophoresis and redigested it with restriction enzymes to map the sites of the initial cut introduced by endonuclease digestion. (Figs. 2 and 3).

These nucleases showed a strong preference for cleavage of the minichromosome in the region between nucleotides 5200 and 400 (SV numbering system, ref. 16). Thirty to 50% of the chromatin molecules were cleaved within this limited region whereas cleavage of purified SV40 DNA by these enzymes showed no preference for this region. As can be seen in Fig. 2, the

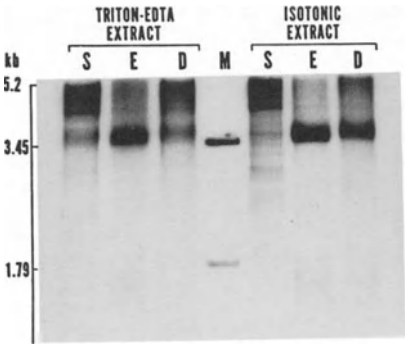


Fig. 2. Location of sites in SV40 chromatin which are preferentially cleaved by endonucleases. 80S nucleoprotein prepared as described in Fig. 1 was incubated with staphylococcal nuclease at 5°C (S), with nuclear extract prepared by the Triton-EDTA procedure from uninfected BSC-1 cells at 37°C (E), or with DNase I at 37°C (D) under conditions which gave maximal yield of full-length linear SV40 DNA. Full-length linear SV40 DNA was isolated by preparative gel electrophoresis and redigested with EcoRI. The resulting DNA fragments were fractionated by electrophoresis on 1.4% agarose. (Adapted with permission from ref. 14).

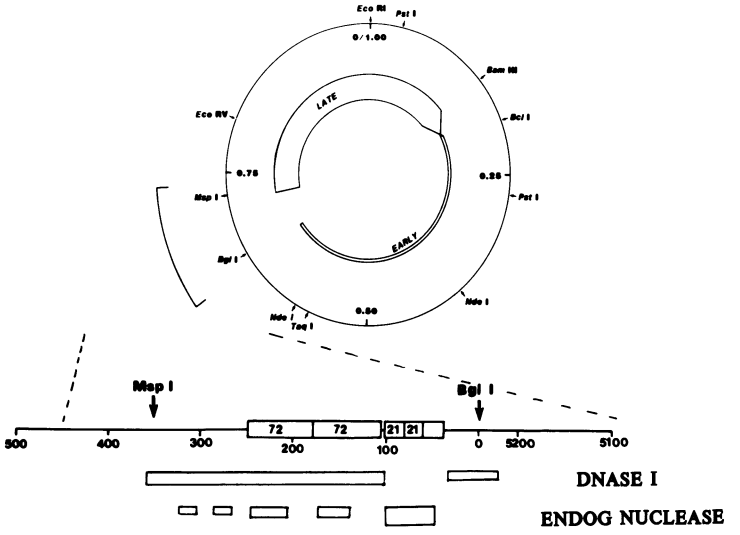


Fig. 3. Map of nuclease-sensitive region of SV40 chromatin. Cleavage patterns for DNase I and endogenous endonuclease represent a composite of many experiments. The DNase I cleavage pattern is also described by Saragosti et al. (17). (Adapted with permission from ref. 18).

specificity of cleavage depends somewhat on the method for isolating the 80S complex and on the enzyme used for analysis; however, we have evidence, as will become apparent in the next section, that at least two distinct subpopulations of viral nucleoprotein exist in the 80S peak which differ with respect to nucleoprotein structure in the region near the origin.

Preference for cleavage in the 5200 to 400 region was also seen with DNase II and with nuclease S1 (when incubated at pH 5.5 but not when incubated at pH 4.5 [Amin and Scott, unpublished]). In our hands, staphylococcal nuclease showed no preference for cleavage in the origin region in chromatin prepared by the isotonic extraction method. When the Triton-EDTA procedure was used, some preference for the origin region was seen; however, this could reflect action of endogenous nuclease during the extraction. (This disagrees with a report by Sundin and Varshavsky [19], probably due to differences in conditions for chromatin isolation or incubation with the enzyme.) Within the region between nucleotide positions

5200 and 400, DNase I and endogenous nuclease cut at different locations (Fig. 3) as do S1 and DNase II (not shown). Presumably each of these enzymes is directed to the nuclease-sensitive region by some aspect of the chromatin structure, but the precise sites where they cut is controlled by steric factors and sequence preferences of the enzymes.

Most of our studies have been done with 80S nucleoprotein which includes molecules involved in transcription and replication. We have also been interested in looking at the nucleoprotein structure which occurs in virus particles, since this represents an inactive state of the chromatin. Disruption of SV40 virions yields core particles which have some properties of SV40 minichromosomes (20-22) although by electron microscopy, the structure does not consist of a typical nucleosome repeat (23). DNase I does not preferentially digest virion cores (or core particles prepared from provirions) in the 5200 to 400 region suggesting that the nuclease-hypersensitive site does not occur in these nucleoproteins (14,24). These results suggest that the hypersensitive site is limited to intracellular chromatin. It presumably plays a role in the function of that structure in the cell.

Barriers to exonuclease digestion in the nuclease-sensitive chromatin structure.

We found that we could obtain additional information about chromatin structure by cleaving within the nuclease-sensitive region with a single-cut restriction enzyme and then digesting the exposed termini with nuclease Bal31 (18). If initial cleavage was carried out with MspI, Bal31 digested only a short distance in the late direction, but some molecules were degraded 400 to 500 base pairs in the early direction (Fig. 4). An analogous result was obtained if the initial cleavage was carried out with BglI -- Bal31 degraded only a short distance in the early direction, but some molecules were degraded 400 to 450 bp in the late direction. These barriers to Bal31 digestion defined the borders of the same nuclease-sensitive region that was identified with DNase I and the endogenous nuclease.

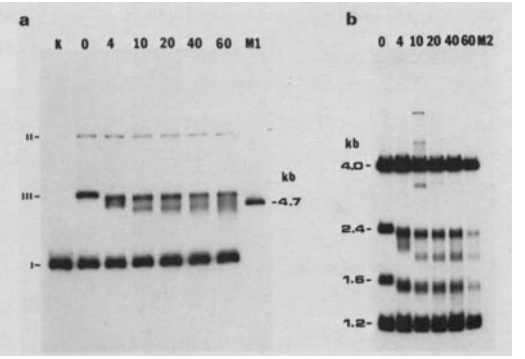


Fig. 4. ^{32}P-labeled SV40 chromatin was digested with MspI for 10 min at 37°C, then NaCl and $CaCl_2$ were added to final concentrations of 250 and 10 mM, respectively. Bal31 was added and the incubation was continued at 30°C. At the times indicated (min), samples were taken, enzyme digestion was stopped by adding sodium dodecyl sulfate and EDTA (a). An additional portion of each sample was treated with RNase A and the DNA was isolated and digested with PstI (b). Electrophoresis in each case was on 1.4% agarose. M1 and M2 are molecular weight standards. I, II and III designate supercoiled-circular, relaxed-circular, and full-length linear DNA, respectively. (Reprinted with permission from ref. 18).

It is clear from gel analysis of the molecules which were cleaved with MspI followed by Bal31 that the SV40 chromatin behaves as if it consists of at least two distinct populations of molecules. One population is shortened by only about 100 bp during exonuclease digestion while the second is shortened by about 550 bp. These subpopulations are always present although the ratio between them varies from preparation to preparation. They have slightly different sedimentation characteristics as evidenced by the fact that fractions from the upper side of the 80S peak are enriched for molecules which are shortened by about 550 bp by Bal31 digestion and those from the lower side of the 80S peak are enriched for molecules shortened by only about 100 bp by Bal31 digestion. This sedimentation difference may reflect a different number of nucleosomes in the two subpopulations. This heterogeneity in our best preparations of SV40 chromatin makes definitive structural analysis complicated.

Analysis of the barriers to Bal31 digestion at higher resolution (Fig. 5) shows that the borders of the nuclease-sensitive region appear as clusters of stop sites for exonuclease digestion rather than as one discrete barrier. The same collection of stop sites (spaced at about 10-bp intervals) was reached whether Bal31 digestion began at the BglI site or at the MspI site. This pattern is reminiscent of patterns obtained by exonuclease digestion of bulk nucleosomes (25-27); however, a contribution of the DNA sequence in this region cannot be excluded. The structures which define the borders of the hypersensitive region may be typical nucleosomes which are held in position by higher order nucleoprotein structure; they may be specialized nucleosomes having unique structural features yet to be defined; or they may be structures unrelated to nucleosomes. As will be shown below, their positioning is controlled by DNA sequences some distance away, within the hypersensitive region.

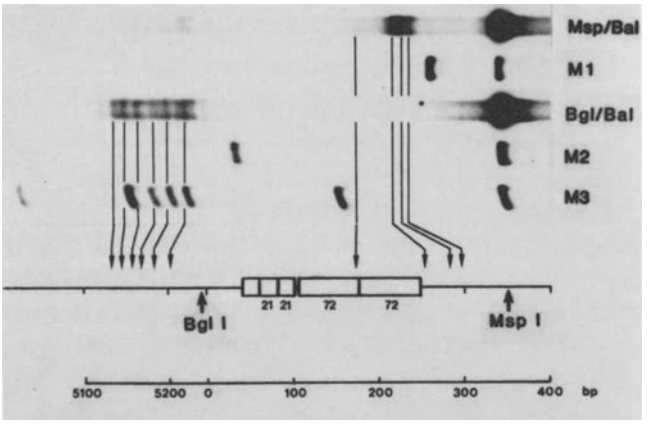


Fig. 5. Stop sites for Bal31 digestion from MspI and BglI sites in SV40 chromatin. ^{32}P-labeled SV40 chromatin was digested with MspI or BglI and Bal31. The DNA was isolated as described in Fig. 4 and further digested with NdeI to map stop sites for Bal31 digestion in the early direction from the initial cleavage sites (electrophoresis was on 4% polyacrylamide). M1, M2 and M3 are molecular weight standards. (Reprinted by permission from ref. 18)

In summary, analysis of SV40 chromatin by cleavage with endonucleases and with exonucleases has identified a nuclease-sensitive region which is precisely positioned over the two SV40 promoters and the origin of replication. The outer borders of this region have been defined by barriers to exonuclease digestion and they bracket the same region which is sensitive to both endogenous nuclease and DNase I. This region extends in the early direction across the origin of replication to a position lying within the leader sequences for early transcripts and about 400 bp in the late direction from the origin to a position lying within the leader sequences for late transcripts. Within the nuclease-sensitive region, specific cleavage sites are different for each endonuclease. As a result, the specific cleavage patterns are difficult to interpret in structural terms. The borders of the nuclease-sensitive region occur as clusters of stop sites for exonuclease -- the structural significance of which is also difficult to interpret unambiguously.

Genetic analysis of the hypersensitive site in SV40 chromatin.

Use of double-origin mutants. Because the nuclease-sensitive feature depends on the protein component of SV40 chromatin and maps uniquely on the viral genome, it was reasonable to propose that specific DNA sequences were responsible. Initially we looked at a number of naturally-occurring partially-duplicated variants of SV40 to see if they contained one or two hypersensitive sites (28,29). In these studies it quickly became apparent that the primary determinants for the nuclease-sensitive feature mapped near the region that was nuclease sensitive.

A double-origin mutant [in(Or)1411] constructed by T. Shenk (30) (Fig. 6) has been particularly useful. To create this mutant, a 273-bp segment of the genome containing the origin of replication and flanking sequences but lacking one copy of the 72-bp repeat was inserted into the HpaI site at nucleotide 2666. The resulting virus can be grown without a helper -- the insertion is at a location which does not disrupt either early or late genes. The genome is kept within the size limit for packaging into virions because of a deletion in the large T intron and deletion of one copy of the 72-bp repeat. Infection of BSC-1 cells with this variant yields chromatin with two nuclease-sensitive sites of about equal accessibility to DNase I suggesting that cis-acting sequences important to the generation of the hypersensitive site are contained within the 273 bp insertion.

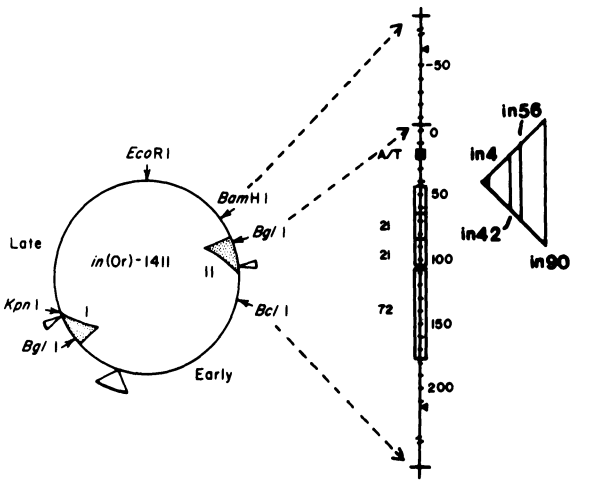


Fig. 6. Map of in(Or)1411 (30). The inserted segment in in(Or)1411 consists of 273 bp (nucleotide 5185 to 286, but lacking one copy of the 72 bp repeat) inserted at the HpaI site (nucleotide 2666). Stipled regions correspond to duplicated segments (designated I and II for later reference). Wedge-shaped extensions from the circle indicate deleted sequences by comparison with wild-type SV40. Numbering on the expanded section of the map is in base pairs from the center of the 27-bp pallindrome. Insertion mutations were prepared by ligating an AluI digest of wild-type SV40 DNA into the NcoI site (nucleotide 37/38) in a cloned segment of in(Or)1411. For the sequences of the inserted segments see ref. 31. (Adapted with permission from ref. 32).

This provided us with a genome which we could modify to further define the role of specific sequences in the formation of this chromatin structure. The fact that in(Or)1411 is viable was important. Mutations which we introduced into the inserted segment (region II in Fig. 6) should also be viable and we could clone them by plaque isolation. This served to screen out mutations in other portions of the genome since nearly all the rest of the mutant genome is essential for viability. We generated deletion mutants in the nuclease-sensitive region by digesting in(Or)1411 chromatin with endonuclease to target the site of initial cleavage. Linear DNA was isolated, transfected into BSC-1 cells, and the resulting plaques were screened for deletions which are presumably generated by action of cellular exonucleases and ligases after the DNA enters the cell (33). Subsequently, we cloned a segment of in(Or)1411 containing region II into a derivative of pBR322 and constructed mutants in E. coli (34). The altered segment was then

excised from the plasmid, recombined with the remaining portion of in(Or)1411 which had also been cloned in a plasmid, and transfected into BSC-1 cells to produce virus.

Insertion mutations between nucleotides 37 and 38. Figure 6 shows a set of insertion mutants in which segments of DNA were inserted into the cleavage site for NcoI. The sequence of inserted DNA was different in each case (31), but the effect on the chromatin structure was similar (Fig. 7). The pattern of cleavage sites by endogenous endonuclease was unchanged in the insertion mutant chromatin except that it was shifted in the late direction with respect to the origin of replication. The same result was seen with DNase I (not shown). This means that cleavage occurred at the same locations in the sequence on the late side of position 37/38, but at different sequences on the early side. These results imply that all of the features of the hypersensitive site are controlled by sequences on the late side of nucleotide 37 and that, by some mechanism, these sequences can keep nucleosomes from entering a region spanning about 80 base pairs on the early side of that position (including the origin of replication).

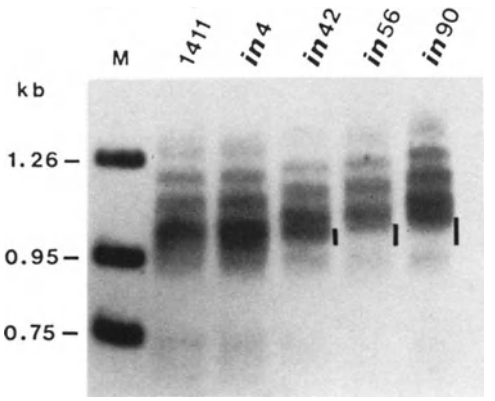


Fig. 7. Distribution of cleavage sites by an endonuclease activity endogenous to BSC-1 cells in in(Or)1411 and insertion mutant chromatin. Full-length linear viral DNA generated by endogenous endonuclease cleavage of viral chromatin was isolated by preparative gel electrophoresis, redigested with EcoRI, fractionated by electrophoresis on 2% agarose, blotted to nitrocellulose, and hybridized with ^{32}P-labeled probe (PstI to BamHI fragment; nucleotides 1988 to 2534 in the SV40 genome). Dark vertical bars to the right of some lanes indicate the positions of the inserted DNA segments in the cleavage patterns. (Reprinted with permission from ref. 31).

Effect of insertion mutations on DNA replication. A plasmid containing the inserted segment in in(Or)1411 can replicate in COS 1 cells since a replication origin is contained within that segment. The insertion mutations described above were each tested as plasmid clones to determine whether DNA replication in COS 1 cells was altered (31). Replication was substantially reduced in those mutants which contained an insertion 90 bp or longer (insertions of up to 390 bp were tested). The fact that short insertions had little effect suggested that no signal critical for replication had been disrupted. In mutants with insertions of 90 bp or more, the replication origin is no longer contained within a nuclease-sensitive chromatin structure. This may result in suboptimal function of the replication origin because of reduced access by protein factors necessary for initiation of replication.

Deletion mutations within region II. Deletion mutations which inactivate the origin of replication have no detectable effect on the nuclease-sensitive chromatin structure (29,33) indicating that the sequences in the core origin region are not responsible for this feature. This is in agreement with the conclusion drawn above that the sequences responsible for this structure map on the late side of nucleotide 37/38.

A number of mutants have been isolated with deletions in region II and have provided insight into the role of DNA sequences in forming the nuclease-sensitive site (32,33). Analysis of four mutants is shown in Fig. 8. In each case, virus stocks containing the mutant genomes were used to infect BSC-1 cells, virus chromatin was extracted and incubated with divalent cations to allow endogenous endonuclease to introduce one cut into a majority of the chromatin molecules. Full-length linear DNA was isolated and the position of cleavage was mapped by redigestion with EcoRI and gel electrophoresis. Each mutant had effects which extended over the entire nuclease-sensitive region. Although details in the cleavage pattern differed between mutants, the most useful information was obtained by comparing the frequency of cleavage in region II with that in the unmodified region I.

Analysis of these and other mutants allows us to draw several conclusions.

(i) The origin of replication and the AT-rich region alone are not sufficient for anything but a very weak hypersensitive site.

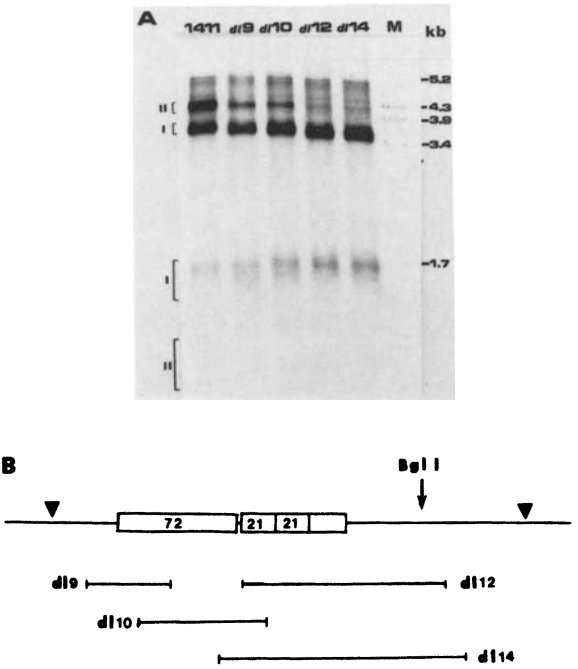


Fig. 8. Distribution of endogenous endonuclease cleavage sites in in(Or)1411 and deletion mutants. A) Full-length linear viral DNA generated by endogenous endonuclease cleavage of ^{32}P-labeled viral chromatin was redigested with EcoRI and fractionated by electrophoresis on 1.4% agarose. I and II refer to DNA fragments generated by initial cleavage within nuclease-sensitive regions I and II in the in(Or)1411 genome (Fig. 6). B) Segments of region II deleted in the mutants analyzed in (A). (Reprinted with permission from ref. 33).

(ii) Sequences in the distal part of the enhancer region are important for the nuclease-sensitive chromatin structure and these sequences have an effect extending to the BglI site -- a distance of about 180 bp.

(iii) No single short deletion results in complete loss of nuclease sensitivity suggesting that more than one genetic element is responsible for this chromatin structure. Longer deletions which result in complete or nearly complete loss of this feature (such as dl12 and dl14 shown in Fig. 8) may do so because they lack multiple genetic elements which can each generate a nuclease-sensitive structure.

(iv) Although sequences at various positions within the nuclease-

sensitive region clearly contribute to the formation of the hypersensitive site, the 21-bp repeat region plays a dominant role. One mutation lacking only 28 bp (including three of the hexanucleotide GGGCGG elements from the 21-bp repeat region) causes a substantial reduction in nuclease sensitivity.

Results reported by Jongstra et al. (35) and by Fromm and Berg (36) reinforce these conclusions. In addition, these investigators have shown that the enhancer region alone is sufficient to form a hypersensitive site. [Our results with d112 seemed to suggest that the 72-bp region alone was not sufficient; however, Fromm and Berg have shown (36) that enhancer sequences inserted in this position and orientation have atypical behavior.] At its simplest, the genetic determinants responsible for the nuclease-sensitive site in SV40 include two components. One maps to the enhancer region and the second, to the 21-bp repeat region. This latter element is probably responsible for the nuclease-sensitive structure over the replication origin. The two elements must work synergistically since the combination is more nuclease-sensitive than either alone. Further subdivision of the genetic determinants has not been ruled out. Clearly the genetics of this chromatin structure is more complex than was anticipated.

DISCUSSION

A portion of the 80S nucleoprotein isolated from nuclei of SV40-infected cells contains a hypersensitive site located between nucleotide positions 5200 and 400 on the viral genome. This is demonstrated by direct accessibility to endonucleases and by accessibility to exonuclease after the chromatin molecules have been cleaved within the nuclease-sensitive region by a restriction enzyme. The nuclease-sensitive structure is not seen in core particles released from virions or intracellular provirions suggesting that this feature is only a property of minichromosomes. Of the intracellular 80S nucleoprotein, only a portion exhibits this feature -- suggesting a structural heterogeneity which may underlie the functional heterogeneity of this nucleoprotein population.

Genetic determinants responsible for the hypersensitive site map between nucleotide positions 37 and 287. Within this region there are at least two elements, either of which can generate a hypersensitive site. One maps to the enhancer region and the other to the 21-bp repeat sequences. One possible interpretation of this organization is that there are really two

hypersensitive sites in SV40 and that they lie very close to each other and are not distinguished as two. If this is true, the two sites interact, since removal of sequences in the enhancer region reduces nuclease-sensitivity across the entire region and removal of sequences over the 21-bp repeat region likewise affects cleavage over the rest of the nuclease-sensitive region.

The role of the nuclease-sensitive chromatin structure.

We have accumulated a certain amount of structural information about this interesting feature of eukaryotic chromatin; however, we do not yet have an answer to the most important question which is: what is the biological significance? Circumstantial evidence that the nuclease-sensitive chromatin structure is biologically relevant is overwhelming. This structure precisely covers a portion of the genome containing all of the signals for DNA replication and gene transcription and it occurs in the subfraction of viral nucleoproteins which is engaged in transcription and replication (It has recently been shown [37] that essentially all SV40 chromatin molecules active in transcription contain a hypersensitive site. Since actively transcribing molecules account for only a small proportion of the total chromatin, some molecules which are not involved in transcription must also contain hypersensitive sites.)

Numerous studies on the chromatin structure around cellular genes have indicated a correlation between the presence of a hypersensitive site over the promoter and the potential for transcriptional activity (4-6); however, quantitative correlation between transcription and the presence of a hypersensitive site is not good. It may be that many genes have the potential for transcription but are regulated by other mechanisms which do not involve change in the chromatin structure over the promoter detectable by these methods.

The occurrence of a special chromatin structure over the sites for initiation of transcription and replication makes sense. Such a structure may facilitate entry of proteins which must bind to the initiation signals in order to get these processes started. Nucleosomal organization of the DNA may present obstacles to the binding of these factors and render initiation signals inactive. Our best evidence for this model comes from our observation that the origin of replication functions with much reduced efficiency when it is moved out of the nuclease-sensitive region.

During virion assembly the viral chromatin structure becomes associated with capsid proteins and eventually becomes a core structure within the virus particle. DNA within the core is more accessible to endonuclease cleavage (38) and to reaction with psoralen (39) than it is as viral chromatin; however, the differential accessibility of the region adjacent to the origin of replication is lost. This feature may be lost very early in assembly accounting for the observation that a large fraction of the 80S nucleoprotein lacks a hypersensitive site. Ambrose et al. (40) have recently shown that minichromosomes formed by a virus mutant which is blocked in encapsidation have a much higher proportion of molecules exhibiting the hypersensitive site. Presumably, this mutant does not even form the earliest intermediates in virion assembly.

Genetic determinants responsible for the nuclease-sensitive chromatin structure.

If the hypersensitive site is involved in initiation of transcription and/or replication we might expect to find that mutants deficient in this structure would be identified as promoter mutants or replication mutants. Mutations which contain defective replication origins, in general, show no alteration in the nuclease-sensitive chromatin structure; however, sequences mapping in the 21-bp repeat region have been shown to play an auxillary role in replication initiation (41) and a major role in formation of the hypersensitive site. It is possible that proteins needed for replication initiation bind inefficiently when the nuclease-sensitive chromatin structure is not present but that binding is facilitated by the presence of the hypersensitive site.

Genetic elements responsible for the hypersensitive chromatin structure correlate approximately with two important transcription signals (the enhancer and the 21-bp repeat sequences); however, the analysis is not yet detailed enough to determine whether these signals function, in part, through their ability to generate a specific chromatin structure. That conclusion will have to wait for more extensive genetic analysis -- perhaps with the use of point mutations -- to determine whether the presence of a hypersensitive site correlates with enhancer function and/or promoter function.

How is the nuclease-sensitive structure established?

It is clear that the formation of a hypersensitive site is controlled by

DNA sequence. This may be mediated through binding of specific proteins which creates a three-dimensional configuration that excludes nucleosomes, or the DNA sequence itself may be poorly compatible with nucleosome formation (e.g., by the formation of a non-B DNA structure.) Weintraub (42) has shown that, under the appropriate conditions, nuclease S1 will cleave naked DNA at the same positions that are nuclease-sensitive in chromatin. This specificity is dependent on the presence of supercoil tension. He has suggested that loss of a nucleosome from a region of chromatin might release supercoil tension which converts certain regions of the DNA into configurations not compatible with nucleosome organization.

While DNA structure may play an important role in formation of the hypersensitive site, it seems clear that the protein component plays a role that is more specific than simply providing differential protection of DNA sequences. For example, only a portion of the 80S nucleoprotein contains a hypersensitive site and that portion varies from experiment to experiment. If DNA sequence were the sole determinant, all chromatin molecules should have the same probability of forming the nuclease-sensitive structure. In addition, our experiments with exonuclease indicate that well-defined borders to the hypersensitive site persist even after the DNA has been cut with a restriction enzyme. If tension in the DNA molecule plays a role in this structure, it must be stabilized by protein-protein interactions or the structure would not be stable after the tension is released.

Whether primary determinants for the hypersensitive site operate at the level of DNA structure or of protein binding sites on the DNA, these determinants must lie between nucleotides 37 and 287. Other structural features must explain the fact that the nuclease-sensitive region extends from approximately nucleotides 5200 to 400. In some way a three-dimensional structure is generated which prevents nucleosomes from entering any nearer than 80 to 120 bp from the primary determinants. These considerations suggest a multicomponent nucleoprotein complex, yet the region appears (in the electron microscope) as if it were a segment of DNA free of bound proteins (43,44). Of course, loosely bound proteins may come off during preparation of the sample for electron microscope observation. Loosely bound proteins may also fail to protect the DNA from nuclease digestion by comparison with more tightly bound proteins located elsewhere on the viral genome.

ACKNOWLEDGMENTS

I would like to thank Jahanshah Amin, Lilia Babe, Judy Bigness, Michael Brown, Byron Cryer, Richard Eaton, Robert Gerard, John Hartmann, Jeffrey Innis, Suzanne Matsuura, Beth Montelone, Kenneth Ness, Judy Pauley, Charles Walter, Dianne Wigmore, and Mary Woodworth-Gutai who provided the experimental basis for this review.

This work was supported by grant number AI-12852 from the National Institutes of Health and grants PCM78-06847 and DCM84-08619 from the National Science Foundation. I was recipient of a Research Career Development Award from the NIH (AM-00549).

REFERENCES

1. Scott, W.A. and Wigmore, D.J. Cell 15:1511-1518, 1978.
2. Waldeck, W., Fohring, B., Chowdhury, K., Gruss, P. and Sauer, G. Proc. Natl. Acad. Sci. USA 75:5964-5968, 1978.
3. Varshavsky, A.J., Sundin, O. and Bohn, M. Cell 16:453-466, 1979.
4. Elgin, S.C.R. Cell 27:413-415, 1981.
5. Cartwright, I.L., Abmayr, S.M., Fleischmann, G., Lowenhaupt, K., Elgin, S.C.R., Keene, M.A. and Howard, G.C. CRC Crit. Rev. Biochem. 13:1-86, 1982.
6. Weisbrod, S. Nature (London) 297:289-295, 1982.
7. Shelton, E.R., Wassarman, P.M. and DePamphilis, M.L. J. Biol. Chem. 255:771-782, 1980.
8. Griffith, J.D. Science 187: 1201-1203, 1975.
9. Moyne, G., Freeman, R., Saragosti, S. and Yaniv, M. J. Mol. Biol. 149:735-744, 1981.
10. Garber, E.A., Seidman, M.M. and Levine, A.J. Virology 90:305-316, 1978.
11. Fernandez-Munoz, R., Coca-Prados, M. and Hsu, M.-T. J. Virol. 29:612-623, 1979.
12. Baumgartner, K., Kuhn, C. and Fanning, E. Virology 96:54-63, 1979.
13. Jakobovitz, E.B. and Aloni, Y. Virology 102:107-118, 1980.
14. Hartmann, J.P. and Scott, W.A. J. Virol. 46:1034-1038, 1983.
15. Weintraub, H. and Groudine, M. Science 193:848-856, 1976.
16. Buchman, A.R., Burnett, L. and Berg, P. In: DNA Tumor Viruses: Molecular Biology of Tumor Viruses (2nd ed.) Part 2. (Ed. J. Tooze), Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1980, pp. 799-829.
17. Saragosti, S., Cereghini, S. and Yaniv, M. J. Mol. Biol. 160:133-146, 1982.
18. Scott, W.A., Walter, C.F. and Cryer, B.L. Mol. Cell. Biol. 4:604-610, 1984.
19. Sundin, O. and Varshavsky, A. J. Mol. Biol. 132:535-546, 1979.
20. Lake, R.S., Barban, S. and Salzman, M.P. Biochem. Biophys. Res. Commun. 54:640-647, 1973.
21. Brady, J.N., Winston, V.D. and Consigli, R.A. J. Virol. 23:717-724, 1977.
22. Christiansen, G., Landers, T., Griffith, J. and Berg, P. J. Virol. 21:1079-1084, 1977.
23. Moyne, G., Harper, F., Saragosti, S. and Yaniv, M. Cell 30:123-130, 1982.
24. Hartmann, J.P. and Scott, W.A. J. Virol. 37:908-915, 1981.

25. Riley, D. and Weintraub, H. Cell 13:281-293, 1978.
26. Riley, D.E. Biochemistry 19:2977-2992, 1980.
27. Kunzler, P. and Stein, A. Biochemistry 22:1783-1789, 1983.
28. Scott, W.A., Wigmore, D.J. and Eaton, R.W. Miami Winter Symposia 16:83-98, 1979.
29. Wigmore, D.J., Eaton, R.W. and Scott, W.A. Virology 104:462-473, 1980.
30. Shenk, T. Cell 13:791-798, 1978.
31. Innis, J.W. and Scott, W.A. Mol. Cell. Biol. 4:1499-1507, 1984.
32. Gerard, R.D., Montelone, B.A., Walter, C.F., Innis, J.W. and Scott, W.A. Mol. Cell. Biol. 5:52-58, 1985.
33. Gerard, R.D., Woodworth-Gutai, M. and Scott, W.A. Mol. Cell. Biol. 2:782-788, 1982.
34. Innis, J.W. and Scott, W.A. Mol. Cell. Biol. 3:2203-2210, 1983.
35. Jongstra, J., Reudelhuber, T.L., Oudet, P., Benoist, C., Chae, C.-B., Jeltsch, J.-M., Mathis, D.J. and Chambon, P. Nature (London) 307:708-714, 1984.
36. Fromm, M. and Berg, P. Mol. Cell. Biol. 3:991-999, 1983.
37. Choder, M., Bratosin, S. and Aloni, Y. EMBO J. 3:2929-2936, 1984.
38. Brady, J.N., Radonovich, M., Lavialle, C. and Salzman, N.P. J. Virol. 39:603-611, 1981.
39. Kondoleon, S.K., Robinson, G.W. and Hallick, L.M. Virology 129:261-273, 1983.
40. Ambrose, C., Blasquez, V. and Bina, M. Proc. Natl. Acad. Sci. USA 83:3287-3291, 1986.
41. Bergsma, D.J., Olive, D.M., Hartzell, S.W. and Subramanian, K.N. Proc. Natl. Acad. Sci. USA 79:381-385, 1982.
42. Weintraub, H. Cell 32:1191-1203, 1983.
43. Saragosti, S., Moyne, G. and Yaniv, M. Cell 20:65-73, 1980.
44. Jakobovits, E.B., Bratosin, S. and Aloni, Y. Nature (London) 285:263-265, 1980.

11

SV40 CHROMATIN STRUCTURE AND VIRUS ASSEMBLY

VERONICA BLASQUEZ, CHRISTINE AMBROSE, HENRY LOWMAN AND MINOU BINA
Department of Chemistry, Purdue University, West Lafayette,
IN 47907 USA

ABSTRACT

The process of virion assembly plays an active role in modulating the structure of SV40 chromatin. Experiments with temperature-sensitive mutants of the major capsid protein (VP1) have shown that minichromosomes which contain an exposed regulatory region accumulate when the initiation of SV40 assembly is blocked. VP1, directly or through specific interactions with other proteins, alters the average spacing between nucleosomes and may be involved in generating a protected regulatory region in SV40 chromatin. The DNA sequences near or between positions 640 and 875 interact with capsid proteins in vivo and appear to contain signals used for the initiation of shell assembly. A DNA topoisomerase is encapsidated during virion assembly. This enzyme is probably responsible for the difference in SV40 DNA linking number, observed previously, between unencapsidated chromatin and virion derived chromatin.

INTRODUCTION

Over the past decade, SV40 has served as a paradigm for examining histone-DNA interactions, the structure-function relationship of chromatin, and the molecular mechanisms which promote DNA condensation in eukaryotic cells. The viral DNA is compacted into nucleosomes by the four cellular core histones to form the SV40 minichromosome or chromatin (1). The minichromosomes play a central role in the SV40 life cycle: they are the templates used for SV40 DNA replication, expression of the viral early and late genes, and virion assembly.

From nuclease digestion studies, it is evident that the structure of SV40 chromatin is altered during the infection cycle (2). A population of minichromosomes isolated from cells late in infection contains a region which is hypersensitive to cleavage by endonucleases (2-10). This

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

so-called open region encompasses the DNA sequences involved in the control of replication and transcription (Fig. 1) and it appears nucleosome free under the electron microscope (11,12). In contrast, this structural discontinuity is not present in minichromosomes obtained from disrupted virions (10,13).

Although numerous lines of evidence relate an open chromatin structure to transcription (14), it is not known how this structure is formed, maintained, or lost. Two general models have been considered to explain how the nuclease-hypersensitive region is formed (14). In one model, transcription factors somehow destabilize and remove the nucleosomes present on promoters (nucleosome-dissociation model). This model infers that the minichromosomes which have a protected regulatory region contain more nucleosomes than the molecules which exhibit an open region. In the other model, the open chromatin structure is established during replication (replicative model). This model predicts that nearly all of the minichromosomes should contain an exposed regulatory region after replication.

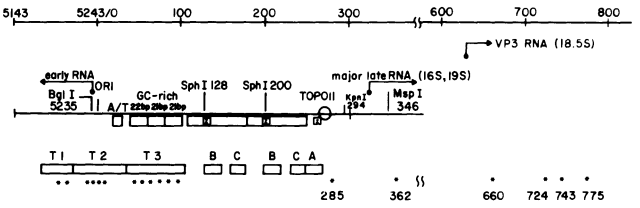


Fig. 1. Landmarks of the open region in SV40 chromatin and the initiation sites of the early and late mRNA transcripts. Shown from left to right are the origin of replication (ori), the TATA-like element (A/T), the three blocks of GC-rich repeats, and the 72-bp repeats (15). Z indicates regions which could form Z-DNA (29). The circle at 270 denotes the position at which a Topo II cleavage site has been mapped (30). The boxes below the diagram indicate the three T-antigen binding sites T1, T2, and T3 (31,32). Marked by a (*) are the positions which contain prototype T-antigen recognition sequences (33). The numbers above the diagram refer to the nucleotide positions in the SV40 sequence (15).

Evaluation of available data supporting one or the other model has not been a straightforward task because the assumption that virion assembly has no effect on the structure of SV40 chromatin may be incorrect. Numerous studies have shown that a massive amount of the virion proteins (VP1, VP2, and VP3) are synthesized late in infection (15) and that these proteins interact with minichromosomes during SV40 assembly (16-23). SV40 chromosomal alterations attributed to the capsid proteins include changes in nucleosome spacing (24,25), DNA supercoiling (26), and histone-DNA interactions (27,28).

The molecular events associated with SV40 assembly has been another fascinating problem and the subject of numerous studies. Since empty capsids were identified in cesium chloride density gradients, earlier reports concluded that during virion assembly, the SV40 DNA is introduced into a pre-formed shell (15). More recent studies have shown that in the course of SV40 maturation, minichromosomes (75 S) are initially condensed into 180 S previrions, then further condensed into young virions, which mature into stable 220 S virions (16-18). These kinetic studies along with the discovery that the pre- and young virions are not stable in cesium chloride gradients (17,18) led to an encapsidation model in which the capsid proteins add to and organize around the 75 S chromatin (16-18). This model predicts that the assembly of previrions must be cooperative--that is, the propagation step of shell assembly is much faster than the initiation event--since the putative intermediates which bridge the chromatin to previrions do not accumulate in cells infected with a wild-type SV40.

Here, we present a summary of our studies directed to (i) gain insight into the mechanism by which an open chromatin structure is established; (ii) examine whether the major capsid protein VP1 plays a role in changing the structure of SV40 chromatin; (iii) learn about the pathway of SV40 assembly; and (iv) examine the factors involved in changing the SV40 DNA topology when the chromatin is condensed into higher order structures during virion assembly.

MATERIAL AND METHODS

Cells and Viruses.

BSC-40 cells were derived from the BSC-1 line of African Green monkey

kidney cells by the selection for growth at 40°C. The SV40 strain 776 was used as the wild type virus. The temperature-sensitive assembly mutants were obtained from Dr. Robert Martin of the National Institutes of Health.

DNA labeling and isolation of SV40 nucleoprotein complexes.

BSC-40 cells infected with wt776, or the desired temperature-sensitive strain, were incubated at the appropriate temperature and labeled with [^{3}H]-thymidine for approximately 12 h prior to harvesting the cells. SV40 nucleoprotein complexes were isolated from the infected cells as previously described (20). Briefly, the harvested cells were concentrated by low-speed centrifugation, equilibrated with buffer B(10 mM HEPES, pH 7.8, 5 mM KCl, 0.5 mM $MgCl_2$) and Dounce homogenized after PMSF addition. The suspension was gently shaken at 4°C for 1.5 h. The cellular chromatin was removed by low-speed centrifugation and the supernatant was sedimented in sucrose gradients (5-31.5% made up in 0.05 M Tris-HCl, pH 7.5). For some experiments, buffer B containing 0.1 M NaCl was employed during the Dounce homogenization step in order to dissociate the capsid proteins from the tsB265 semiassembled particles. The products were fractionated on sucrose gradients as above.

Digestion of SV40 nucleoprotein complexes and analysis of digestion products.

For the DNase I digestion studies, sucrose gradient fractions containing the SV40 nucleoprotein complexes of interest were pooled and adjusted to 0.1 M NaCl and 5 mM $MgCl_2$. Typical DNA concentrations were 1-2 ug/ml. DNase I (Sigma) was then added to a final concentration of 5 units/ml and the mixture was incubated at 37°C for various period of time. All reactions were terminated by adjustment to 20 mM EDTA. In the case of the tsB265 semiassembled particles, the digested samples were refractionated in 5-31.5% sucrose gradient and only the DNA still bound to the capsid shell, evident by its sedimentation at 160 S, was further analyzed. DNA was purified from the samples, concentrated, and examined by the indirect end-labeling procedure (34). Briefly, the concentrated DNA samples were cleaved with AccI, BamHI, or TaqI and fractionated by electrophoresis through a 1.7% agarose gel. The DNA was transferred from the agarose gel to nitrocellulose paper and hybridized to a ^{32}P-labeled SV40 DNA fragment that mapped close to the chosen restriction cut.

Hybridization procedures were carried out as described by Maniatis et al. (35). The DNA bands were visualized by autoradiography.

The protocols used for probing the structure of nucleoprotein complexes with restriction enzymes and micrococcal nuclease have been previously published (10,25).

RESULTS AND DISCUSSION

Minichromosomes containing an open regulatory region are established during replication.

We have detected virions as early as 18 h after infection with wt776 and their appearance coincides with the appearance of the 75S chromatin (Fig. 2). There is a steady increase in the virion to chromatin ratio as infection proceeds (Fig. 2). Previous studies have shown that an appreciable fraction of the 75S chromatin isolated late in infection corresponds to minichromosomes which have been committed to the virion assembly pathway (10,19,26,36).

From the above data, it is evident that the wt chromatin population isolated at any time post-infection contains virion assembly intermediates, which could have different chromatin structures. Therefore, in order to examine the mechanism by which the minichromosomes containing an open regulatory region are formed, it is necessary to uncouple virion assembly from DNA replication. Towards this end, we

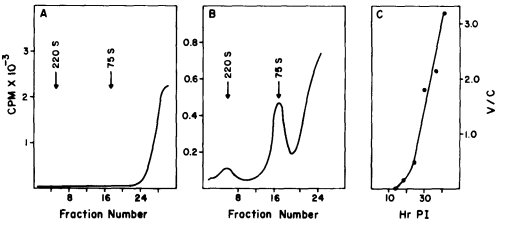


Fig. 2. Time course of chromatin and virion assembly during infection of BSC-40 cells by wt776. Total cellular extracts were prepared at 14 h (A) and 18 h (B) post infection and analyzed on sucrose gradients for chromatin (75 S) and virion (220 S). C represents the relative amounts of virion and chromatin (V/C) assembled at various intervals after infection (Hr PI).

characterized the temperature-sensitive mutants defective in virion assembly (tsC, tsBC, tsB) derived previously by Chou and Martin from mutagenesis of the wt776 virions with hydroxylamine (37). The tsB/C mutations map to the gene coding for the major capsid protein VP1 (38). We have shown that the mutants of the C group assemble normal virions at the permissive temperature (33°C) but are blocked in the initiation step of shell assembly at the nonpermissive temperature (40°C). Unlike the wt chromatin, the tsC219 minichromosomes assembled at 40°C do not contain a detectable amount of the capsid proteins (20).

Using restriction endonuclease digestion analysis, we have examined the effect that a block in the initiation of virus assembly has on the relative amount of minichromosomes which contain open or protected regulatory regions. The tsC219 minichromosomes prepared from cells labeled with [^{3}H]-thymidine at 40°C and at 33° were digested in separate reactions with various restriction enzymes. The DNA was purified and then examined on agarose gels. The fluorograms revealed that more than 95% of the tsC219 chromatin assembled at 40°C was accessible to cleavage by BglI. This accessibility was reduced to 50% when tsC219 chromatin

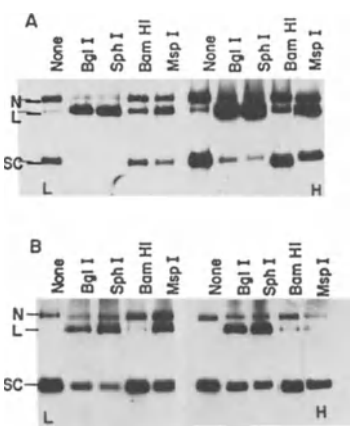


Fig. 3. Restriction enzyme digestion analysis of tsC219 chromatin assembled at the nonpermissive and permissive temperatures. The minichromosomes were extracted from cells labeled at 40°C (A) and 33°C (B), with [^{3}H]thymidine, and fractionated on sucrose gradients. The gradient fractions spanning the light (L) and heavy (H) sides of the 75 S chromatin peak were pooled and digested with the enzyme indicated above each lane. The digestion products were analyzed by agarose gel electrophoresis and detected by fluorography. N, L, and SC denote the mobilities of nicked, linear, and supercoiled SV40 DNA, respectively. (Reprinted with permission from ref. 10).

was prepared fom cells incubated at 33° C (Fig. 3).

A similar dependence on the initiation of virion assembly was also evident in the cleavage pattern of tsC219 chromatin treated with the enhancer-specific enzyme SphI, or with MspI, which recognizes a sequence near the transcription-initiation site of the late genes (Fig. 1), but not for enzymes which cleave the coding sequences (10, Fig. 3). Moreover, the cellular endogenous endonucleases introduce double-stranded cuts in the regulatory region of nearly all of the tsC219 minichromosomes assembled at 40°C (data not shown). As presented in the following section, tsC219 chromatin, blocked in the initiation of virus assembly, is also highly accessible to cleavage by DNase I. These results demonstrate that the minichromosomes containing an open region accumulate when the initiation of shell assembly is blocked and that a protected regulatory region appears with the onset of virion assembly.

If an open region in minichromosomes results from the dissociation of nucleosomes from the promoters, then the tsC219 chromatin assembled at 40°C should, on the average, contain fewer nucleosomes than wt776 chromatin. To resolve this issue, the negatively supercoiled DNA molecules isolated from wt776 and tsC219 chromatin were subjected to electrophoresis on a high resolution agarose gel. The number of supercoils in protein-free SV40 DNA reflects the number of nucleosomes which were present on minichromosomes (39). As Fig. 4 demonstrates, the distributions of the DNA topoisomers extracted from the two chromatin preparations are very similar. In the gel system employed, topological isomers differing by a single linking number could be resolved. We

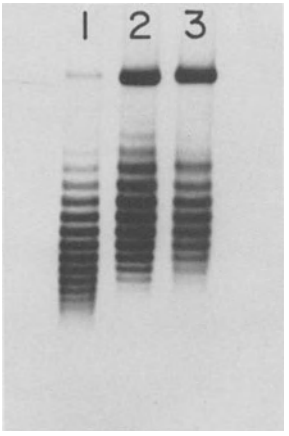


Fig. 4. Analysis of the DNA topoisomers in virion and in unencapsidated SV40 chromatin. DNA labeled with [^{3}H]-thymidine was purified from: wt776 virion (lane 1), wt776 chromatin (lane 2), and the tsC219 chromatin blocked in initiation of virion assembly (lane 3). The samples were run in an agarose gel containing 75 μg chloroquine /ml and visualized by fluorography.

therefore conclude that wt776 chromatin and the tsC219 chromatin, blocked in the initiation of virus assembly, contain the same number of nucleosomes. This finding does not support a model where transcription factors induce the dissociation of nucleosomes from the promoter region to establish an open chromatin structure.

Our data are most consistent with a model in which the open region is established and maintained during replication. Two previous studies have suggested that the open region is established after replication. One study demonstrated that the ori region is less accessible to nuclease attack in replicative intermediates than in mature chromatin (40) and the other found similar accessibility to BglI (41). However, the presence of replication-associated proteins on the DNA (see for example Fig. 8 in ref. 42) makes it unclear whether it is these proteins or the nucleosomes that confer the nuclease resistance. The possible role of replication in establishing minichromosomes containing an open region is further supported by transfection experiments (8). When plasmid DNA molecules harboring the origin-promoter-enhancer region of SV40 are introduced into CV-1 or COS-1 cells, they are rapidly folded by the cellular histones into minichromosomes. However, the pattern of DNase I hypersensitive cleavage is different for non-replicating and post-replicated chromatin. Only the molecules transfected into Cos-1 cells that replicate DNA exhibit the DNase I sensitivity pattern observed in the lytic cycle (8). A similar structure is observed for the DNase I-hypersensitive region of wt776 chromatin and tsC219 chromatin.

The DNase I hypersensitive sites in SV40 chromatin are clustered in two regions which flank the GC-rich repeats (7,8,9). One region (RI) includes the replication origin, the T antigen binding sites T1 and T2, the TATA box, and the transcription initiation sites of the early genes. RII spans the 72 bp repeats and extends toward the major transcription initiation site of the late genes (Fig. 1).

Fig. 5 compares the DNase I digestion profiles of wt776 chromatin and the tsC219 chromatin which was blocked in the initiation of shell assembly. For these experiments, the viral chromatin was treated with DNase I as a function of time. The DNA was purified, digested with AccI and analyzed by the indirect end-labeling technique (34). DNase I linearized nearly all of the tsC219 minichromosomes while only a fraction

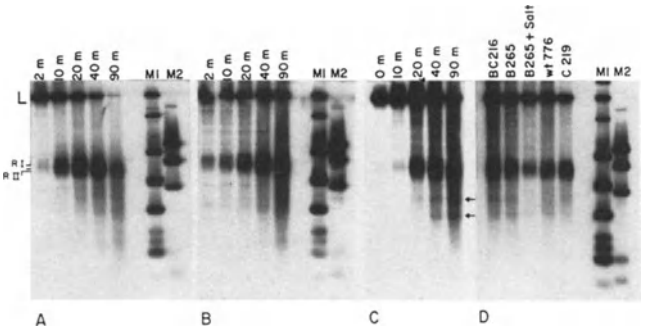


Fig. 5. DNase I digestion analysis of the viral nucleoprotein complexes. The tsC219 chromatin blocked in the initiation of shell assembly (panel A) and wt776 chromatin (panel B) were digested with DNase I as a function of time as indicated above each lane. DNA was purified, digested with AccI, and fractionated on an agarose gel. The bands were visualized by the indirect-end labeling procedure (34), by using a ^{32}P-labeled HaeII-AccI fragment (796 bp) of SV40 as a probe. For the experiments presented in panel C, the tsB265 semiassembled particles produced at the nonpermissive temperature were digested with DNase I as a function of time and the products were refractionated on sucrose gradients to obtain the DNA fragments containing the capsid protein and sedimenting at about 160 S. The DNase I cleavage sites in these fragments were analyzed with respect to the AccI site as above. Panel D shows a comparative analysis of the DNase I digestion products of the tsBC216 capsid-chromatin complexes, the tsB265 semiassembled particles, the tsB265 chromatin from which the capsid was dissociated prior to DNase I digestion, wt776 chromatin, and the tsC219 chromatin blocked in the initiation of shell assembly. The samples were digested by DNase I to the same extent and examined with respect to the AccI site as above. L denotes the position of linear SV40 DNA. RI and RII mark the DNase I-hypersensitive regions. The arrows between panels C and D point to the two DNase I sensitive sites present in the tsB265 semiassembled particles. Lanes M1 and M2 contained ^{32}P-labeled DNA size markers.

of the wt776 chromatin was cleaved by the enzyme. For both preparations, the highest frequency of DNase I cleavage occurred in the regulatory region at nearly similar sites. The RI regions in tsC219 and wt776 minichromosomes both mapped to positions 5200 to 40. The RII regions could be subdivided into subregions RIIa and RIIb. RIIa appears early during digestion; it begins around position 100 and spans 1.5 copies of the 72 bp repeats. The RIIb subregion is cleaved later in digestion; it includes a DNA segment from the late-proximal copy of the 72 bp repeat and extends to about nucleotide 280.

A reduced hypersensitivity to DNase I is apparent in the region beyond the enhancer elements. Near the major transcription initiation site of the late genes, there are minor cleavage sites which map to positions 309 and 370, approximately. In addition, we have detected a DNase I-sensitive site at about position 665, 35 base pairs away from the transcription-initiation site of the 18.5 S mRNA which codes for the minor capsid protein VP3 (15, Fig. 1). We were surprised to find, at position 660, 8 base pairs with a primary sequence identical to the palindromic sequence found around the replication origin, in the T-antigen binding site T2 (Fig. 1). Prototype T-antigen recognition sequences (33) are also found on each side of the major transcription-initiation site of the late genes, at positions 285 and 362 (Fig. 1), and elsewhere in the SV40 genome.

Changes in the average nucleosome repeat length as a function of virion assembly.

In eukaryotic cells, the nucleosome spacing periodicity (repeat length) occurs in a species-, cell type-, and developmental stage-specific manner (43-45). *In vitro* reconstitution studies have demonstrated that two independent mechanisms may be involved in establishing a particular average nucleosome repeat length: (1) DNA base sequence-specific core histone interaction, termed nucleosome positioning (e.g., 46), and (2) nucleosome alignment mediated by the very lysine-rich histones H1 or H5 (47,48).

Micrococcal nuclease digestion studies have shown that the average nucleosome repeat of wt SV40 chromatin is 198 bp, a value longer than the average repeat length obtained for the bulk cellular chromatin (188 bp), (25,49). Since nearly all of the tsC219 minichromosomes assembled at

40°C contain a regulatory region hypersensitive to endonucleolytic cleavage, but the same number of nucleosomes as wt molecules, we probed their overall nucleosome arrangement to investigate whether it was different from the arrangement observed for wt chromatin. Micrococcal nuclease digestion of tsC219 and wt776 minichromosomes in parallel reactions produced a series of DNA bands characteristic of the repeating subunit structure of chromatin (Fig. 6). The average nucleosome spacings were calculated from the slope of a line obtained by plotting DNA length

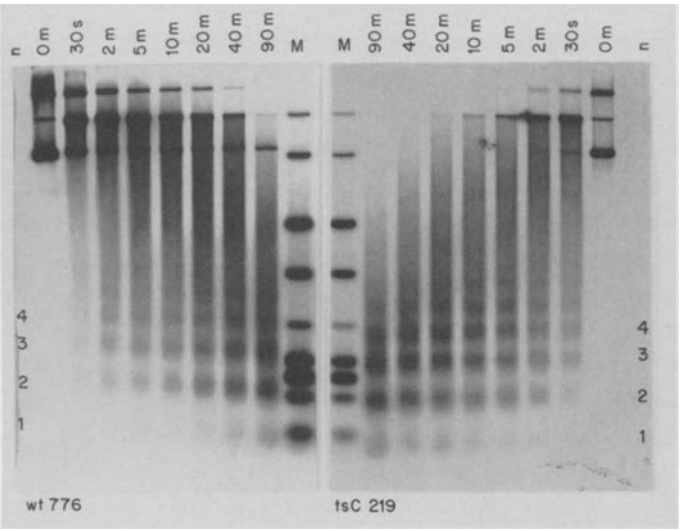


Fig. 6. Comparison of the average nucleosome repeat length in wt776 chromatin and in tsC219 chromatin blocked in the initiation of virus assembly. The minichromosomes were isolated from cells labeled with [^{3}H]-thymidine and digested in parallel reactions with micrococcal nuclease for various periods of time as indicated above. The digestion products were freed from proteins, fractionated in an agarose gel, and visualized by fluorography. The lane marked M consists of ^{32}P-end-labeled size markers (5243, 3273, 1768, 1169, 752, 526, 447, 366, and 215 bp). The lane labeled n denotes the number of nucleosomes corresponding to each oligomeric DNA band. (Reprinted with permission from ref. 25).

against nucleosome oligomer number. This method avoids extrapolation procedures to correct for the micrococcal nuclease nibbling activity (45). The results indicate that the average nucleosome repeat of tsC219 chromatin assembled at 40°C is considerable shorter than the repeat of wt776 chromatin (177 + 4 bp as opposed to 198 + 4 bp). A longer repeat is also observed in tsC219 chromatin assembled at the permissive temperature and in the intermediates which accumulate in cells infected with VP1 mutants blocked in the propagation step of virus assembly at 40°C (25).

The results presented above demonstrate that: (i) the SV40 chromatin undergoes nucleosomal rearrangement, increasing the average repeat length upon the onset of virion assembly initiation, and (ii) this chromosome reorganization is somehow made possible by a function localized in the tsC domain of VP1. A hypothesis consistent with the data is that after replication, VP1, directly or through specific interactions with other proteins, binds to the SV40 DNA and slides the nucleosomes apart. Such a sliding mechanism may be involved in producing the minichromosomes which contain a protected regulatory region. Nucleosomes present over the regulatory region may act to repress gene activity and to commit minichromosomes to the virion assembly pathway (Fig. 7). This model is consistent with the nucleosome sliding mechanism proposed previously by Coca-Prados et al. (17) to explain the finding that the average nucleosome repeat length in previrions is longer than the average repeat observed for minichromosomes.

Positioning of nucleosomes on SV40 minichromosomes.

The position of nucleosomes on SV40 DNA has been previously predicted from the periodicity of certain dinucleotides along the sequence (50). The fact that the average nucleosome repeat in SV40 chromatin changes as a function of virion assembly makes it unlikely that *in vivo* the nucleosomes occupy unique positions on the DNA. Nonetheless, the results of numerous studies indicate that nucleosomes are distributed non-randomly with respect to the hypersensitive region (3-12). Moreover, a non-random distribution of nucleosomes in the coding regions has been deduced from mapping the micrococcal nuclease cleavage sites on SV40 chromatin (7,51,52). In agreement with these results, we find, on the

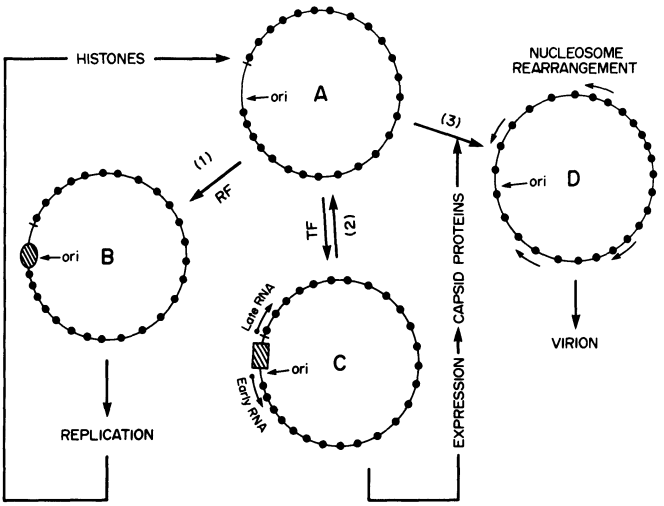


Fig. 7. Model for channeling of SV40 chromatin into transcription, replication, and encapsidation pathways. We proposed that the minichromosomes containing a nuclease-hypersensitive region (species A) are established during replication through interactions of DNA with nonhistone regulatory proteins and possibly by the positioning of nucleosomes in the DNA segments which flank the open region. We envisage that the replication and transcription factors compete with the capsid proteins for the newly assembled minichromosomes. For example, species B is a molecule committed to replication as a result of the interactions of its replication origin with the replication factors (RF) and species C is a transcription complex. VP1, directly or indirectly, rearranges the SV40 chromatin structure by sliding the nucleosomes toward the open region (24,25) to form the molecules committed to the encapsidation pathway (species D). The proposed mechanism would result in channeling SV40 molecules into the transcription, replication, and encapsidation pathways in a coordinated fashion (Reprinted with permission from ref. 25).

autoradiograms obtained from the DNase I digestion studies, well-defined subbands which map outside of the hypersensitive region (Fig. 5). As expected, different electrophoretic mobilities are observed for the subbands generated from the digestion of wt776 chromatin and tsC219 chromatin assembled at 40°C. The data suggest that on SV40 DNA, there are alternative nucleosome positioning sites. Because of their good resolution, we have determined the map location of the subbands observed in DNase I digestion studies of the wt776 chromatin (Fig. 5). The results suggest that in the coding regions, nucleosomes can be found in the region which includes nucleotides 594, 745, 1065, 3450, and 4280. Nucleosomes arranged in non-random arrays could also be mapped to regions between nucleotides 1160-1840, 2300-2900 and 3500-4093.

The initiation and propagation of shell assembly on SV40 chromatin.

To follow the pathway of SV40 assembly, we have examined the intermediates assembled in cells infected with the tsBC and tsB mutants at 40°C (21,22). One of the tsBC mutants (tsBC11) appears to be defective in the initiation of shell assembly and only produces the 75S chromatin (16,22). In cells infected with the other BC mutants, nucleoprotein complexes which sediment around 120 S accumulate. These complexes contain viral capsid proteins and appear under the electron microscope as a small protein cluster bound to the chromatin (22). The intermediates which accumulate in tsB-infected cells appear as semiassembled virions. They sediment as a broad peak around 160 S, and consist of capsid proteins, with varying degrees of polymerization, bound to the chromatin (21). The nature of complexes assembled in tsBC and tsB-infected cells demonstrates that, during virion assembly, the capsid proteins interact with SV40 chromatin and polymerize as incoming proteins add onto the nucleated complexes. The incomplete polymerization observed in the mutant-infected cells is probably due to alterations in protein-protein contacts (53).

We digested the semiassembled nucleoprotein complexes formed in tsB265-infected cells with DNase I to determine whether the partially polymerized capsid interacts with specific SV40 DNA sequences. The chromatin fragments containing the bound capsid proteins were isolated by fractionation of the DNase I digestion products on sucrose gradients. We found that a DNase I hypersensitive region was present in a fraction of

the fragments (Fig. 5C). In addition, we noted the presence of two DNase I-sensitive sites (at positions 640 and 875) which appeared after 10 min of digestion and became more pronounced after 20 min. (Fig. 5C). DNase I also introduced cuts at similar positions in tsBC216 capsid-chromatin complexes, but not in the tsC219 chromatin which was blocked in the initiation step of shell assembly and which consequently did not contain bound capsid protein (Fig. 5D). The capsid can be dissociated from the tsB265 chromatin by including 0.1 M NaCl in the buffer used for nucleoprotein complex isolation. When this procedure was followed, the isolated tsB265 chromatin was not cleaved by DNase I at positions 640 and 875 (Fig. 5D). The results indicate that these positions, or the sequences around them, are somehow involved in capsid protein-chromatin interactions.

Inspection of the SV40 nucleotide sequence revealed that the region between the two DNase I-sensitive sites, at positions 640 and 875, contains four blocks of base pairs whose sequences resemble the pentanucleotide motif recognized by the large T antigen in vitro (33). The first block, located at nucleotide 660, is the sequence which contains 8 base pairs identical to the palindromic sequence found around the replication origin. The other three blocks contain the sequences GAGGC, GCCTC, and GGGGC, at nucleotides 724, 743, and 775, respectively. Other prototype T antigen recognition sequences are found at nucleotides 1402, 1508, 1538, 2197, 4471, 4936.

Since SV40 DNA recombinants constructed from the SV40 ori region (containing the T antigen binding sites) and eukaryotic genes can produce plaques when complemented with a wt SV40, it has been suggested that SV40 assembly does not require a packaging signal (54). However, the efficiency of packaging was not determined. Our results suggest that another region of SV40 DNA which contains a cluster of prototype T antigen recognition sequences interacts with the capsid proteins. Thus it is plausible that such sequences might serve as signals for the initiation of shell assembly.

Chromatin topology and virion assembly

Chen and Hsu have recently discovered that the higher order compaction of SV40 chromatin during virion assembly coincides with an increase in the SV40-DNA linking number (26). Since this phenomenon was not detected

in previous studies (55,56), we compared the relative electrophoretic mobility of DNA topoisomers extracted from the 75 S chromatin and from the 220 S virion in an agarose gel containing 75 µg chloroquine/ml (Fig. 4). At this concentration of chloroquine, the negative superhelical turns present in SV40 DNA are removed and positive superhelical turns are generated (56). Thus, the molecules which exhibit the greatest mobility have the highest linking number.

The results indicate that during encapsidation, the linking number of the 75S chromatin is increased by about 2 turns (Fig. 4). It appears that the capsid proteins wind the SV40 DNA or alter histone-DNA interactions as they polymerize around the minichromosomes. A topoisomerase must be present during virion assembly to translate such changes in the SV40 chromatin structure into a change in linking number of SV40 DNA. This mechanism is supported by our previous discovery of a topoisomerase activity in mature virions (57).

The encapsidated topoisomerase can be detected when concentrated virion preparations are disrupted with a reducing agent and a chelator of divalent ions. The biochemical properties of the enzyme suggest that is a type I DNA topoisomerase. It does not require ATP to relax supercoiled DNA. It is active in the absence of divalent ions although its activity is enhanced in the presence of Mg^{2+} (57). The enzyme readily relaxes both positive and negative supercoiled DNA (data not shown). Kasamatsu and Wu previously isolated a protein covalently linked to SV40 DNA from virions disrupted in SDS (58). It is likely that this protein corresponds to the encapsidated topoisomerase (57).

CONCLUSIONS

During the lytic cycle, the SV40 chromatin undergoes several structural changes. The results of our studies on the SV40 assembly mutants suggest that the minichromosomes which contain an open regulatory region are established directly during replication. This can happen, for example, if the interactions of the transcription and replication factors with DNA exclude nucleosome assembly, and/or if the core histones exhibit selectivity for the DNA sequences which flank the ori region.

The major capsid protein VP1, directly or through specific interactions with other proteins, alters the spacing between nucleosomes,

perhaps by nucleosome sliding. Sliding of nucleosomes toward the regulatory region provides a mechanism for the repression of gene activity and may serve as a signal for channelling minichromosomes into the virion assembly pathway.

The signals used for the initiation of SV40 chromatin packaging are not known. It appears that the capsid proteins interact preferentially with the prototype T antigen recognition sequences. Such interactions may play a role in triggering the initiation of shell assembly.

The capsid proteins polymerize around the SV40 chromatin during the propagation step of SV40 assembly. A DNA topoisomerase plays an active role during this process. This enzyme changes the linking number of SV40 DNA as the capsid proteins fold the minichromosomes into higher order structures. It remains to be seen whether a change in linking number also occurs when the cellular chromatin is organized into higher order structures.

This work was supported by research grants form the National Science Foundation.

REFERENCES

1. Griffith, J. D. Science 87:1202-1203, 1976.
2. Cremisi, C. Nucleic Acids Res. 9:5949-5964, 1981.
3. Scott, W. A. and Wigmore, D. J. Cell 15:1511-1518, 1978.
4. Waldeck, W., Fohring, B., Chowdhury, K., Gruss, P. and Sauer,G. Proc. Natl. Acad. Sci. USA 75:5964-5968, 1978.
5. Varshavsky, A. J., Sundin, O. H. and Bohn, M. J. Nucleic Acids Res. 5:3469-3477, 1978.
6. Varshavsky, A. J., Sundin, O. H. and Bohn, M. J. Cell 16:453-466, 1979.
7. Cereghini, S., Herbomel, P., Jouanneau, J., Saragosti, S., Kanitna, M., Bourachot, B., DeCrombrugghe, B., and Yaniv, M. Cold Spring Harbor Symp. Quant. Biol. XLVII:935-944, 1982.
8. Cereghini, S., and Yaniv, M. The EMBO J. 3:1243-1253, 1984.
9. Jongstra, J., Reudelhuber, T. L., Oudet, P., Benoist, C., Chae, C. B., Jeltsch, J. M., Mathis, D. J. and Chambon, P. Nature 307:708-714, 1984.
10. Ambrose, C., Blasquez, V. and Bina, M. Proc. Natl. Acad. Sci. USA 83:3287-3291, 1986.
11. Jakobovits, E. B., Bratosin, S. and Aloni, Y. Nature 285:263-265, 1980.
12. Saragosti, S., Moyne, G. and Yaniv, M. Cell 20:65-73, 1980.
13. Hartman, J. P. and Scott, W. A. J. Virol. 37:90-8-915, 1981.

14. Eissenberg, J. C., Cartwright, I. L., Thomas, G. H. and Elgin, S. C. R. Ann. Rev. Genet. 19:485-536, 1985.
15. Tooze, J. Molecular Biology of Tumour Viruses, Part 2, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1981.
16. Garber, E., Seidman, M. and Levine, A. J. Virology 90:305-316, 1979.
17. Coca-Prados, N. and Hsu, M.-T. J. Virol. 31:199-208, 1979.
18. Fanning, E. and Baumgartner, I. Virology 102:1-12, 1980.
19. Milavetz, B. and Hopkins, N. J. Virol. 43:830-839, 1982.
20. Bina, M., Blasquez, V., Ng, S.-C. and Beecher, S. Cold Spring Harbor Symp. Quant. Biol. XLVII:565-569, 1982.
21. Blasquez, V., Beecher, S. and Bina, M. J. Biol. Chem. 258:8477-8484, 1983.
22. Ng, S.-C. and Bina, M. J. Virol. 50:471-477, 1984.
23. Blasquez, V. and Bina, M. FEBS Lett. 181:64-68, 1985.
24. Coca-Prados, M., Yu, H. and Hsu, M.-T. J. Virol. 44:603-609, 1982.
25. Blasquez, V., Stein, A., Ambrose, C. and Bina, M. J. Mol. Biol. 191 :97-106, 1986.
26. Chen, S. S. and Hsu, M.-T. J. Virol. 51:14-19, 1984.
27. Moyne, G., Harper, F., Saragosti, S. and Yaniv, M. Cell 30:123-130, 1982.
28. Brady, J. N., Lavialle, C. and Salzman, N. P. J. Virol. 35:371-381, 1980.
29. Nordheim, A. and Rich, A. Proc. Natl. Sci. USA 80:1821-1825, 1983.
30. Yang, L., Rowe, T. C., Nelson, E. M. and Liu, L. F. Cell 41:127-132, 1985.
31. DeLucia, A. L., Lewton, B. A., Tjian, R. and Tegtmeyer, P. J. Virol. 46:143-150, 1983.
32. Tjian, R. Cell 26:1-2 (1981).
33. Wright, P. J., DeLucia, A. L. and Tegtmeyer, P. Mol. Cell. Biol. 4:2631-2638, 1984.
34. Wu, C. Nature 286:854-860, 1980.
35. Maniatis, T., Fritsch, E. F. and Sambrook, J. Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1982.
36. Garber, E. A., Seidman, M. and Levine, A. J. Virol. 107:389-401, 1980.
37. Chou, J. Y. and Martin, R. G. J. Virol. 15:127-136, 1975.
38. Lai, C-J., and Nathans, D. Virology 75:335-345, 1976.
39. Germond, J. E., Hirt, B., Oudet, P., Gross-Bellard, M., and Chambon, P. Proc. Natl. Acad. Sci. USA 72:1843-1847, 1975.
40. Varshavsky, A., Levinger, L., Sundin, O., Barsoum, J., Ozkaynak, E., Swerdlow, P. and Finely, O. Cold Spring Harbor Sym. Quant. Biol. XLVII:511-528, 1982.
41. Tack, L. C., Wassarman, P. M. and DePamphilis, M. L. J. Biol. Chem. 255:771-782, 1980.
42. Stillman, B. Cell 45:555-565, 1986.
43. Shaw, B. R., Cognetti, G., Sholes, W. M. and Richards, R. G. Biochemistry 20:4971-4978, 1981.
44. Arceci, R. J. and Gross, P. R. Dev. Biol. 80:186-209, 1980.
45. Thomas, J. O. and Thompson, R. J. Cell 10:633-640, 1977.
46. Simpson, R. T., Thomas, F. and Brubaker, J. M. Cell 42:799-808, 1985.
47. Stein, A. and Kunzler, P. Nature 302:549-550, 1983.

48. Stein, A. and Bina, M. J. Mol. Biol. 178:341-363, 1984.
49. Shelton, E. R., Wassarman, P. M. and DePamphilis, M. L. J. Biol. Chem. 255:771-782, 1980.
50. Mengeritsky, G. and Trifonov, N. Cell Biophys. 6:1-8, 1984.
51. Sundin, O. and Varshavsky, A. J. Mol. Biol. 132:535-546, 1979.
52. Nedospasov, S. A. and Georgiev, G. P. Biochem. Biophys. Res. Commun. 92:532-539, 1980.
53. Bina, M., Blasquez, V. and Ambrose, C. Biophys. J. 49:38-40, 1986.
54. Hamer, D. DNA cloning in mammalian cells with SV40 vectors, p. 83-101.In J. K. Stelow and A. Hollander (ed.), Genetic engineering, Vol. 2. Plenum Publishing Corp. New York, 1980.
55. Keller, W., Muller, U., Eicken, I., Wendel, I. and Zentgraf, H. Cold Spring Harbor Symp. Quant. Biol. XLII:227-244, 1977.
56. Shure, M., Pulleyblank, D. E. and Vinograd, J. Nucleic Acids Res. 4:1183-1205, 1977.
57. Bina, M., Beecher, S. and Blasquez, V. Biochemistry 21:3057-3063, 1982.
58.Kasamatsu, H. and Wu, M. Proc. Natl. Acad. Sci. USA 73:1945-1949, 1976.

12

THE COMPLEX CELLULAR NETWORKS IN THE CONTROL OF SV40 GENE EXPRESSION

A. BEN-ZE'EV

Department of Genetics, Weizmann Institute of Science, Rehovot 76100, Israel

ABSTRACT

Electronmicroscopical and biochemical studies have shown an association between SV40 viral proteins, viral assembly, and the biogenesis of late SV40 RNA with the complex cellular networks of the nucleus and of the cytoplasm. Pulse-chase experiments revealed a rapid transport of viral capsid proteins from the cytoplasm to the nuclear matrix, where virus assembly occurs. The nuclear matrices of SV40-infected cells retain the majority of mature SV40 virions and the rapidly labeled SV40 nuclear RNA. Isolated nuclei from SV40-infected cells when incubated in the presence of [α-^{32}P]UTP elongate the in vivo preinitiated SV40 RNA synthesizing both long viral RNA molecules and a 94 nucleotide long promoter-proximal viral RNA species (attenuator RNA). In contrast to rapidly labeled viral nuclear RNA, the attenuator RNA is not associated with the nuclear matrix. Pretreating the cells with proflavine that interferes with RNA secondary structure increases the amount of long viral RNA molecules which become associated with the nuclear matrix. Pretreatment with DRB (1-β-ribofuranosylbenzimidazole) which enhances premature termination of RNA synthesis enhances the accumulation of the attenuator RNA in the nuclei in a DNase and salt soluble fraction which is not associated with the matrix.

Isolated nuclear matrices that contain about 2%-6% of the viral DNA are capable of synthesizing in vitro in the presence of [α-^{32}P]UTP between 35%-70% of the viral RNA synthesized in isolated nuclei implying a dramatic enrichment in the nuclear matrix of transcriptionally active ternary complexes. The processing and transport of the viral nuclear RNA occur also in association with the matrix. In the cytoplasm, the newly

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

synthesized RNA appears in association with viral-specific polyribosomes on the cytoskeletal framework where it is engaged in the synthesis of viral proteins. The association of the viral RNA with the cytoskeletal framework appears to be obligatory for translation, since virtually all the viral specific polyribosomes are associated with the cellular framework. The results suggest a major role for the complex cellular networks in the localization of virus assembly and in the control of SV40 gene expression.

INTRODUCTION

One of the central questions in cell biology is the molecular basis for pattern formation and tissue morphogenesis. Cellular organization and tissue architecture, and the principles that determine the assembly and function of cellular structure, are major subjects of contemporary biology.

Recent studies of cell structure and organization have shown the existence of complex subcellular proteinacious networks in the cytoplasm (cytoskeleton) and the nucleus (nuclear matrix) of eukaryotic cells (for recent papers on the cytoskeleton and the nuclear matrix see 1,2). The development of sequential fractionation techniques that made possible the isolation of these structures, enabled the investigation of their organization and biochemical composition. Using such approaches it was found that all animal viruses use the cytoplasmic and/or the nuclear skeletons during their replication at one or several steps of their macromolecular metabolism (for reviews see 3,4).

In the present chapter I summarize studies where we used the various steps of SV40 macromolecular metabolism to investigate topographic aspects of viral behavior inside the infected cell, in relation to the subcellular structures. By studying the regulation of viral gene products in relation to these networks we hope also to understand the possible involvement of these complex cellular structures in the regulation of cellular gene expression.

MATERIALS AND METHODS

Cells and virus

BSC-1 cells were grown in Dulbecco's modified Eagle's medium (DMEM) containing 10% calf serum. Confluent cultures were infected with plaque-purified SV40 (strain 777) at 10-50 p.f.u./cell in the same medium containing 2% calf serum.

Pulse-chase protocol

SV40-infected cells were treated with 20 mM glucosamine in DMEM containing 2% calf serum for 60 min. The cells were washed twice with fresh medium and labeled for 10 min with 200 μCi/ml of [5-^{3}H]uridine (35.1Ci/mmol, The Radiochemical Centre, Amersham, UK). Following the pulse, the cells were washed twice with DMEM and the chase was initiated by incubating the cultures with fresh medium containing 20 mM glucosamine, 5 mM uridine, 5 mM cytidine, and 2.5 mM thymidine.

Cell fractionation and RNA extraction

The fractionation into the soluble and cytoskeletal fractions was similar to the procedure described by Ben-Ze'ev et al (16). Briefly, the cells were washed twice with cold phosphate-buffered saline and once with extraction buffer [50 mM NaCl, 10 mM Hepes, pH 7.4, 2.5 mM $MgCl_2$, 300 mM sucrose, and 1 mM phenylmethyl sulphonyl fluoride (PMSF)]. Extraction was achieved by adding to the washed cells the above buffer with 1% Triton X-100 for 5 min on ice. The Triton X-100 soluble fraction was removed. The cytoskeletons were solubilized by scraping them into DOC/Tween buffer (10 mM NaCl, 10 mM Hepes pH 7.4, 1.5 mM $MgCl_2$, 0.5% deoxycholate, 1% Tween 40, and 1 mM PMSF) followed by homogenization either by pipetting or using a Teflon motor-driven homogenizer. The nuclei were pelleted at 2000 r.p.m. for 2 min at 4^{o}C and the solubilized cytoskeletons were removed.

The nuclei were resuspended in 80 μl of a 2 mg/ml DNase I solution containing 0.15 M NaCl, 5 mM $MgCl_2$, 1 mM PMSF. The DNase was previously chromatographed on agarose-5'-uridine-2'(3') phosphate and further treated with iodoacetate to remove traces of RNase. The nuclei were incubated on ice for 30-60 min. The nuclear suspension was brought to 0.5 M $(NH_4)_2SO_4$ in 10 mM Hepes pH 7.4 and kept on ice for 5 min. The

matrices were then collected by centrifugation at 1000 r.p.m. for 5 min at 4°C. The DNase and salt-soluble fraction was removed and the nuclear matrix was resuspended in SDS buffer (100 mM NaCl, 0.5% SDS 10 mM Tris pH 7.4, 5 mM EDTA). For RNA extraction, SDS was added to all fractions to 0.5% and RNA was extracted with phenol chloroform and chloroform-isoamyl alcohol. The RNA preparations were denatured with 90% formamide at 37°C for 3 min before sedimentation through 15-30% sucrose gradients in SDS buffer in the SW 27.1 rotor at 25 000 r.p.m. for 14 h at 20°C.

RNA-DNA hybridization

Each fraction of a sucrose gradient was brought to 4xSSC (SSC is 0.15 M NaCl, 0.015 M sodium citrate) and was incubated in a final volume of 250 µl at 68°C for 24 h with 1 µg SV40 single-stranded (ss) DNA immobilized onto 7-mm nitrocellulose filters.After 24 h the filters were washed three times with 2xSSC, treated with RNase A (10 µg/ml) for 1 h at 22°C, washed again with 2xSSC, dried, and counted.

Transcription in vitro in isolated nuclei and in nuclear matrices

Isolated nuclei and nuclear matrices prepared as described above were resuspended in 100 mM $(NH_4)_2SO_4$, 6 mM KCl, 30 mM Hepes-NaOH (pH 8.0), 12.5% (v/v)glycerol, 1.5 mM $MnCl_2$, 1.2 mM dithiothreitol as described (Hay et al. (25)). The non-radioactive nucleotides (ATP, CTP, GTP) were added to a final concentration of 0.4 mM each and finally, 100 µCi of [α-^{32}P]UTP (410 Ci/mmol; Amersham, England) were added. The reaction mixture was incubated at 26°C for 5 min. The nuclei or nuclear matrices were separated from the mixture by centrifugation at 500 g and RNA was extracted from nuclei or nuclear matrices in SDS buffer and phenol chloroform. Alternatively, the nuclear matrix was prepared from the isolated nuclei after transcription in vitro, as described above, and the RNA was extracted from the nuclear matrix and from the DNase and salt soluble fraction. For desalting the DNase and salt soluble fraction before RNA extraction, 0.5% (w/v) SDS was added and the salt was removed by precipitating twice at 10,000 g at 4°C for 10 min. DNA was removed from the phenol-extracted and ethanol-precipitated RNA by digestion with 100 µg DNase I/ml in TKM (50 mM-Tris-HCl (pH 6.7), 25 mM-KCl,2.5 mM $MgCl_2$) at 4°C for 60 min and then by a second extraction with phenol chloroform and SDS. To remove nucleotides, the RNA was chromatographed through a

Sephadex G-25 column and the RNA in the void volume was precipitated with ethanol.

Analysis of RNA by RNA-DNA hybridization and gel electrophoresis

The ^{32}P-labeled RNA was analyzed by hybridization to SV40 DNA restriction fragments generated by digesting SV40 DNA with EcoRI, BglI and HpaI restriction enzymes. To analyze the virus-specific RNA by gel electrophoresis, [^{32}P]RNA was hybridized to SV40 DNA filters in 70% (v/v) formamide, 0.3 M NaCl, 0.5% SDS, 10 mM Tris HCl (pH 7.4), 1 mM EDTA and the virus-specific RNA was eluted in 90% formamide, 1 mM EDTA, 10 mM Tris-HCl (pH 8.3), 100 mM-boric acid, 2 mM EDTA, 10 M urea, incubated for 5 min at 65°C and subjected to electrophoresis on 7 M urea/12% (w/v) acrylamide gels.

Protein gel electrophoresis(SDS-PAGE)

Samples for SDS-polyacrylamide gel electrophoresis were concentrated by ethanol precipitation (3 volumes) at -20°C, resuspended in sample buffer (2% SDS, 10% glycerol, 5% β-mercaptoethanol), and analyzed in 15% acrylamide gels.

Electron microscopy

Samples were fixed with 2% glutaraldehyde in 0.2 M cacodylate buffer, and post-fixed with 0.5% osmium tetroxide in 0.2 M cacodylate buffer. After graded dehydration in ethanol the pellets were embedded in Epon and sectioned. Photographs were taken in a Phillips 300 electron microscope.Cells grown on gold grids were also preparedfor visualization in the electron microscope by the whole mount critical point drying technique as described (2).

RESULTS

Electron microscopic visualization of the subcellular networks in SV40 infected cells

In order to examine the possible involvement of the cytoplasmic and nuclear networks in SV40 macromolecular metabolism we applied a sequential fractionation scheme that separates the various subcellular structural fractions and preserves morphologic constituents as follows: At late times after infection with SV40 the cells were exposed to a buffer

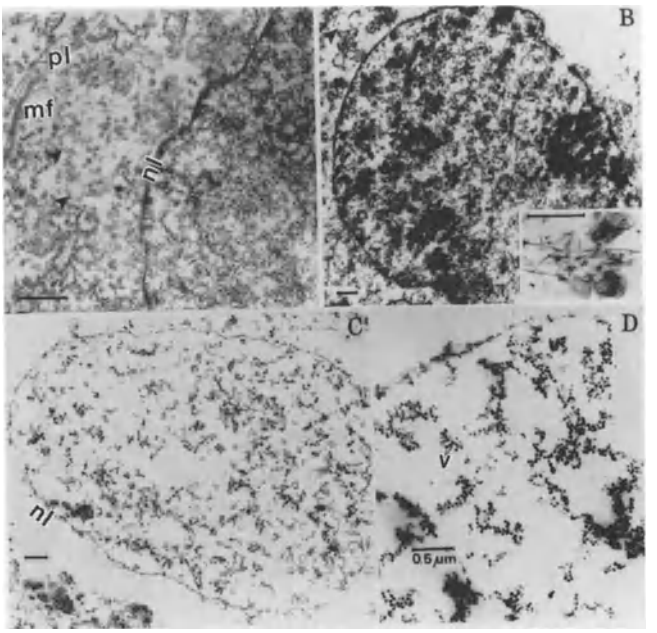


Fig. 1. Electron microscopic characterization of structural frameworks in the cytoplasm and the nucleus of SV40-infected BSC-1 cells at 42 h post-infection. (A) Cytoskeleon of SV40-infected BSC-1 cells prepared by Triton X-100 extraction. (B) Nuclei after DOC/Tween homogenization of the cytoskeletal framework. (C) Nuclear matrix obtained after DNase and salt extraction of nuclei prepared as in (B). (D) same as (C) at higher magnification. Inset and arrows show 'vacuoles' induced by SV40 infection and surrouned by intermediate filaments co-sedimenting with nuclei. v - virions, mf- microfilaments, nl - nuclear lamina, pl - plasma lamina. Bars 0.5 μM. (Reprinted with permission from Ben-Ze'ev et al (6)).

containing Triton X-100. This procedure removes over 90% of the lipids together with about 50-60% of total cellular protein depending on the cell culture density. The release of soluble proteins leaves behind a self-supporting interconnected cellular framework (Fig. 1A). This framework contains the nucleus which is packed with virions and which is bounded by the nuclear lamina. In the cytoplasmic side, the filamentous network to which the polyribosomes are attached can also be seen. A much better demonstration of the filamentous characteristic of the cytoplasmic framework can be seen in the whole-mount embedment-free electron micrograph shown in Fig. 2A, where the abundance of filaments is by far more impressive than in the conventional plastic embedded thin section

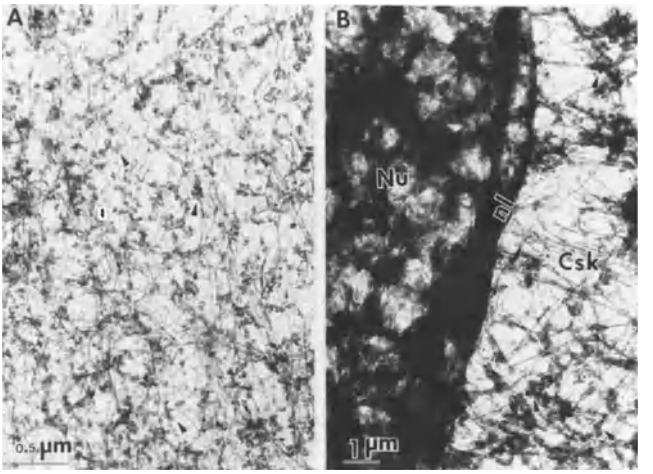


Fig. 2. The cytoplasmic framework visualized by the whole mount embedment-free technique. Cells grown on gold grids were extracted with Triton X-100 and prepared for electron microscopy by the critical point drying technique (2). A - the cytoplasmic framework showing the polyribosomes (arrowheads) attached to the cytoskeletal framework, B - area close to the nucleus showing the termination of the cytoskeletal filaments at the nuclear lamina (NL), Nu - nucleus, Csk - cytoskeleton.

shown in Fig. 1A. The attachment of the polyribosome to the cytoplasmic framework can also be better visualized by this technique (Fig. 2A), which shows the termination of the cytoplasmic filaments at the nuclear lamina boundary (Fig. 2B). The cellular framework maintains overall cell shape and detailed cellular morphology after the removal of lipids and soluble cellular proteins and is covered by a proteinacious lamina which is derived from plasma membrane proteins [Fig. 1A and (5)]. This lamina contains the binding sites for certain viruses (5), glycoproteins and also the submembranal proteins involved in the formation of adhesion plaques.

In the next step we separated the majority of the cytoplasmic framework from the remnant nucleus by extraction with a mixture of deoxycholate and Tween-40 (Fig. 1B). This treatment removes the cytoplasmic constituents except the intermediate filaments (IF) which are insoluble under the extraction procedures used here and which cosediment with "vacuoles" characteristic of SV40 infected cells (Fig. 1B inset).

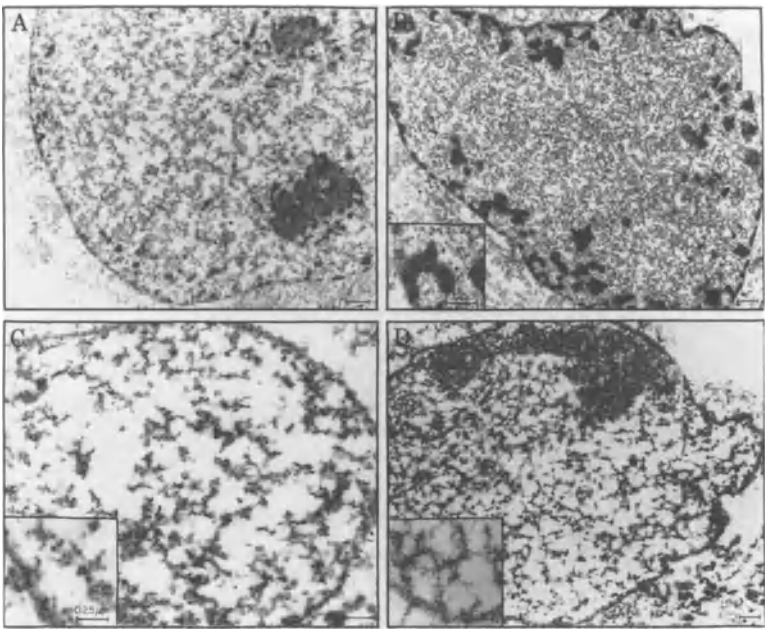


Fig. 3. Electron microscopic characterization of steps in the preparation of nuclear matrices from uninfected and SV40-infected BSC-1 cells. (a) Nucleus from SV40-infected cell after extraction with Triton X-100/Saponine (12). (b) Nucleus from (a) after treatment with DNase I showing the electron-dense histone patches. Inset, magnification showing the residual nuclear lamina and histones. (c) Nuclear matrix obtained from (b) after treatment with 0.25 M ammonium sulfate. The inset is a magnification of nuclear matrix from SV40-infected cells showing a filamentous network that holds the virions. (d) Nuclear matrix from uninfected BSC-1 cells obtained as in (c). The inset shows the filamentous property of the matrix. The bars represent 0.25 μm. (Reprinted with permission from Abulafia et al (26)).

In order to study the existence of a possible association between SV40 macromolecules and the nuclear matrix we have treated the nuclei as obtained in Fig. 1B with DNase and then with salt (2 M NaCl) to remove over 95% of DNA and histones. The nuclear matrix obtained by this procedure (Fig. 1C) consists of an outer boundary known as the nuclear lamina and an internal network which is covered by virions (Fig. 1C and 1D). The three dimensional organization of the nuclear matrix network is again poorly visualized in the thin section plastic-embedded electron micrograph shown in Figs 1C and 1D. Figure 3A-D is much better in visualizing the intranuclear structure since the extraction procedure and

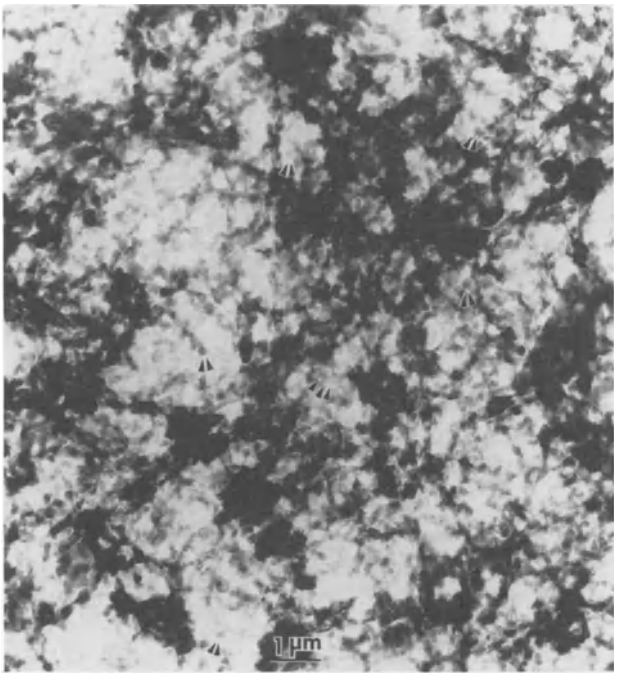


Fig. 4. The nuclear matrix of SV40-infected cells visualized by the whole-mount critical-point drying technique. Arrows point to virions associated with the nuclear matrix.

the salt concentration used were milder. In Figure 3D a considerable amount of intranuclear structure including the remnant nucleolus is seen in uninfected cells (see also inset of Fig. 3D), and these structures are covered extensively by SV40 virions (Fig. 2C, 2D and 3C). The abundance of structure and the filamentous nature of the nuclear matrix can be visualized best by the resinless whole-mount electron microscopy, as shown in the electron micrograph of a portion of the nuclear matrix of SV40 infected cells (Fig. 4). Here the abundance of intranuclear structure and its complexity makes the identification of virions difficult. Nevertheless this approach for visualization gives a much more realistic view of the organization of the nuclear matrix (see also (2)).

Association of SV40 and viral proteins with the nuclear matrix

The association of mature SV40 with preparations of nuclear matrix was demonstrated by showing a marked increase in the percentage of DNA resistant to treatment with DNase I in SV40-infected cells (20%), when compared to uninfected cells (2-5%) (6). This DNase resistant DNA was interpreted as encapsidated viral DNA in mature virions which is protected against the enzymatic attack. Experiments with a tsB mutant of SV40 (strain 777) further support this interpretation, since a short shift up (1 h) to the nonpermissive temperature in cells infected with this tsB mutant dramatically reduce the percentage of DNase I resistant DNA to the level of uninfected cells. This results from virus disassembly at the nonpermissive temmperature (6), that makes the previously encapsidated viral DNA amenable for enzymatic attack. The cofractionation and possible association of SV40 particles with the nuclear matrix is further supported by the analysis of equivalent amounts of each of the fractions, obtained as described in Fig. 1, on SDS-PAGE (Fig. 5). Both Coomassie blue staining (Fig. 5I) and autoradiography of ^{3}H-leucine-labeled cells (Fig. 5II) show the retention of the major viral capsid proteins VP1 and VP3 with the nuclear matrix fraction and very little, if any, of these viral proteins were detected in other subcellular fractions (see Fig. 5a-f). Note also the protection of a small amount of histones in nuclear matrix preparations of SV40 infected cells (Fig. 5D-F) and not in nuclear matrices of uninfected cells (Fig. 5D'-F'). This fraction of histones represent histones sequestered in mature virions with viral DNA which is organized in the form of minichromosomes.

The kinetics of association of newly synthesized viral capsid proteins with the nuclear matrix and their assembly into virions was followed in "pulse-chase" experiments. SV40-infected cells were labeled with ^{35}S-methionine and then a portion of the cells were washed extensively and were incubated for various periods of time in a medium with excess unlabeled methionine (Fig. 6). The cells were fractionated into the cytoplasmic soluble (S), cytoskeletal (SK), chromatin (D) and nuclear matrix (M) fractions as described in Fig. 1, and equivalent amounts of radioactive proteins were analyzed by SDS-PAGE. It is evident from Fig. 6 that even after a short pulse (Fig. 6A) the majority of the

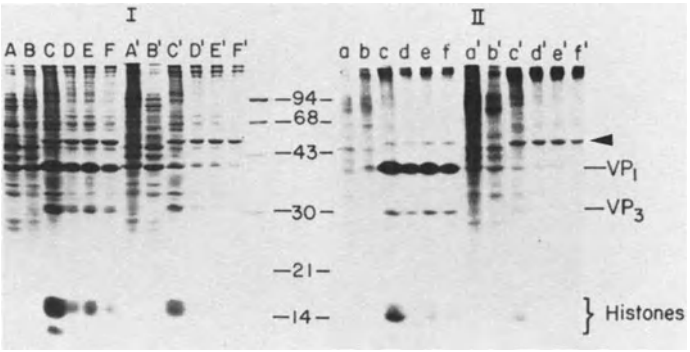


Fig. 5. Polyacrylamide gel electrophoresis of proteins of the various fractions obtained during preparation of subcellular frameworks. SV40-infected and mock-infected BSC-1 cells were labeled at 42 h post-infection for 3 h with [^{3}H]leucine (100 μCi/ml). Equivalent amounts of material from the various fractions were analyzed on 15% acrylamide gels. (I) Coomassie blue staining,(II) autoradiography of the same lanes as in I. (A) Triton X-100 soluble material (cytoplasm), (B) DOC/Tween solubilized cytoskeleton, (C) nuclei obtained after DOC/Tween treatment, (D) nuclear matrix obtained with DNase and 1 M NaCl, (E) as (D) obtained with DNase and 0.3 M $(NH_4)_2SO_4$, (F) as (D) but with DNase and 0.5 M $(NH_4)_2SO_4$. A' - F' the same as A - F but of mock-infected cells. a - f the autoradiogram of A - F, and a' - f' the autoradiogram of A' - F'. Arrow indicates position of the 58-KDa intermediate filament protein vimentin. (Reprinted with permission from Ben-Ze'ev et al (6)).

viral capsid proteins VP1 and VP3 are already associated with the nuclear matrix fraction. The radioactively labeled histones, on the other hand, were found in the chromatin fraction. At later times after the initiation of the "chase" (Fig. 6E-H) the retention of some of the histones in the nuclear matrix fraction could be observed, representing most probably histones incorporated into the maturing virions. The kinetics of virus maturation estimated by this procedure were similar to the rate determined by other techniques (7). These results together with the electronmicroscopic observations suggest a functional association between SV40 capsid proteins and their maturation into virions with the nuclear matrix fraction.

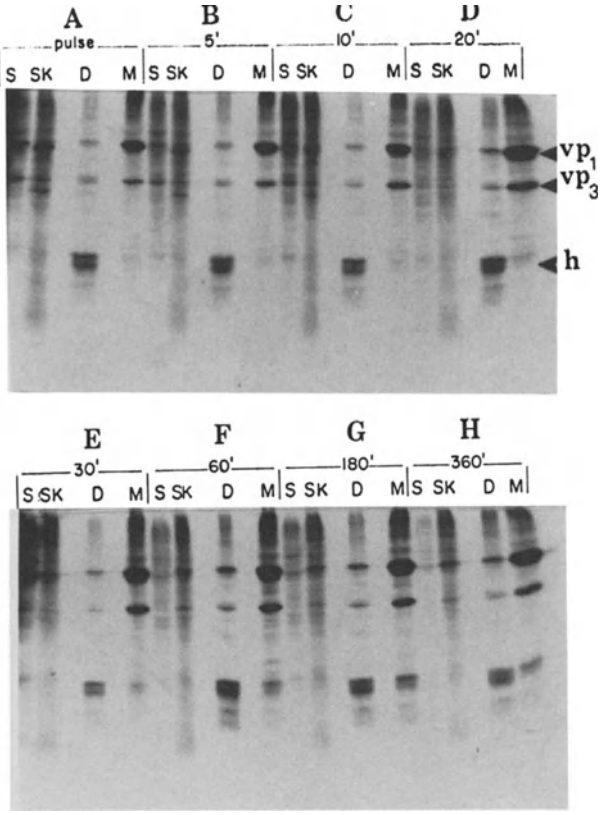


Fig. 6. The assembly of SV40 virions is associated with the nuclear matrix. SV40-infected cells were labeled for 10 min in methionine-deficient medium with 100 μCi [^{35}S]methionine. The chase was initiated by washing the plates with fresh medium and incubating the cells in medium containing unlabeled methionine. At different times after the initiation of the chase, from equal numbers of cells, the following cell fractions were prepared: S - a Triton X-100 soluble cytoplasmic fraction, SK - a DOC/Tween cytoskeletal fraction, D - a DNase and high salt soluble nuclear extract, M - a nuclear matrix. The proteins were concentrated and analysed on acrylamide gels. (A) 10 min [^{35}S]methionine pulse, (B) 5 min, (C) 10 min, (D) 20 min, (E) 30 min, (F) 60 min, (G) 180 min, (H) 360 min chase. VP1 and VP3 are the major viral capsid proteins, h - histones. (Reprinted with permission from Ben-Ze'ev et al (6)).

SV40 RNA biogenesis and its association with the nuclear matrix

The abundance of viral SV40 RNA transcripts in the nucleus at late times after infection enabled the following of synthesis, processing and transport of the newly synthesized viral RNA and the relationships between these processes and the nuclear matrix and the cytoskeletal framework. A pulse chase protocol was used which consisted of depleting the cells of the unlabeled intracellular UTP by incubating the cells for 60 min with glucosamine. The infected cells were then pulse labeled for 10 min with [5-^{3}H]uridine. The chase was initiated by washing the cells and incubating them in a medium containing glucosamine and unlabeled uridine, cytidine and thymidine to avoid extensive labeling of viral DNA (8). After different times of "chase" the cells were fractionated as shown in Figs. 1 and 6. The RNA of each fraction was extracted and separated on sucrose gradients and the virus specific ^{3}H-uridine labeled RNA in each fraction across the gradient was determined by hybridization with excess single-stranded (ss) SV40 DNA immobilized on filters. The newly synthesized viral RNA in the nucleus, with a characteristic major peak of 19S, was found in association with the nuclear-matrix fraction (Fig. 7A). At the end of 3 h chase there was a 5-6 fold decrease in the amount of viral RNA associated with the nuclear matrix, due to processing and transport to the cytoplasm (9) (Fig. 7B-E). In the chromatin fraction (Fig. 7F-J) only a minor proportion of the viral RNA was found either after a short pulse or during the chase. The results thus show a quantitative association of the synthesis, processing and transport of SV40 RNA with the nuclear matrix of infected cells. The data are in agreement with other studies showing an association between the newly synthesized cellular and several viral RNA species with the nuclear matrix in a variety of systems (10-15). However the pulse-chase experiments presented here are the first to demonstrate the association of the viral RNA during synthesis processing as well as during transport with the nuclear-matrix fraction.

Newly synthesized SV40 RNA appears in the cytoplasm in association with the cytoskeletal framework

When the distribution of virus-specific polyribosomes was analyzed in the Triton X-100 soluble cytoplasmic fraction and the DOC/Tween-40 soluble cytoskeletal fraction, over 80% of the virus specific polyribo-

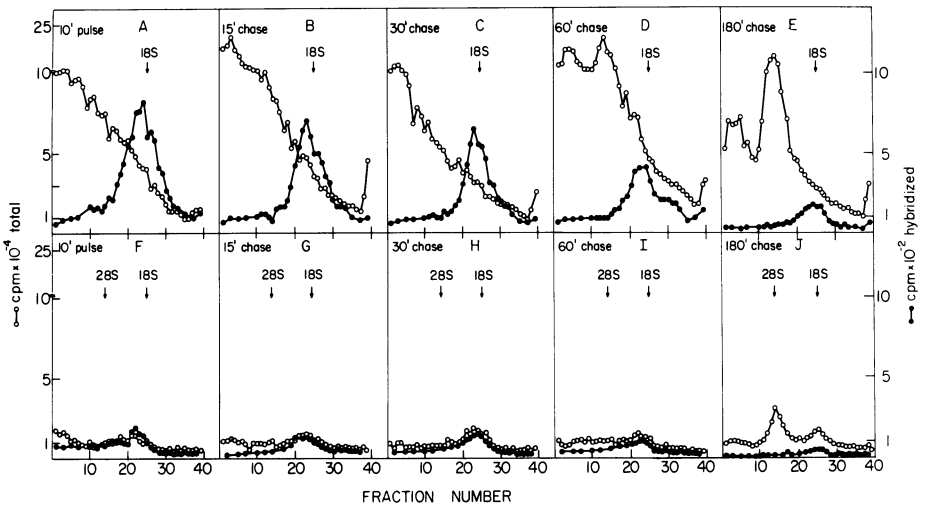


Fig. 7. SV40 RNA is associated with the nuclear matrix during its synthesis and processing in infected cells. At 42 h post-infection, SV40 infected BSC-1 cells were labeled with 200 μCi/ml of [5-^{3}H]uridine for 10 min. A chase was initiated by incubating the cultures with glucosamine and unlabeled uridine, cytidine, and thymidine. At different times during the chase the RNA from the nuclear matrix (A-E) and the DNase and salt soluble fraction (F-J) was extracted and analyzed by sedimentation through sucrose gradients. The amount of total RNA was determined by TCA precipitation and one half of each fraction was taken to hybridization with 1 μg of ss SV40 DNA immobilized on filters. O—O total RNA, ●—● SV40-specific RNA (Reprinted with permission from Ben-Ze'ev et al (6)).

somes were found in the cytoskeletal fraction (Fig. 8B) and only about 20% were found in the initial Triton X-100 soluble fraction (Fig. 8A). By following the distribution of viral RNA in these subcytoplasmic fractions in pulse-chase experiments, after a 10 min pulse and 15 min chase 19S viral RNA was found in association with the cytoskeleton (Fig. 9A) and almost no viral RNA in the Triton X-100 soluble fraction (Fig. 9B). During a 3 h chase period viral RNA continues to accumulate in the cytoskeletal fraction (Fig. 9C) in association with the polyribosomes, and the 16S viral RNA becomes its major constituent, while the Triton X-100 soluble fraction continues to have only a minor part (~15%) of the viral RNA (Fig. 9D). About 30-50% of the viral cytoplasmic RNA is poly(A)$^-$ RNA and this RNA also accumulates in the first 3 h of the chase in association with the cytoskeleton (16). The distribution of the viral

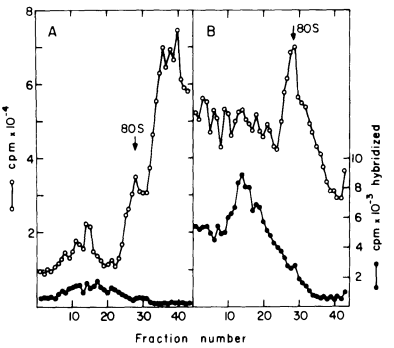


Fig. 8. Analysis of viral polyribosomes associated with the cytoskeletal framework and the soluble fraction. At 45 hr after infection the cells were labeled with 50 μCi/ml of [^{3}H]uridine for 3 hr and the soluble (A) and cytoskeletal framework (B) fractions were prepared. The two fractions were centrifuged through sucrose gradients in the SW 41 rotor for 60 min at 40,000 rpm and fractions were collected. Aliquots (100 μl) of each fraction were precipitated with TCA and counted (O). The same size aliquots were brought to 0.5% SDS and hybridized with SV40 DNA filters (●). (Reprinted with permission from Ben-Ze'ev et al (16)).

RNA in the various nuclear and cytoplasmic fractions in the pulse-chase experiments shown in Figs 7 and 9 is summarized in Fig. 10. These kinetics studies suggest that the rapidly labeled viral RNA is synthesized, processed and transported in association with the nuclear matrix, and after its appearance in the cytoplasm it becomes first associated with the cytoskeleton where the virus specific polyribosomes are localized. However, after longer periods of chase, both the poly(A)$^+$ and the poly(A)$^-$ viral RNA move into the soluble fraction of the cytoplasm (Fig. 11a), to reach, at steady state, about an equal distribution of viral RNA between the soluble and the cytoskeletal fractions in the cytoplasm (Fig. 11b). Since the cytoskeletal fraction was shown to contain all of the polyribosomes (17,18) the viral RNA has to be attached to the cytoskeletal framework in order to be translated. Studies with several cytoplasmic RNA viruses (19,20) and with the DNA nuclear viruses adeno (21,22) and herpes simplex (23,24) reached similar conclusions.

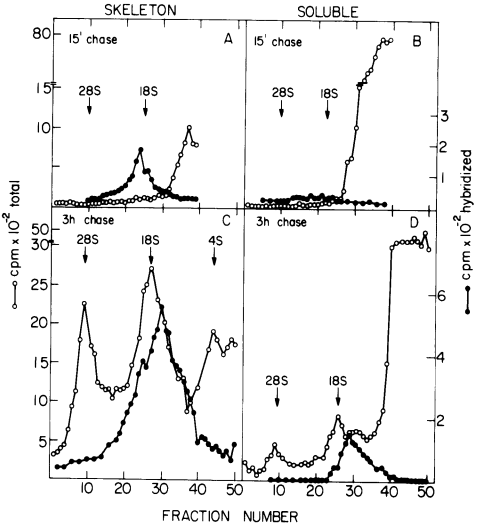


Fig. 9. Newly synthesized cytoplasmic SV40 RNA is associated with the cytoskeleton. SV40-infected BSC-1 cells were pulsed for 10 min with [5-^{3}H]uridine and chased with unlabeled uridine. After 15 min chase (A,B) and 3 h chase (C,D) the RNA from the Triton X-100 soluble fraction in the cytoplasm (B,D) and the cytoskeletal fraction obtained with DOC/Tween (A,C) was extracted and analysed on 15-30% sucrose gradients. SV40-specific RNA was determined by hybridizing one half of each fraction with 1 μg of ss SV40 DNA bound to filters. O—O. Total RNA determined by TCA precipitation, ●—● SV40-specific RNA. (Reprinted with permission from Ben-Ze'ev et al (6)).

The promoter-proximal nascent RNA of late SV40 transcription is not associated with the nuclear matrix

In order to study the active involvement of the nuclear matrix in the regulation of gene expression, we initiated experiments in in vitro systems (i.e. isolated nuclei and isolated nuclear matrices) to ask whether actively transcribing RNA polymerase molecules and transcriptionally active DNA sequences are associated with the nuclear matrix. A very convenient system for these studies was provided by the premature transcription termination that occurs during SV40 late transcription (25). In this system it was demonstrated that nascent RNA molecules which initiated in vivo at the major initiation site, were synthesized in isolated nuclei to a length of 94 nucleotides before terminating at a

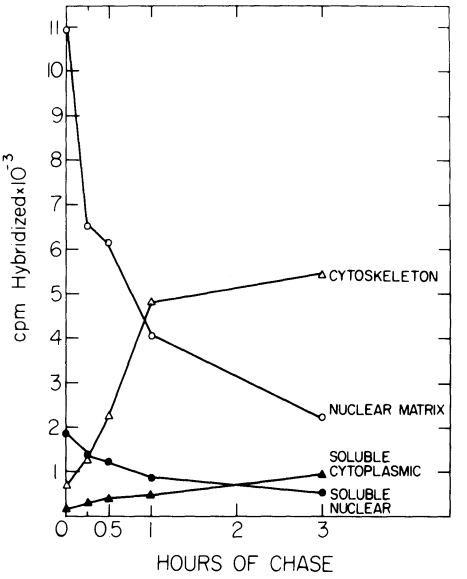


Fig. 10. The synthesis and processing of SV40 RNA are associated with the nuclear matrix and the mature SV40 RNA is transported to the cytoskeleton. SV40-infected BSC-1 cells were pulsed for 10 min with [5'-^{3}H]uridine and chased as described in Figure 7. At each time point, cells were fractionated into a Triton X-100 soluble cytoplasmic fraction, a DOC/Tween homogenate that contains the cytoskeleton, a DNase and salt-soluble nuclear fraction, and a nuclear matrix. The RNA from each fraction was extracted and analyzed on 15-30% sucrose gradients. Each fraction across a sucrose gradient was hybridized with ss SV40 DNA bound to filters. The SV40-specific radioactivity is summed across each gradient in each subcellular fraction after 10 min pulse, and after 15 min, 30 min, 60 min, and 180 min of chase. O—O SV40 RNA from the nuclear matrix, Δ—Δ SV40 RNA associated with the cytoskeleton, ●—● SV40 RNA removed from nuclei by DNase and 0.5 M $(NH_4)_2SO_4$, ▲—▲ SV40 RNA from the cytoplasm obtained in the initial Triton X-100 lysis. (Reprinted with permission from Ben-Ze'ev et al (6)).

typical transcription-termination structure (25). Pretreating the cells with DRB was found to enhance the premature termination, while the treatment with proflavine, which is known to interfere with RNA secondary structure, allows readthrough (25).

In order to analyze the association of the *in vitro* elongated RNA with the nuclear matrix, isolated nuclei from SV40 infected cells were incubated briefly (for 5 min) *in vitro* with [α-^{32}P]UTP to label the *in vivo* initiated nascent RNA molecules. The RNA from the nuclear-matrix fraction and from the chromatin fraction (DNase and salt soluble) prepared from these ^{32}P-UTP labeled nuclei was isolated and analyzed on

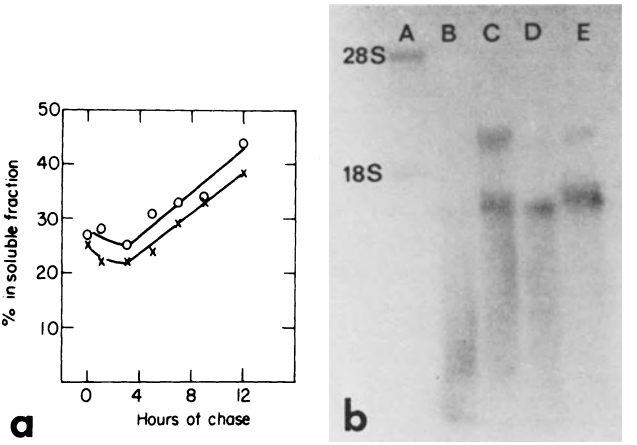


Fig. 11 (a). Transit of virus-specific RNA from the cytoskeletal framework to the soluble fraction. SV40-infected cells were pulse labeled for 20 min with [^{3}H]uridine and chased for various times in the presence of glucosamine and cold uridine. The amount of labeled poly(A)$^+$ (O) and poly(A)$^-$ (x) viral RNAs in the cytoskeletal framework and soluble fractions at each time point during the chase were determined by hybridization to excess SV40 DNA bound to nitrocellulose filters. The summation of hybridizable counts per minute of the cytoskeletal framework and soluble RNAs at each time point were taken as 100%.

Fig. 11 (b). Viral RNA analysis by the RNA transfer method. The poly(A)$^+$ and poly(A)$^-$ RNA molecules in each of the subcytoplasmic fractions were analyzed by RNA blotting and hybridization to nick translated SV40-^{32}P-labeled DNA. (A) ^{32}P-labeled 18 S and 28 S ribosomal RNA; (B) cytoskeletal poly(A)$^-$ RNA; (C) cytoskeletal poly(A)$^+$ RNA; (D) soluble poly(A)$^-$ RNA; (E) soluble poly(A)$^+$ RNA. (Reprinted with permission from Ben-Ze'ev et al (16)).

sucrose gradients (Fig. 12). While the majority of the labeled RNA molecules in both fractions consisted of short RNA molecules, the nuclear matrix fraction contained also long viral RNA molecules (Fig. 12A), whereas the chromatin fraction (Fig. 12B) contained exclusively the short viral RNA molecules. These radioactively labeled molecules hybridize preferentially with the promoter-proximal region of the late SV40 transcribed genome (restriction fragment e) (Fig. 12B inset). The labeled RNA molecules associated with the nuclear matrix fraction, which are much longer, hybridize with restriction DNA fragments which are distal to the major initiation site (Fragments b and d, Fig. 12A inset).

The 94 nucleotide long promoter-proximal attenuator RNA which accumulates in nuclei of SV40 infected cells (Fig. 13B,N) maps between the major initiation site at nucleotide 243 and terminates at nucleotide 336

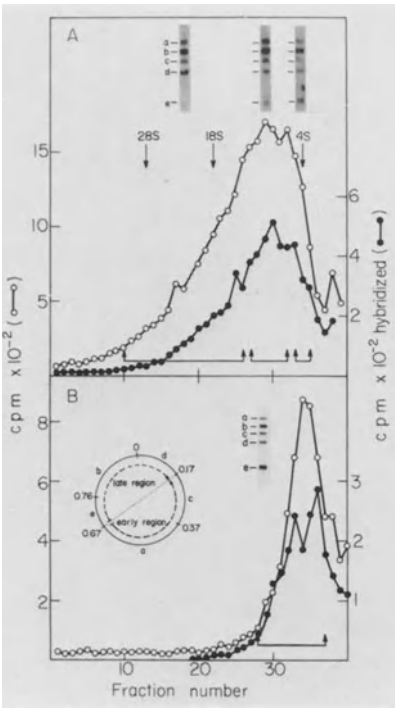


Fig. 12. The in vitro transcribed SV40 RNA shorter than 5S is not associated with the nuclear matrix. Isolated nuclei obtained from SV40-infected cells by the Triton X-100/Saponine method were incubated in vitro for 5 min with a transcription mixture containing [α-^{32}P]UTP. From the ^{32}P-labeled nuclei, (a) a nuclear matrix fraction and (b) a DNase plus salt soluble fraction were prepared and [^{32}P]RNA extracted from each fraction was analyzed by sedimentation through SDS/sucrose gradients. Total RNA (O) in each fraction was determined by precipitation with TCA and SV40-specific RNA (●) was identified by hybridizing each fraction with SV40 DNA on filters. Fractions across the sucrose the sucrose gradients were pooled as indicated in the Figure and hybridized to SV40 DNA restriction fragments obtained by cleavage of SV40 DNA with EcoRI, BglI land HpaI restriction enzymes. (Reprinted with permission from Abulafia et al (26)).

(25). This RNA is not bound to the nuclear matrix (Fig. 13B,M), but it is found in the chromatin fraction (Fig. 13B,S). The pretreatment of cells with DRB enhances the accumulation of the attenuator RNA, which is the almost exclusive RNA found in the nuclei (Fig. 13A, DRB), but again this RNA is found in the DNase and salt soluble fraction (Fig. 13B, DRB,

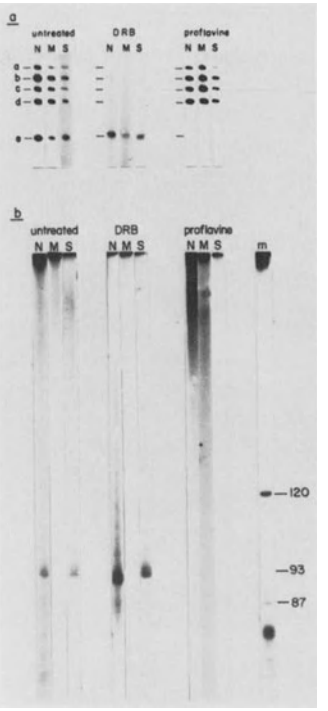


Fig. 13. A 94-nucleotide long SV40 RNA that hybridizes to the SV40 promoter-proximal DNA is not associated with the nuclear matrix. (a) Isolated nuclei (N) prepared from an equal number of untreated, DRB-treated (75 μM, 40 min) or proflavine-treated (80 mM, 2 min) SV40-infected cells were incubated with [α-^{32}P]UTP. The nuclear matrix fraction (M) and a DNase plus salt-soluble fraction (S) were prepared, and ^{32}P-labeled RNA was extracted and hybridized to the 5 restriction fragments as described in Fig. 12. (b) An equal portion of each RNA sample was hybridized to, and eluted from SV40 DNA on filters, and the SV40-specific RNA molecules were analyzed by gel electrophoresis in 7 M-urea/12% acrylamide gels. m, ^{32}P-labeled tRNA markers. (Reprinted with permission from Abulafia et al (26)).

S). In cells pretreated with proflavine the *in vitro* synthesized ^{32}P-labeled RNA does not hybridize with the promoter-proximal restriction fragment e (Fig. 13A) and on the gel there is no detectable 94 nucleotide long band.

The relative amount of *in vitro* labeled viral RNA in the DNase and salt-soluble fraction increases by 50% in DRB pretreated cells (Table 1), while in proflavine-pretreated cells there is a twofold decrease in

TABLE 1 The distribution of SV40 RNA synthesized in isolated nuclei between the nuclear matrix and DNase plus salt soluble fraction in infected cells, and untreated, or pretreated with DRB or proflavine

	Untreated		DRB treated		Proflavine	
	cts/min hydridized	%	cts/min hybridized	%	cts/min hybridized	%
Total	66,800	100	46,178	100	92,418	100
Matrix	47,070	70.5	26,033	56.4	78,675	85.1
Soluble	19,730	29.5	20,145	43.6	13,743	14.9

Nuclei were isolated from SV40 infected cells, untreated or pretreated with DRB or proflavine and after incubation with ^{32}P-UTP the nuclear matrix fraction and the DNase plus salt soluble fraction were prepared. ^{32}P-RNA was isolated from whole nuclei or from the two subnuclear fractions. Total radioactive RNA was determined by TCA precipitation, while the amount of SV40 specific RNA was determined by hybridization to excess SV40 DNA bound to filters. RNA from an equal number of cells was analyzed in the hybridization experiments.

the proportion of labeled RNA in the soluble fraction as compared to control cells (Table 1). Concomitantly, there is an increase in the amount of labeled viral RNA in the nuclear-matrix fraction, suggesting that, while the promoter-proximal attenuator RNA accumulates in the soluble fraction, RNA molecules elongated beyond the attenuation site are bound to the nuclear matrix.

The promoter-proximal and other actively transcribed viral DNA sequences are associated with the nuclear matrix.

In the previous experiments the nascent RNA molecules were elongated in vitro in isolated nuclei and the partitioning of the labeled viral RNA species between the nuclear matrix and the DNase and salt-soluble chromatin fraction was determined. In the following experiments we used isolated nuclear matrices to ask whether actively transcribing RNA polymerase molecules and transcriptionally active DNA sequences are found in association with these isolated matrices. When the matrix preparations [prepared in 250 mM $(NH_4)_2SO_4$] were incubated with [α-^{32}P]UTP they were capable of synthesizing the promoter-proximal attenuator RNA molecules as indicated by the extensive hybridization of the labeled RNA with the e fragment (Fig. 14 I) and the appearance of the 94 nucleotide long band on gels (26). However, the promoter-proximal RNA molecules were not tightly bound to the matrix, since they were preferentially removed from the nuclear matrix with a wash step in 250 mM $(NH_4)_2SO_4$ (Fig. 14 III). The longer promoter-distal viral RNA molecules remained associated with the nuclear matrix (Fig. 14 II) and so did the cellular small molecular weight RNA species varying in size between 80-200 nucleotides (27,28).

The above results indicate that active RNA polymerase molecules are associated with purified matrices. Table 2 summarizes quantitative analyses of transcribed and total DNA sequences associated with the nuclear matrix. Whole nuclei and nuclear matrices were prepared from infected cells which were pulse labeled with [^{3}H]thymidine in the presence or the absence of DRB and proflavine. The amount of virus-specific DNA in the various fractions was determined by DNA-DNA hybridization with ss SV40 DNA bound to filters. Nuclear matrices from treated and control cells contain between 4.0% and 8.7% of the labeled nuclear DNA. The labeled viral DNA comprises about 50% of the nuclear or the nuclear matrix asso-

TABLE 2 Quantitative analyses of SV40 DNA and transcription activity on isolated matrices

Treatment	Nuclear fraction	Total ^{3}H - DNA cts/min	%	Viral ^{3}H - DNA cts/min	%	Total ^{32}P - UMP cts/min	%	Viral ^{32}P- UMP cts/min	%
Untreated	Total	2,575,800	100	1,603,767	100	3,244,800	100	34,699	100
	Matrix	105,204	4.0	46,239	2.8	693,937	21.3	12,611	36.3
DRB	Total	2,339,400	100	1,474,363	100	2,360,994	100	31,892	100
	Matrix	107,000	4.5	45,172	3.0	579,372	24.5	16,801	52.6
Proflavine	Total	3,558,800	100	2,567,050	100	1,902,826	100	57,026	100
	fraction	309,620	8.7	158,745	6.1	950,225	49.9	28,052	49.1

Nuclei or nuclear matrices from SV40 infected cells, untreated or treated with DRB or proflavine were prepared as described in TABLE 1. For measuring the amount of SV40 DNA remaining with the nuclear matrix preparations the infected cells were labeled for 2 hr with (^{3}H)-thymidine. Total DNA was determined by precipitation of the purified DNA and SV40 specific DNA was determined by DNA-DNA hybridization of an equal amount of TCA precipitable radioactivity to filters containing SV40 DNA. For determining the *in vitro* trascriptional activity, isolated nuclei or nuclear matrices were incubated with ^{32}P-UMP and the ^{32}P-labeled RNA was isolated. Total incorporation was determined by TCA precipitation and the amount of SV40 specific RNA was quantitated by hybridizing RNA from an equal number of cells with SV40 DNA on filters.

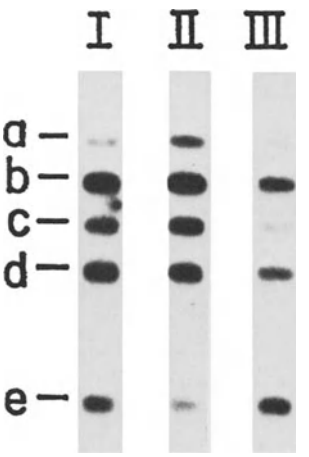


Fig. 14. The promoter-proximal viral RNA molecules are not tightly associated with the nuclear matrix. Nuclear matrices prepared from SV40-infected cells by DNase digestion followed by washing with 250 mM-ammonium sulfate were incubated with [α-^{32}P]UTP for 5 min. Following incubation, the matrices were washed again with 250 mM-ammonium sulfate for removing the weakly bound RNA molecules. [^{32}P]RNA was extracted from (I) whole matrices, (II) washed matrices and (III) from the ammonium sulfate soluble fraction, and hybridized to SV40 restriction fragments as in Fig. 12. (Reprinted with permission from Abulafia et al (26)).

ciated labeled DNA, and only about 2.8% to 6.1% of the total nuclear labeled viral DNA is associated with the nuclear matrix. RNA polymerase activity in the various nuclear matrix fractions was determined by a brief (5 min) incubation of the isolated matrix preparations with [^{32}P]UTP. The results in Table 2 show that 2.8% to 6.1% of the viral nuclear matrix-associated DNA synthesized between 36.2% to 52.6% of the labeled viral RNA as compared to whole nuclear preparations. Since the nuclear matrix-associated DNA is composed of short DNA fragments (those remaining after DNase and salt treatment), it is possible that part of the RNA polymerase molecules run off their template during five minutes of incubation. This might cause an underestimation of RNA polymerase activity. The kinetics studies of [^{32}P]UMP incorporation show that during very short times of incubation (0.5 to 1 min) between 70% to 90% of the viral RNA synthesized in whole nuclei is synthesized by purified nuclear matrices (Fig. 15 inset). At longer periods of incubation at least 35% of the RNA polymerase activity found in isolated nuclei is

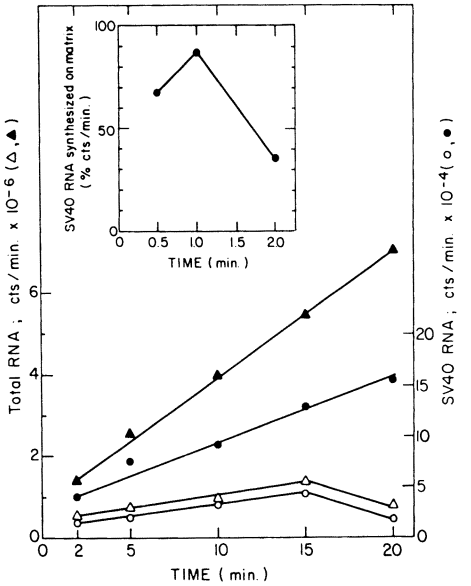


Fig. 15. Kinetics of total and viral RNA synthesis in whole nuclei and in nuclear matrices. Nuclei and nuclear matrices isolated from an equal number of SV40-infected cells were incubated in vitro with [α-^{32}P]UTP. ^{32}P-labeled RNA was isolated after various durations of incubation and the amount of total RNA was determined by precipitation with TCA. SV40-specific RNA was determined by hybridizing RNA from an equal number of cells with SV40 DNA on filters. (▲) Total nuclear RNA. (●) SV40-specific nuclear RNA. (Δ) total matrix RNA. (O) SV40-specific matrix RNA. inset - Kinetics for short period of incubation. (Modified after Abulafia et al(26)).

associated with the nuclear matrix (Fig. 15). Hybridization of the ^{32}P-labeled RNA synthesized by purified nuclear matrices and whole nuclei from DRB, proflavine and control cells with viral DNA restriction fragments showed an identical pattern between matrices and whole nuclei of the respective treatments (26). This suggests that all or almost all actively transcribing viral DNA sequences are associated with the nuclear matrix together with active RNA polymerase molecules.

DISCUSSION

In the experiments described in this chapter we have used various steps in the replication of SV40 in monkey kidney cells as a model system to study the involvement of the complex cellular networks in the regulation of gene expression. The electronmicroscope analyses have shown an association between mature SV40 virions and the nuclear matrix. The use of the sequential fractionation scheme employed in this study is a powerful tool in the investigation of the function and the molecular details of the role of the cellular networks of the cytoplasm and of the nucleus in gene expression and viral replication. Pulse-chase experiments followed by this sequential fractionation procedure revealed that the encapsidation of the viral minichromosome into mature virus is associated with the nuclear matrix.

Our in vitro studies with isolated nuclear matrices have shown a more direct role for the nuclear marix in the control of SV40 gene expression. The study of the premature transcription-termination process during late SV40 transcription has shown that the nascent viral RNA does not associate immediately with the nuclear matrix, since the 94-nucleotide long attenuator RNA is not associated with the nuclear matrix. The association occurs only after the nascent RNA has elongated beyond a critical length. In the case of SV40, it is possible that the attachment of the nascent RNA to the matrix occurs only after certain antiattenuation factor(s) influence RNA secondary structure, perhaps relieving constrains (29) that may prevent the binding to the matrix. Following the association of the nascent RNA with the nuclear matrix, transcription proceeds and the primary viral transcript is produced and transported into the cytoplasm. The pulse-chase experiment has indicated an association of the newly synthesized viral RNA with the nuclear matrix throughout these steps. The viral mRNA, after reaching the cytoplasm, was first found in association with the cytoskeletal framework, where it is involved in viral protein synthesis. Since the transcription initiation factors (30) as well as aminoacyl-transfer RNA synthetase complexes (31) and all the polyribosomes (17) are associated with this framework, there seems to be an obligatory association between the viral mRNA and the cytoskeletal framework in order for the former to be translated, although this binding alone might not be sufficient (32).

The binding of mRNA to elements of the structural framework may have topographical significance as suggested by recent studies showing a non random localization of various mRNA molecules in the cytoplasm employing in situ hybridization experiments (33).

Our studies have also shown that purified nuclear matrices qualitatively and quantitatively preserve the ability to synthesize RNA similarly to isolated nuclei, thus implying that transcription of all or almost all viral DNA sequences occurs in association with the nuclear matrix. While this result is in agreement with that of several other laboratories (15,27,34-37), our is the first direct demonstration that transcription occurs on the matrix. The model that we suggest for the matrix associated transcription is based on the fixed transcription concept (Fig. 16). According to this model the viral DNA can associate with the matrix at various sites, and at these same sites the transcribed molecules are attached to the matrix. These sites harbor the factors necessary for the initiation of transcription including the RNA polymerase molecules. The transcripts are generated as the DNA is reeled through the fixed transcription complex sites (11). The replication of DNA was also suggested to occur on the nuclear matrix at fixed sites (38) and an enrichment of the α DNA polymerase on the nuclear matrix was reported (39,40). An interesting possibility would be to have a common fixed site for transcription and replication. In this regard it is interesting to note that the same sequences near the origin of replication of SV40 DNA were shown to be necessary for both SV40 DNA replication and transcription (41). These might be the sequences associated with the fixed site of replication and transcription via the viral T-antigen, which was also found in association with the nuclear matrix (42). Deletion of these sequences would therefore inhibit both replication and transcription. The molecular components and mechanisms underlying these associations between the nuclear matrix, the cytoskeletal framework and viral gene replication, expression and virus assembly are yet to be elucidated. The morphologic tools together with improved cell fractionation schemes will allow the detailed analysis at the biochemical and the molecular levels of these aspects of cell and virus research.

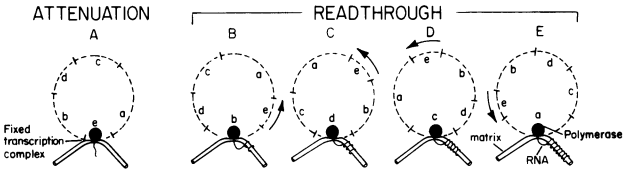


Fig. 16. Schematic representation of the fixed-site transcription model in late SV40 transcription.

ACKNOWLEDGMENT. The studies presented in this chapter were supported by a grant from the USA-Israel Binational Foundation, BSF, Jerusalem, Israel.

REFERENCES

1. The Journal of Cell Biology. 99: 1s-238s, 1984.

2. Fey, E.G., Krochmalnic, G. and Penman, S. J. Cell Biol. 102: 1654-1665, 1986.

3. Luftig, R. B. J. Theor. Biol. 99: 173-191, 1982.

4. Penman, S. In. Virology (Ed. B. N.Fields et al.) Raven Press, New York, 1985, pp. 169-182.

5. Ben-Ze'ev, A., Duerr, A., Solomon, F. and Penman, S. Cell 17: 859-865, 1979.

6. Ben-Ze'ev, A., Abulafia, R. and Aloni, Y. EMBO J. 1: 1225-1231, 1982.

7. Jakobovits, E. B. and Aloni, Y. Virology 102: 107-118, 1980.

8. Ben-Ze'ev, A. and Aloni, Y. Virology 125: 475-479, 1983.

9. Chiu, N. H., Radonovich, M. F., Thoren, M. M. and Salzman, N. P. J. Virol. 28: 590-599, 1978.

10. Herman, R., Weymouth, L. and Penman, S. J. Cell Biol. 78: 663-674, 1978.

11. Jackson, D. A., McCready, S. J. and Cook, P. R. Nature 292: 552-555, 1981.

12. Long, B. H., Huang, C.-Y. and Pogo, A. O. Cell 18: 1079-1090, 1979.

13. van Eekelen, C. A. G. and van Venrooij, W. J. J. Cell Biol. 88: 554-563, 1981.

14. Gallinaro, H., Puvion, E., Kister, L. and Jacob, M. EMBO J. 2: 953-960, 1983.

15. Ross, D. A., Yen, R. W. and Chae, C.B. Biochemistry 21: 764-771, 1982.

16. Ben-Ze'ev, A., Horowitz, M., Skolnik, H., Abulafia, R., Laub, O. and Aloni, Y. Virology 111: 475-487, 1981.

17. Cervera, M., Dreyfuss, G. and Penman, S. Cell 23: 113-120, 1981.

18. Lenk, R., Ransom, L., Kaufmann, Y. and Penman, S. Cell 10: 67-78, 1977.

19. Lenk, R. and Penman, S. Cell 16: 289-301, 1979.

20. Bonneau, A. M., Darveau, A. and Sonenberg, N. J. Cell Biol. 100: 1209-1218, 1985.

21. van Venrooij, W. J., Sillekens, P. T. G., van Eekelen, A. G. and Reinders, R. J. Exp. Cell Res. 135: 79-91, 1981.

22. van Eekelen, C. A. G., Ohlsson, R., Phillipson, L., Mariman, E., van Beek, R. and van Venrooij, W. Nucleic Acids Res. 10: 7115-7131, 1982.

23. Ben-Ze'ev, A., Abulafia, R. and Bratosin, S. Virology 129: 501-507, 1983.

24. Quinlan, M. and Knipe, D. Mol. Cell. Biol. 3: 315-324, 1983.

25. Hay, N., Skolnik-David, H. and Aloni, Y. Cell 29: 183-193, 1982.

26. Abulafia, R., Ben-Ze'ev, A., Hay, N. and Aloni, Y. J. Mol. Biol. 172: 467-487, 1984.

27. Maundrell, K., Maxwell, E. S., Puvion, E. and Scherrer, K. Exp. Cell Res. 136: 435-445, 1981.

28. Miller, T., Huang, C. Y. and Pogo, A. O. J. Cell Biol. 76: 692-702, 1978.

29. Aloni, Y., Hay, N., Skolnik-David, H., Pfeiffer, P., Abulafia, R., Ben-Asher, E., Jakobovits, E. B., Laub, O. and Ben-Ze'ev, A. In: Developments in Molecular Virology (Ed. Y. Becker), Martinus Nijhoff, Boston, vol. 4, 1983, pp. 1-48.

30. Howe, J. G. and Hershey, J.W. Cell 37: 85-93, 1984.

31. Mirande, M., Lecorre, D., Louvard, D., Reggio, H., Pailliez, J. P. and Waller, J. P. Exp. Cell Res. 156: 91-102, 1985.

32. Nielsen, P., Goetz, S. and Trachsel, H. Cell Biol. Int. Rep. 7: 245-254, 1983.

33. Lawrence, J. B. and Singer, R. H. Cell 45: 407-415, 1986.

34. Ciejek, E. M., Tsai, M.-J. and O'Malley, B. W. Nature (Lond.) 306: 607-609, 1983.

35. Hentzen, P. C., Rho, J. H. and Bekhor, I. Proc. Natl. Acad. Sci. USA 81: 304-307, 1984.

36. Robinson, S. I., Nelkin, B. D. and Vogelstein, B. Cell 28: 99-106, 1982.

37. Small, D., Nelkin, B. and Vogelstein, B. Nucleic Acids Res. 13: 2413-2431, 1985.

38. Pardoll, D. M., Vogelstein, B. and Coffey, D. S. Cell 19: 527-536, 1980.

39. Smith, H. C. and Berezney, R. Biochemistry 21: 6751-6761, 1982.

40. Jones, C. and Su, R. T. Nucleic Acids Res. 10: 5517-5532, 1982.

41. Contreras, R., Gheysen, D., Knowland, J., Van der Voorde, A. and Fiers, W. Nature(Lond) 300: 500-505, 1982.

42. Verderame, M. F., Kohtz, D. S. and Pollack, R. E. J. Virol. 46: 575-583, 1983.

13

MOLECULAR BIOLOGY OF PAPILLOMA VIRUS

H. PFISTER, E. KLEINER, G. LANG, G. SAGNER, W. DIETRICH, P.G. FUCHS
Institut für Klinische Virologie, Universität Erlangen-Nürnberg, D-8520 Erlangen, Federal Republic of Germany

ABSTRACT

Papillomaviruses do not replicate in tissue culture. Most viral DNAs were therefore directly cloned from biopsy material and characterized by hybridization to known reference DNAs, and by sequencing. There seems to be only one sense strand. The genome can be subdivided into a 0.4 - 1 kb regulatory region (origin of replication, enhancer and promoter elements), an early region, comprising about 4 kb, and a late region, consisting of 3 kb. At least 42 human papillomavirus (HPV) types are differentiated on the basis of less than 50% crosshybridization. They show various degrees of homology among each other, sometimes in different areas of the genome. The viral DNA usually persists extrachromosomally in tumor cells. Only with HPV16 and 18 integration was consistently observed when tumors became malignant. In transformed cells there are rather low levels of viral transcripts, although an active enhancer in front of the early transcription unit is stimulated by a trans-activating viral protein. These positive regulatory signals are counteracted by negative control elements. So far, there is no evidence that activation of cellular oncogenes is required for transformation by papillomaviruses.

INTRODUCTION

Papillomaviruses cause tumors (warts) of the skin and mucosa. Most representatives are strictly epitheliotropic and induce hyper-plasia of cells in the spinous layer. The

Aloni, Y (ed), Molecular Aspects of Papovaviruses.

virus abortively infects cells of the basal layer and seems to interfere with proper keratinocyte differentiation. As differentiation proceeds, the cells finally become fully permissive. Viral DNA replication first takes place in cells of the spinous layer and structural proteins appear in the granular and horny layer (for review, see 1).

Some papillomaviruses from animals induce fibropapillomas. Infection first leads to massive fibroplasia due to transformation of fibroblasts, which can be reproduced in vitro. The overlying epithelium does not become affected before 4 - 6 weeks (2).

A number of papillomaviruses induce tumors that may progress to carcinomas (3). Malignant conversion generally occurs on the basis of long persisting papillomas, which indicates that additional events are required. Chemical or physical carcinogens consequently increase the risk of malignant conversion and reduce the latency period. The DNA of human papillomaviruses (HPV) was detected in carcinomas of the cervix uteri, in skin carcinomas of patients with epidermodysplasia verruciformis (ev), and in cancers of the larynx (4) and the oral cavity (5).

Papillomaviruses represent the second genus of the papovaviridae family (6). They share genus-specific antigens and their DNAs crosshybridize under conditions of low stringency (1). The icosahedral capsid is 55 nm in diameter and the double-stranded circular DNA consists of about 8.0 kb. There is still no in vitro system for propagation of papillomaviruses, which considerably hindered studies on molecular biology. These are mainly based on persistently infected cell lines, obtained either by in vitro transformation of fibroblasts with bovine papillomavirus (BPV) 1 or from tumor biopsies (human cervical cancer or rabbit tumors). In analogy with polyoma viruses, those parts of the genome, which are expressed in such cell lines or in the basal, non-productive part of a wart are referred to as "early". Those genes, which code for structural proteins are called late. A genetic

analysis of BPV1 revealed the functions of some proteins encoded by open reading frames (ORF) from the early region. ORF E1 is important for extrachromosomal replication of viral DNA (7), E7 for high-copy maintenance (7), E2 for trans-activation of transcription (8, 9), and E6 and E5 for transformation of fibroblasts (9, 10, 11).

RESULTS

Cloning and classification of papillomaviruses.

Papillomaviruses cannot be propagated in tissue culture. HPV DNA is therefore directly cloned from DNA extracts from biopsy material. Viral DNA is first revealed by hybridization under relaxed conditions (T_m - 40°C, where T_m is the melting temperature of the DNA) using available HPV DNAs as probes. Restriction enzymes, which are suitable for cloning, can be identified by a couple of test cleavages. In the case of large amounts of viral DNA it will be inserted directly into pBR322. Otherwise it is recommendable to construct a library of biopsy DNA in phage Lambda and to select HPV DNA-positive clones by plaque hybridization. HPV2a, 2c, 3, 5, 8, 10, 13, 19, 20, and 25 DNAs were isolated in our laboratory (12, 13, 14, 15, A. Gassenmaier and H. Pfister, unpublished). Different isolates are compared by DNA cross-hybridization in liquid phase and are regarded as independent types in the case of less than 50% cross-reactivity (16). DNAs of some types do not reanneal at all under stringent conditions (T_m - 20°C) whereas others do cross-hybridize between 5% and 30% (Table 1).

Genome organization and patterns of relationship.

We determined the nucleotide sequence of HPV8, which induces macular skin lesions in patients with ev and appears to be associated with skin cancers linked to this syndrome. The genome consisted of 7654 base pairs (17). All major ORFs but one were located on one DNA strand, which is regarded as sense strand (Fig. 1). The region between ORFs L1 and E6, spanning 397 base pairs, showed no major reading frames. In this area there are numerous A-T rich sequences and two copies

Table 1 : Cross-hybridization between DNAs of different HPV types.

	1	2a	2c	6	11	13	8	19	25
HPV 1	100	0	0	0	0	0	0	0	0
HPV 2a		100	55	0	0	0	0	0	0
HPV 2c			100	0	0	0	0	0	0
HPV 6				100	25	4	0	0	0
HPV 11					100	3	0	0	0
HPV 13						100	0	0	0
HPV 8							100	10	29
HPV 19								100	25
HPV 25									100

Cross-hybridization was determined from the kinetics of DNA reassociation in liquid phase (12, 13, 15). The numbers represent the percentage of cross-reactivity, with 100% representing the homologous hybridization. Values of calf thymus DNA hybridization were defined as 0%.

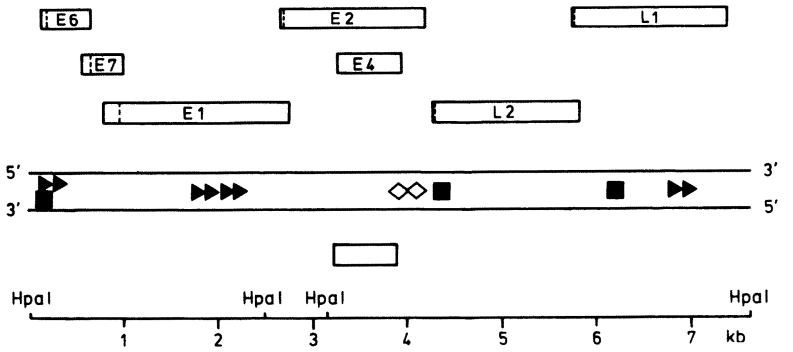


Fig. 1. Genome organization of human papillomavirus 8. Open reading frames were displayed by means of the computer program "FRAMES" (18) and are indicated by open bars. Dotted lines within the frames show the first methionine codon. The reading frames are designated E1 to E7 and L1 or L2 according to currently accepted nomenclature (19).

of the inverted repeat ACCGNNNNCGGT. Potential RNA polymerase II promotor sequences, consisting of CAAT-box and TATA-motif, are situated at the 5' end of E6. Additional putative promotor sequences appear within ORF E1 and close to the 3' end of L1 (Fig. 1). Signals for termination of transcription and polyadenylation are at the 5' end of E6 and in the 5' part of L2 and L1. A possible RNA polymerase III promotor is close to the 3' end of ORF E2.

The sequence of HPV8 was compared with those of other papillomaviruses. All sequences showed a colinear organization. As a consistent feature the 3' moiety of E1, the 5' moiety and the 3' end of E2, the 5' and 3' ends of L2, and the entire L1 appeared highly conserved. A moderate homology was noted for ORFs E6, E7, and the 5' part of L2, whereas most viruses differed considerably in ORF E4, in the 3' part of L2, and in the non-coding region. Comparing highly homologous HPV DNAs, we observed some deviations from the above scheme. HPV8 and 25, for example, seem to be closely related in the 5'-half of E1 and less homologous in the 3'-part of this ORF (J. Krubke, H. Delius, and H. Pfister, unpublished). Similarly, HPV19 and 25 cross-hybridize clearly more efficiently within the non-coding region and ORF E4 than within ORF E1.

We extended the comparison of HPV genomes to types, which were not yet sequenced, and looked for peaks of homology by Southern blot hybridization under conditions of slightly reduced stringency. HPV 2 DNA hybridized with DNA of HPV 3, 6, 10, 11, 13, 18, and 26 (Fig. 2). The genomes of HPV2 and 8 were aligned by hybridization with subgenomic DNA fragments under relaxed conditions to determine the genome organization of HPV2. According to this alignment cross-hybridization between HPV2 and other viruses was mainly confined to ORF L2 and L1. Only HPV3 and 10 hybridized in addition to sequences from ORFs E1 and E2. Fig. 3 shows an analogous analysis for HPV13, 16, and 26, hybridized to HPV18 DNA. The early region of HPV 18 revealed significant homology to HPV 26 and HPV 16 (Fig. 4). It became obvious particularly from this analysis

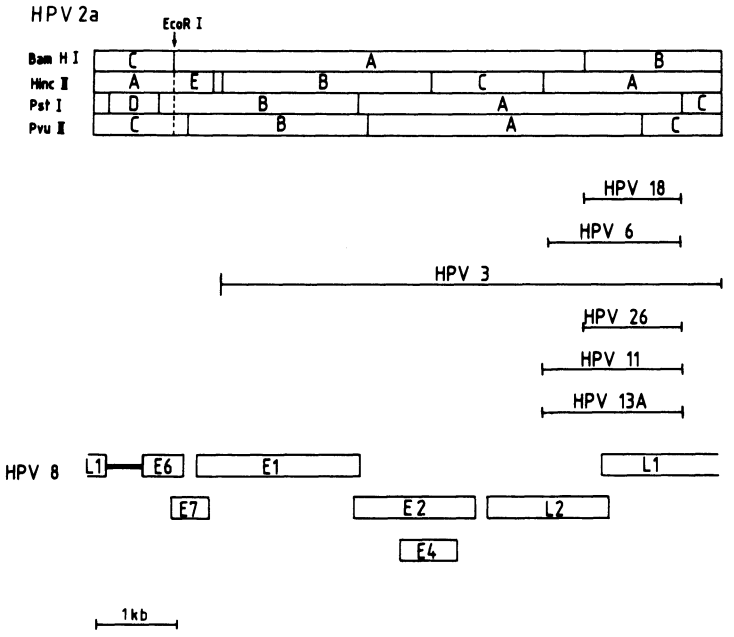


Fig. 2. Areas of cross-hybridization between HPV2a, 3, 6, 11, 13, 18, and 26. Only the BamHI A fragment hybridized in the case of HPV13. HPV2a DNA, which was separated from the vector by EcoRI cleavage was digested with BamHI, HindII, PstI, and PvuII. Southern blots with the resulting fragments were hybridized to labelled DNA of the other HPV types in 50% formamide and 5 x SSC at 37 °C. Filters were washed in 6 x SSC, 0.5 % SDS at 65 °C. The cross-hybridizing regions were narrowed down by subtracting the negative fragments of the individual digests. Aligned with HPV2a, the HPV8 genome is shown at the bottom.

that different pairs of HPV types cross-hybridize in different areas of the genome.

Persistence of viral DNA.

We examined the physical state of viral DNA in BPV1 induced hamster tumors (21), BPV1 transformed mouse fibroblasts, and in an HPV5 DNA positive squamous cell carcinoma of an ev patient (22). The viral DNA persisted extrachromosomally with a high copy number (20 - 200) per cell. No deletions or rearrangements were observed in our systems. Concatemers were detected in small amounts.

The picture is somewhat different with HPV 16 and 18 in genital tumors. In premalignant lesions the DNA of these

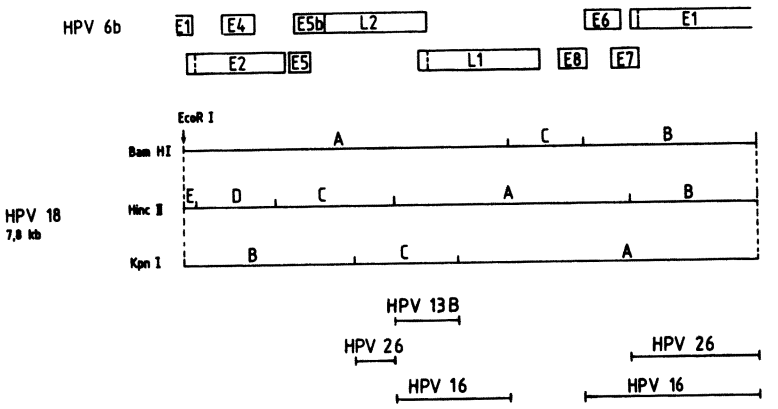


Fig. 3. Areas of cross-hybridization between HPV18, 13, 16, and 26. Only the BamHI B fragment hybridized in the case of HPV13. The physical map of HPV18 and the alignment with HPV6b (20) is shown on top. For experimental details, see legend to Fig. 2.

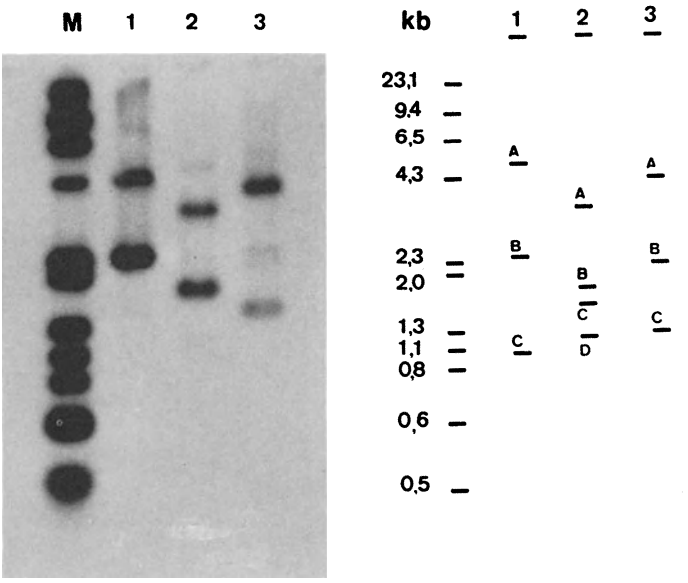


Fig. 4. Southern blot hybridization of ^{32}P-labelled HPV16 DNA with HPV18 DNA, which was separated from the vector by EcoRI cleavage and digested with BamHI (1), HincII (2), and KpnI (3). The size of the length standard DNA and the position of the HPV18 DNA fragments are given at the right. Hybridization conditions are described in the legend to Fig. 2.

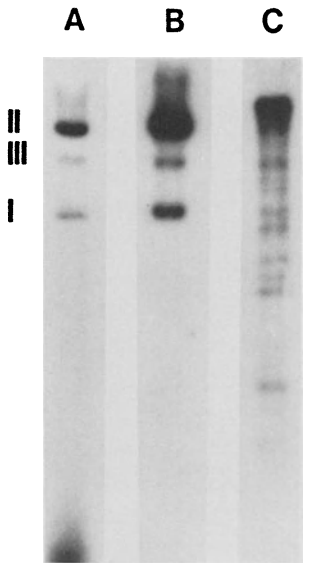


Fig. 5. Physical state of HPV18 (A) and HPV16 (B, C) DNA in genital tumors: cervical intraepithelial neoplasia (CIN) grade II (A), CIN I (B), and invasive squamous cell carcinoma (C). One half of biopsies was processed for histological examination and DNA was extracted from the second one. Southern blots were hybridized to labelled HPV18 and HPV16 DNA, respectively. The position of form I, II, and III of plasmid DNA is indicated to the left. DNA of lane C was partially digested with PstI and shows integrated, high molecular weight viral DNA.

types also persists as a plasmid but in frankly malignant tumors most of the viral DNA appeared integrated into the cellular genome (Fig.5).

Viral DNA may persist without inducing any clinical symptoms. When screening biopsies from clinically and histologically normal cervical mucosa of unselected women by DNA-hybridization we observed HPV DNA in 12 out of 29 cases. HPV 6 and 11 were detected in 4 cases, HPV 16 in 7, and HPV 18 in 2 (P.G. Fuchs, F. Girardi, and H. Pfister, unpublished). This indicates that asymptomatic infections with HPV are widespread in the general population.

Gene expression.

Virus-specific transcripts, which are all polyadenylated, could be demonstrated on Northern blots of total RNA from BPV 1 induced hamster tumors (23) and BPV 1 transformed mouse fibroblasts (Fig. 6). The RNAs are transcribed from the early part of the BPV 1 genome and appear in small amounts, which is in contrast to the usually high copy number of viral DNA (21, 23). Basically there seems to be no strict correlation between DNA and RNA levels when comparing various clones of BPV 1 transformed C127 cells and embryonic mouse or hamster fibroblasts

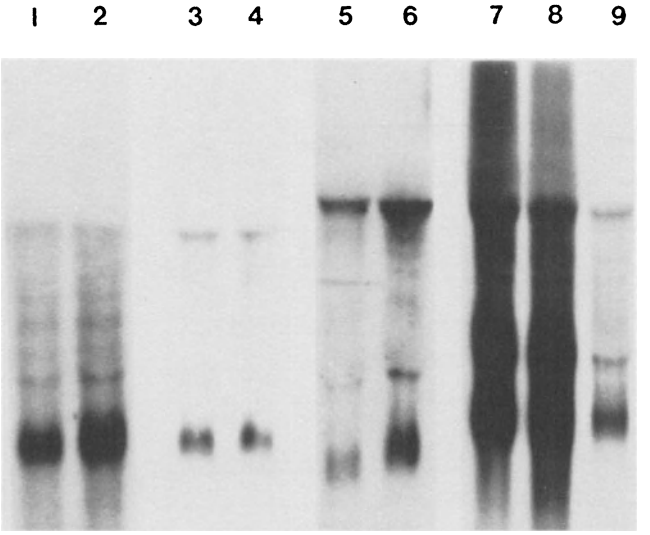


Fig. 6. Virus-specific transcripts in BPV1 transformed cell lines after treatment with different drugs. Total RNA was isolated and displayed by the Northern blot technique as described elsewhere (24). (1) C127 B81 iododeoxyuridine, 1 µg/ml, 4d (2) untreated control, (3) C127 B81, TPA, 20 ng/ml, 4d (4) untreated control, (5) embryo-fibroblasts of the DBA mouse, infected at the second passage with BPV1 virions and treated with TPA (20 ng/ml) from day 1 to 3 after infection, (6) untreated control, (7) C127 B81,puromycin, 100 µg/ml, 1 h, (8) C127 B81, cycloheximide, 25 µg/ml, 1 h, (9) untreated control.

(data not shown). The low RNA levels are partially due to a rapid degradation. The half-life of early transcripts was 35 min. in C127 cells and 65 min in DBA mouse embryo fibroblasts as determined by the decrease of RNA after blocking of transcription by actinomycin D (24). Attempts failed to increase the level of viral RNA by treatment with iododeoxyuridine (1 µg/ml) or TPA (12-O-tetradecanoylphorbol-13-acetate at a concentration of 20 ng/ml (Fig. 6). We infected embryonic fibroblasts of the DBA mouse at the second passage with BPV 1 virions and looked for an effect of TPA on transformation . The drug was added to one culture flask (20 ng/ml) one day after infection and was left for three days. Morphologically transformed cells, which were able to form colonies in soft

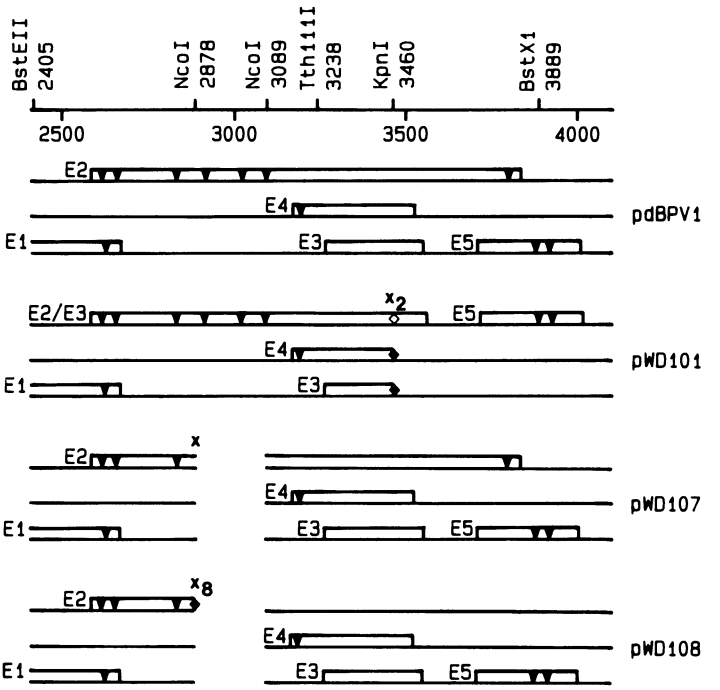


Fig. 7. BPV1 mutants affected in the 3'-part of the early region. A physical map of that area is shown on top. For the construction of pWD101, the plasmid pdBPV1 was cleaved with KpnI, treated with T4 DNA polymerase in the presence of all four dNTP's and ligated with phosphorylated Xba I linker octamers. The inserted linkers were removed except for two (x_2) by SacI cleavage and religation. Sequence analysis revealed the unexpected loss of 1 bp each to the left and right of the linker insertion site. Downstream of the linkers a stretch of 30 bases showed sequence alterations. This resulted in a frameshift fusing ORF E2 to the modified 3' end of E3. The 211 bp NcoI fragment was deleted to construct plasmids pWD107 and 108, the protruding ends of the NcoI cleavage sites were removed by S1 nuclease and one (x) or eight (x_8) XbaI linkers were inserted. In pWD107 the construction created an in frame deletion in the 5' half of E2, in pWD108, ORF E2 is terminated by the stop codon of a linker.
▭ ORF's, ▼ ATG codon, frameshift, ◇ ◆ termination codon introduced by XbaI linker.

agar, grew out with and without TPA treatment. DNA was extracted at passage 4, 6, 8, and 10. No difference was observed in the content of viral DNA of TPA treated cells and of untreated controls. When tested at passage 8, both cultures revealed similar levels of viral RNA (Fig. 6), indicating

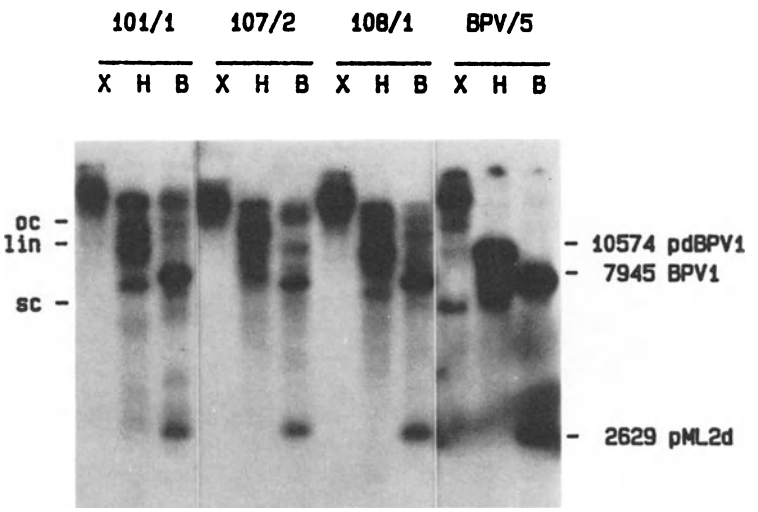


Fig. 8. Autoradiographic detection of BPV1-specific DNA in total cellular DNA from cloned foci of C127 mouse cells, which were morphologically transformed by different mutants (pWD101, 107, 108) or cloned wild-type DNA (pdBPV1). 10 µgDNA each were cleaved with XhoI (X, noncutter for plasmid DNA), HindIII (H, single cutter), or BamHI (B, separates vector and viral insert), electrophoresed on a 1 % agarose gel, transferred to nitrocellulose and hybridized to ^{32}P labelled pdBPV1 DNA. The positions of open circles (oc), supercoiled (sc) and linearized (lin) plasmid DNA are indicated. Size markers to the right correspond to linear pdBPV1-, BPV1-, and pML2d-DNA. The patterns observed are consistent with integration of mutant DNA in tandem orientation at several sites and extrachromosomal persistence of monomeric and multimeric BPV1 DNA.

that TPA has no effect on the establishment of BPV 1 in mouse fibroblasts.

In view of the low levels of viral RNA it was surprising to detect DNA sequences, which stimulate BPV1 transcription in cis and trans. An enhancer element in the non-coding region is activated by a protein encoded by ORF E2. This was first shown by the stimulation of the chloramphenicol acetyl transferase (CAT) gene expression under control of BPV1 sequences (8). We confirmed the stimulatory effect of E2 by analyzing authentic transcription of BPV1 E2 mutants. These were generated by insertion of XbaI-linkers carrying a TAG stop codon (Fig. 7). The DNAs were transfected to C127 cells and trans-

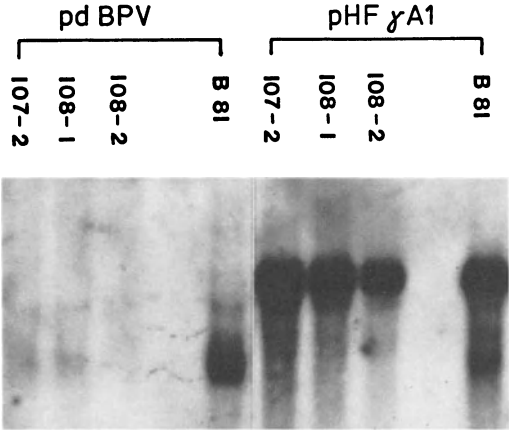


Fig. 9. Amount of BPV1 specific transcripts in C127 cells transformed by BPV1 mutants (107-2,108-1, and 108-2) or BPV1 wild-type (B81). The mutant DNAs are described in Figure 7. Total RNA was electrophoresed under denaturing conditions, transferred to nitrocellulose and hybridized with nick-translated pdBPV1 DNA. After exposure to X-ray film the nitrocellulose filter was boiled in water for 10 min to remove the radioactive DNA. It was subsequently hybridized with ^{32}P-labelled DNA, specific for γ-actin (pHF-gammaA1) to check for the amount of total RNA in each lane.

formed foci selected from soft agar. They contained integrated DNA at similar copy number as wild-type transformed cells (Fig. 8). In contrast, the amount of viral RNA in mutant transformed cells was at least tenfold reduced when compared to BPV1 wild-type (Fig. 9).

Looking for similar control elements in HPV8 we detected an enhancer element in comparable map position, which can be activated by the BPV1 E2 product (R. Seeberger and H. Pfister, in preparation). The HPV8 EcoRI- PvuII DNA fragment, covering the non-coding region, was cloned in both orientations upstream of the CAT gene controlled by the SV40 promotor. The constructs were transfected into C127 and HeLa cells and were inactive by themselves. Cotransfection of a BPV1 E2 expression vector however, led to a clear-cut stimulation of CAT activity.

The positive regulatory elements seem to be counteracted by negative control functions (24). If protein synthesis of BPV1

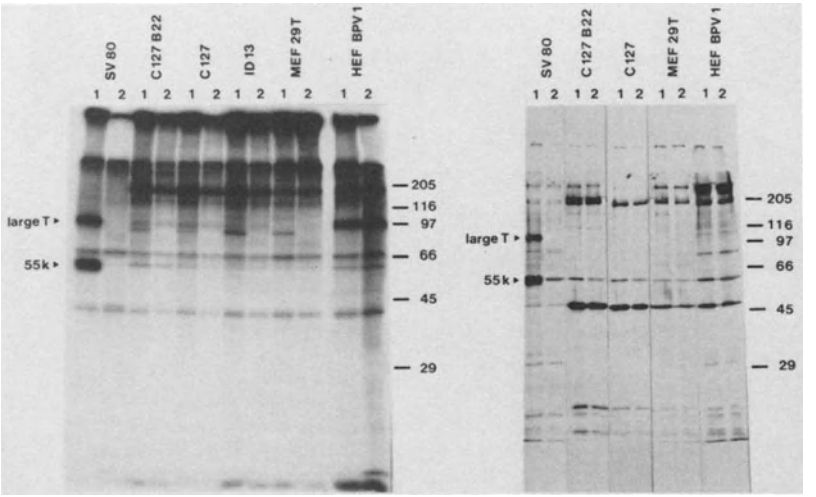


Fig. 10. Immunoprecipitation of extracts from BPV1-transformed cell lines with monoclonal antibodies directed against p53.
(A) BPV1-transformed cell lines (C127 B22, 1D13, DBA mouse embryo-fibroblasts (MEF) 29T, hamster embryo fibroblasts (HEF) BPV1), a SV40-transformed cell line (SV80), and non-transformed mouse fibroblasts (C127) were labelled for 4 h in the presence of ^{35}S-methionine. Cells were lysed in buffer (50 mM Tris (pH 7.5), 5 mM EDTA, 150 mM NaCl, 0.5% NP40), sonicated and proteins immunoprecipitated with either monoclonal antibody against p53 (1) (27) or control antibody (2). The radiolabelled products were separated by electrophoresis on 15% SDS-polyacrylamide gels and detected by autoradiography.
(B) Immunoprecipitation as in (A), after labelling of cells with ^{32}P. The positions of the cellular p53, the SV40 large T antigen, and molecular weight markers are indicated.

transformed cells was inhibited by addition of cycloheximide (25 µg/ml) for only one hour, the amount of viral transcripts increased about 10-fold (Fig. 6). Puromycin (100 µg/ml) had the same effect. Protein synthesis inhibition by puromycin leads to early release of ribosomes, which indicates that cycloheximide does not just act via RNA protection by accumulating ribosomes. Nuclear run-on experiments showed a roughly 7-fold

increase in transcriptional activity and the half-life of viral RNAs was about two-fold prolonged. These data indicate that rather labile proteins control both RNA synthesis and degradation.

The control of gene expression turned out to be not restricted to BPV1. We cloned the entire early region of HPV8 into the pSV2 neo plasmid and transfected C127 cells. HPV8-neo DNA-positive clones were expanded after G418 selection and were shown to express HPV8 specific RNA. Cycloheximide treatment for 2 h again increased the level of viral transcripts about 10-fold.

In contrast to these findings it is interesting to note that cycloheximide treatment had no effect on the amount of HPV 18 and HPV 16 transcripts in HeLa, C4-1, or SiHa cells, which are all derived from human cervical carcinomas. No changes in the level of viral RNA occurred even after 7 h cycloheximide treatment (24).

The cellular p53 protein in BPV1 transformed cells.

The cellular p53 is found at elevated levels in a wide variety of transformed cells (25) and was shown to be physically associated with the SV40 large T antigen and the adenovirus E1b - 58 k protein (26). We used monoclonal antibodies against p53 for immunoprecipitation from protein extracts from BPV1 transformed rodent cells (Fig. 10). The level of p53 was not elevated in comparison to untransformed cells except for one (not shown) out of two BPV1-transformed embryonic fibroblast lines of DBA mice. These results imply that elevation of p53 is not required for transformation by BPV1. No BPV1-specific protein could be identified in association with p53.

DISCUSSION

Papillomaviruses turned out to be very heterogeneous. Based on a rather strict type criterion 42 HPV types can be differentiated at the moment (28), many of which do not cross-hybridize under stringent conditions. Genes involved in replication (7), transcription control (8), and capsid forma-

tion (29) appear most conserved. HPV-subgenera can be distinguished based on DNA relationship (Table 1) and the members of one subgenus usually share common biological and pathogenetic properties (28). Some viruses such as HPV18 show limited homology with representatives from several groups within different genome regions (Fig. 3). In some cases this may help to correlate the genotype with biological characteristics. It was for example interesting to note homology between HPV18 and 16 at the 5' end of the early region, which codes for the potentially transforming gene E6 (10), because both types are frequently associated with cervical cancer (20, 30).

In spite of nucleotide sequence heterogeneity the general genome organization of papillomaviruses is rather uniform and differs from that of other members of the papovavirus family (3, 19). As a characteristic feature all major ORFs are located on one DNA strand. This is in line with all RNA species being transcribed from the same DNA strand (31). The early transcription unit is mainly defined by a promotor in front of ORF E6 and a polyadenylation signal at the end of ORF E2 (Fig.1). A variety of m-RNAs is generated by differential splicing and additional promotors within the early region also seem to be active (for review see 3). The most abundant RNA in BPV1-transformed cells is theoretically able to code for an E1/E4 fusion protein and for an E5 protein (32, 33). The transcription of the early region is stimulated by an enhancer in the non-coding region, which is in turn activated by the viral E2 gene product (8, 24). A functional E2 achieves a roughly ten-fold increase in RNA levels in transformed cells. Its effect is balanced by a negative control leading to a decrease by a factor of ten (24). After release from negative control by cycloheximide treatment, E2 mutants of BPV1 show similar RNA levels as BPV1 wild-type.

The switch to late gene expression, which depends on keratinocyte differentiation, is not yet understood. In

the case of BPV1 the transcripts appear under control of a promotor sequence upstream of the early promotor, which is highly reminiscent of the SV40 late promotor (GGTACACATCC vs GGTACCTAACC) (C.C. Baker and P.M. Howley, personal communication).

A latent infection, where no viral transcripts were detectable in spite of a high DNA copy number, was described for a papillomavirus of the multimammate rat (34). A similar state seems possible in obviously numerous cases where HPV types persist without inducing any clinical symptoms (35, 36, this paper).

In view of the clinical relevance of HPV, the transforming functions and the carcinogenic progression associated with papillomavirus infection are of special interest. Two genes of BPV1, namely E6 and E5 are independently able to induce morphologic transformation of fibroblasts (9, 10, 11). It was already pointed out, however, that all HPVs are strictly epitheliotropic and it is not for sure that keratinocytes respond to E6 and E5 functions in the same way as fibroblasts. As a matter of fact there is no ORF E5 in HPV8 (17) and in HPV5 (K. Zachow and R. Ostrow, personal communication). On the other hand, ORF E6 appears moderately conserved among all papillomaviruses sequenced so far. Most cysteine residues are arranged into the repetitive motif C-X-X-C, which appears at least four times with exactly conserved spacing, suggesting a rather conserved tertiary structure. An E6 specific protein was recently identified in BPV1 transformed cells (37). It has a relative molecular mass of 15.5 kD and is almost equally distributed between nucleus and non-nuclear membranes. Viral transcripts, which could code for an E6 protein, were detected in HPV18 DNA containing cervical cancer cell lines (38) and in papillomavirus induced skin cancers of rabbits (39). This may indicate that the E6 gene also plays a role in oncogenesis by epitheliotropic papillomaviruses.

The analysis of HPV16- and 18-associated tumors suggests that integration of viral DNA into the cellular genome may

have an impact on carcinogenic progression (20, 30). As with all other papillomaviruses the DNA of these types persists extrachromosomally in benign lesions. The replication of BPV1 was shown to depend on functions encoded by ORFs E1 and E7 (7). In malignant tumors, however, HPV16 and 18 DNA occurs. partially or completely integrated. The integration site into cellular DNA seems to be random but the circular viral DNA preferentially opens within the 3' part of the early region thus disrupting the early transcription unit (20, 30, 38, 40). The viral-cellular fusion transcripts, which arise from the early viral promotor, could gain an altered stability due to the exchange of the 3' end, which would in turn affect the level of potentially transforming proteins. The half-life of HPV18 RNA in HeLa cells indeed turned out to be about 5-times as long as that of BPV1 RNA in mouse fibroblasts (24). In other malignancies HPV-DNA integration is apparently not essential, however. The DNAs of HPV5, 8, 10, 11, or 33 stay extrachromosomal within the limits of test sensitivity (3, 41, 42, 43). This indicates that integration represents no general mechanism of carcinogenic progression of HPV-associated tumors.

Events, which have to occur during the long latency periods between primary infection and malignant conversion may aim at cellular genes, which can trigger oncogenic activities. We observed no consistent increase in the level of p53 in BPV1-transformed cells nor was there an increase in the amount of the src gene product in BPV1-transformed hamster cells (44). The specific phosphotransferase activity of the $pp60^{c\text{-}src}$ was increased in some transformed cell clones whereas no elevation was observed in others. This excludes that an influence on $pp60^{c\text{-}src}$ is necessary for transformation. No association between viral proteins and p53 or $pp60^{c\text{-}src}$ could be demonstrated.

The copy number of the c-myc and/or c-ras oncogenes is elevated in advanced cervical cancers (45) but appears normal in early malignancies, indicating that these oncogenes may become activated during tumor progression but are not likely

to trigger early events.

Some information on HPV properties, which are relevant for malignant conversion, may be obtained from a detailed comparison of viruses, which are closely related but differ in their association with malignant tumors. Such candidates are HPV16 and 31 or HPV5, 8, and 25 (13, 46). At the moment a comparison can be based on sequence analysis and on *in vitro* tests for replication, gene expression, and transformation. Only the cottontail rabbit papillomavirus system allows a direct analysis of the effect of mutations on tumor progression.

ACKNOWLEDGEMENTS

We are indebted to Drs. P.M. Howley, R. Ostrow, and H. zur Hausen for providing pdBPV1, HPV26, 16 and 18 DNA. This work was supported by the Deutsche Forschungsgemeinschaft (Pf 123/3-1) and by the Wilhelm-Sander Stiftung (83.016.1).

REFERENCES

1. Pfister, H. Rev. Physiol. Biochem. Pharmacol. 99: 111-181, 1984.
2. Olson, C., Gordon, D.E., Robl, M.G. and Lee, K.P. Arch. Environ. Health 19: 827-837, 1969.
3. Evered, D. and Clark, S. (eds) Papillomaviruses, Ciba Foundation Symp. 120. J. Wiley & Sons, Chichester, 1986.
4. Kahn, T., Schwarz, E. and zur Hausen, H. Int. J. Cancer 37: 61-65, 1986.
5. De Villiers, E.-M., Weidauer, H., Otto, H. and zur Hausen, H. Int. J. Cancer 36: 575-578, 1985.
6. Matthews, R.E.F. Intervirology 17: 1-199, 1982.
7. Lusky, M. and Botchan, M.R. J.Virol. 53: 955-965, 1985.
8. Spalholz, B.A., Yang, Y.-C. and Howley, P.M. Cell 42: 183-191, 1985.
9. Yang, Y.-C., Spalholz, B.A., Rabson, M.S. and Howley, P.M. Nature 318: 575-577, 1985.
10. Schiller, J.T., Vass, W.C. and Lowy, D.R. Proc. Natl. Acad. Sci. USA 81: 7880-7884, 1984.
11. Schiller, J.T., Vass, W.C., Vousden, K.H. and Lowy, D.R. J. Virol. 57: 1-6, 1986.
12. Fuchs, P.G. and Pfister, H. Intervirology 22: 177-180, 1984.
13. Gassenmaier, A., Lammel, M. and Pfister, H. J. Virol. 52: 1019-1023, 1984.

14. Pfister, H., Nürnberger, F., Gissmann, L. and zur Hausen, H. Int. J. Cancer 27: 645-650, 1981.
15. Pfister, H., Hettich, I., Runne, U., Gissmann, L. and Chilf, G.N. J. Virol. 47: 363-366, 1983.
16. Coggin, J.R. and zur Hausen, H. Cancer Res. 39: 545-546, 1979.
17. Fuchs, P.G., Iftner, Th., Weninger, J. and Pfister, H. J. Virol. 58: 626-634, 1986.
18. Devereux, J., Haeberli, P. and Smithies, O. Nucleic Acid Res. 12: 387-395, 1983.
19. Danos, O., Engel, L.W., Chen, E.Y., Yanif, M. and Howley, P.M. J. Virol. 46: 557-566, 1983.
20. Boshart, M., Gissmann, L., Ikenberg, H., Kleinheinz, A., Scheurlen, W. and zur Hausen, H. EMBO J. 3: 1151-1157, 1984.
21. Pfister, H., Fink, B. and Thomas, C. Virology 115: 414-418, 1981.
22. Pfister, H., Gassenmaier, A., Nürnberger, F. and Stüttgen, G. Cancer Res. 43: 1436-1441, 1983.
23. Freese, U.K., Schulte, P. and Pfister, H. Virology 117: 257-261, 1982.
24. Kleiner, E., Dietrich, W. and Pfister, H. EMBO J., in press.
25. Klein, G. (ed) Advances in Viral Oncology, Vol. 2, Raven Press, New York, 1982.
26. Rotter, V. and Wolf, D. Adv. Cancer Res. 43: 113-141, 1985.
27. Gurney, E., Harrison, R. and Fenno, J.J.Virol. 34: 752-763, 1980.
28. Pfister, H. Zbl. Haut 152: 193-202, 1986.
29. Pilancinski, W.P., Glassman, D.L., Krzyzek, R.A., Sadowski, P.L. and Robbins, A.K. Biotechnology 1: 356-360, 1984.
30. Dürst, M., Kleinheinz, A., Hotz, M. and Gissmann, L. J. gen. Virol. 66: 1515-1522, 1985.
31. Engel, L.W., Heilman, C.A. and Howley, P.M. J.Virol. 47: 516-528, 1983.
32. Stenlund, A., Zabielski, J., Ahola, H., Moreno-Lopez, J. and Pettersson, U. J. Mol. Biol. 182: 541-554, 1985.
33. Yang, Y.C., Okayama, H. and Howley, P.M. Proc. Natl. Acad. Sci. USA 82: 1030-1034, 1985.
34. Amtmann, E., Volm, M. and Wayss, K. Nature 308: 291-292, 1984.
35. Ferenczy, A., Mitao, M., Nagai, N., Siverstein, S.J. and Crum, C.P. N.Engl.J.Med. 313: 784-788, 1985.
36. Steinberg, B.M., Topp, W.C., Schneider, P.S. and Abramson, A.L. N.Engl.J.Med. 308: 1261-1264, 1983.
37. Androphy, E.J., Schiller, J.T., Lowy, D.R. Science 230: 442-445, 1985.
38. Schwarz, E., Freese, U.K., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. and zur Hausen, H. Nature 314: 111-114, 1985.
39. Danos, O., Georges, E., Orth, G. and Yaniv, M. J.Virol. 53: 735-741, 1985.

40. Lehn, H., Krieg, P. and Sauer, G. Proc.Natl.Acad.Sci. USA 82: 5540-5544, 1985.
41. Green, M., Brackmann, K.H., Sanders, P.R., Loewenstein, P.M., Freel, J.H., Eisinger, M. and Switlyk, S.A. Proc.Natl.Acad.Sci. USA 79: 4437-4441, 1982.
42. Lancaster, W.D., Kurman, R.J., Sanz, L.E., Perry, S., Jenson, A.B. Intervirology 20: 202-212, 1983.
43. Beaudenon, S., Kremsdorf, D., Croissant, O., Jablonska, S., Wain-Hobson, S. and Orth, G. Nature 321: 246-249, 1986.
44. Amini, S., Lewis, A.M., Israel, M.A., Butel, J.S. and Bolen, J.B. J. Virol. 57: 357-361, 1986.
45. Riou, G., Barrois, M., Tordjman, I., Dutronquay, V. and Orth, G. C.R.Acad.Sc.Paris 299: 575-580, 1984.
46. Lorincz, A.T., Lancaster, W.D., Temple, G.F. J. Virol 58: 225-229, 1986.

INDEX