VIRAL MESSENGER RNA

DEVELOPMENTS IN MOLECULAR VIROLOGY

Yechiel Becker, Series Editor
Julia Hadar, Managing Editor

Other books in the series:
Becker, Y. (ed.) *Herpesvirus DNA* (1981)
Becker, Y. (ed.) *Replication of Viral and Cellular Genomes* (1983)
Kohn, A. and Fuchs, P. (eds.) *Mechanisms of Viral Pathogenesis: From Gene to Pathogen* (1984)
Becker, Y. (ed.) *Antiviral Drugs and Interferon: The Molecular Basis of their Activity* (1984)
Becker, Y. (ed.) *Recombinant DNA Research and Viruses: Cloning and Expression of Viral Genes* (1985)
Feitelson, M. *Molecular Components of Hepatitis B Virus* (1985)

VIRAL MESSENGER RNA

Transcription, Processing, Splicing and Molecular Structure

edited by

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

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

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

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

The figure on the cover is reprinted with permission from the Journal of Molecular Biology *(Volume 166, page 497, 1983) and from the* Cold Spring Harbor Symposia on Quantitative Biology *(Volume 47, page 1002, 1983). This figure is also reproduced on page 56 of this book.*

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

Viral messenger RNA.

(Developments in molecular virology)
Includes bibliographies and index.
1. Messenger ribonucleic acid. 2. Genetic transcription. 3. Viruses, RNA. 4. Viral genetics. I. Becker, Yechiel. II. Title: Viral messenger RNA. III. Series.
[DNLM: 1. RNA, Messenger—genetics. 2. RNA, Viral—genetics. 3. Transcription, Genetic. W1 DE998DG / QW 168 V813]
QR458.V57 1985 576'.64 *84-27206*

ISBN-13: 978-1-4612-9621-8 e-ISBN-13: 978-1-4613-2585-7
DOI: 10.1007/978-1-4613-2585-7

CONTENTS

CONTRIBUTORS

Amiya K. Banerjee
Roche Institute of Molecular Biology
Roche Research Center
Nutley, New Jersey 07110, USA

Ann Beaton
Molecular Biology and Virology Program
Graduate School of Medical Sciences
Memorial Sloan-Kettering Cancer Center
New York, New York 10021, USA

Janet Braam
Molecular Biology and Virology Program
Graduate School of Medical Sciences
Memorial Sloan-Kettering Cancer Center
New York, New York 10021, USA

E. Buetti
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Robert H. Costa
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

Jean-Luc Darlix
Department of Molecular Biology
University of Geneva
30, quai Ernest-Ansermet
1211 Geneva 4, SWITZERLAND

Bishnu P. De
Roche Institute of Molecular Biology
Roche Research Center
Nutley, New Jersey 07110, USA

B. Gayathri Devi
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

H. Diggelmann
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Kenneth G. Draper
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

N. Fasel
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

J. Firzlaff
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Raymond J. Frink
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

Linda M. Hall
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

Alik Honigman
Department of Molecular Genetics
The Hebrew University-Hadassah Medical School
P.O.B. 1172
91 010 Jerusalem, ISRAEL

Hana C. Hyman
Department of Molecular Genetics
The Hebrew University-Hadassah Medical School
P.O.B. 1172
91 010 Jerusalem, ISRAEL

A. Jalanko
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
00380 Helsinki 38, FINLAND

W. K. Joklik
Department of Microbiology and Immunology
Duke University Medical Center
Durham, North Carolina 27710, USA

Raymond Kaempfer
Department of Molecular Virology
The Hebrew University-Hadassah Medical School
P.O.B. 1172
91 010 Jerusalem, ISRAEL

N. Kalkkinen
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
00380 Helsinki 38, FINLAND

Michael Katze
Molecular Biology and Virology Program
Graduate School of Medical Sciences
Memorial Sloan-Kettering Cancer Center
New York, New York 10021, USA

Daniel Kolakofsky
Department of Microbiology
University of Geneva Medical School
64, Ave. de la Roseraie
1205 Geneva, SWITZERLAND

Robert M. Krug
Molecular Biology and Virology Program
Graduate School of Medical Sciences
Memorial Sloan-Kettering Cancer Center
New York, New York 10021, USA

Esa Kuismanen
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
SF-00380 Helsinki 38, FINLAND

Gerald Lancz
Department of Medical Microbiology and Immunology
The University of South Florida
Tampa, Florida 33612, USA

Jose L. La Torre
Centro de Virologia Animal
Serrano 661 - 1414 Capital, ARGENTINA

E. C. Mariman
Department of Biochemistry
University of Nijmegen
Geert Grooteplein N21
6525 EZ Nijmegen, THE NETHERLANDS

Claude Meric
Department of Molecular Biology
University of Geneva
30, quai Ernest-Ansermet
1211 Geneva 4, SWITZERLAND

Joseph R. Nevins
The Rockefeller University
New York, New York 10021, USA

Meihan Nonoyama
Showa University Research Institute for Biomedicine in Florida
Clearwater, Florida 33520, USA

Amos B. Oppenheim
Department of Molecular Genetics
The Hebrew University-Hadassah Medical School
P.O.B. 1172
91 010 Jerusalem, ISRAEL

D. Owen
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Jean L. Patterson
Dept. of Microbiology
University of Geneva Medical School
64, Ave. de la Roseraie
1205 Geneva, SWITZERLAND

Present address:
Dept. of Microbiology
and Molecular Genetics
Harvard Medical School
Boston, MA 02115, USA

K. Pearson
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Ralf F. Pettersson
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
SF-00380 Helsinki 38, FINLAND

Marcia K. Rice
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

Ragna Ronnholm
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
SF-00380 Helsinki 38, FINLAND

Jeff Sample
Department of Medical Microbiology and Immunology
The University of South Florida
Tampa, Florida 33612, USA

Angeles Sanchez
Roche Institute of Molecular Biology
Roche Research Center
Nutley, New Jersey 07110, USA

Eduardo A. Scodeller
Centro de Virologia Animal
Serrano 661 - 1414 Capital, ARGENTINA

James R. Smiley
Pathology Department
McMaster University
Hamilton, Ontario, CANADA

H. Soderlund
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
00380 Helsinki 38, FINLAND

Pierre-Francois Spahr
Department of Molecular Biology
University of Geneva
30, quai Ernest-Ansermet
1211 Geneva 4, SWITZERLAND

William Steinhart
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

F. William Studier
Biology Department
Brookhaven National Laboratory
Upton, New York 11973, USA

K. Takkinen
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
00380 Helsinki 38, FINLAND

Ismo Ulmanen
Recombinant DNA Laboratory
University of Helsinki
Valimotie 7
SF-00380 Helsinki 38, FINLAND

W. J. van Venrooij
Department of Biochemistry
University of Nijmegen
Geert Grooteplein N21
6525 EZ Nijmegen, THE NETHERLANDS

R. Verheijen
Department of Biochemistry
University of Nijmegen
Geert Grooteplein N21
6525 EZ Nijmegen, THE NETHERLANDS

A. L. Vessaz
Swiss Institute for Experimental Cancer Research
CH-1066 Epalinges, SWITZERLAND

Edward K. Wagner
Department of Molecular Biology and Biochemistry
University of California
Irvine, California 92717, USA

PREFACE

The nucleotide sequence of the gene from which messenger RNA molecules are transcribed is in a form that can be translated by cellular ribosomes into the amino acid sequence of a particular polypeptide, the product of the gene. The discovery of messenger RNA more than twenty years ago led to a series of studies on its organization and function in cells in the presence of infecting viruses.

This volume is devoted to current studies in the field of cellular and viral messenger RNA. The studies presented provide an insight into molecular and genetic aspects of messenger RNA. Special attention was paid by the authors to the molecular organization of mRNA species, to the processing of mRNA molecules, and to the different strategies employed by DNA and RNA viruses in the synthesis of their mRNA. The ability of a virus to take over the protein-synthesizing mechanisms of an infected cell depends on its ability to produce mRNA molecules which can affect the host mRNA or utilize cellular components more efficiently. The differences between, and similarities of, the strategies of mRNA synthesis devised by various DNA and RNA viruses are described herein.

This book should be of interest to all students of cellular and viral genes and scientists in the field. It is suitable as a textbook for workshops and courses on mRNA.

I wish to thank the authors for their fine contributions and for their interest.

Yechiel Becker
Jerusalem

VIRAL MESSENGER RNA

General

1

BIOGENESIS OF MESSENGER RNA

JOSEPH R. NEVINS

The expression of genes in a eukaryotic cell, either cellular or viral, involves a number of biochemical events that culminate with the appearance in the cytoplasm of a functional mRNA engaging ribosomes. The definition of this pathway of mRNA formation has been largely performed through the study of viral systems, adenovirus, SV40 and polyoma. In this chapter I will discuss our experiments that have contributed to an understanding of the complexity of eukaryotic mRNA biogenesis.

1. INTRODUCTION

The adenoviruses are double-stranded DNA-containing viruses that replicate in the nucleus of infected cells (1). Because of these characteristics as well as the fact that they bring no enzymes into the cell, this group of viruses provide an excellent model system for the study of the formation of a eukaryotic mRNA. That is, those aspects that characterize the formation of adenovirus RNA likely are a reflection of the events producing a eukaryotic messenger RNA. Due to the difficulties in measuring early events of mRNA biosynthesis, the viral system offers clear advantages over the study of single copy cellular genes. Many of the primary discoveries, such as definition of transcription units encoding mRNA, the precise localization of transcription initiation sites, temporal order of RNA processing events, mechanism of formation of poly(A) addition sites and RNA splicing were first documented with adenovirus. The purpose of this chapter is to present an overview of the process of mRNA biogenesis - that is, a detailing of the steps involved in producing a functional cytoplasmic mRNA. Other chapters will deal with certain specific aspects of this process, and thus in greater detail. Most of what will be discussed here stems from our work employing adenovirus as a system of study. What should be emphasized, however, is the fact that in every case thus far examined, the

Y. Becker (ed.), VIRAL MESSENGER RNA.

details orginating from adenovirus experiments have been the rule for cellular genes as well.

2. MATERIALS AND METHODS

All procedures have been described in previous publications (2-5).

3. RESULTS AND DISCUSSION

The general pathway for the generation of a eukaryotic mRNA is depicted in Figure 1. Over the past few years experiments performed in our laboratory have been focused on many of these events resulting in the elucidation of several details of mRNA biogenesis. Thus the following discussion presents our experiments which have contributed to the knowledge of the details of this pathway.

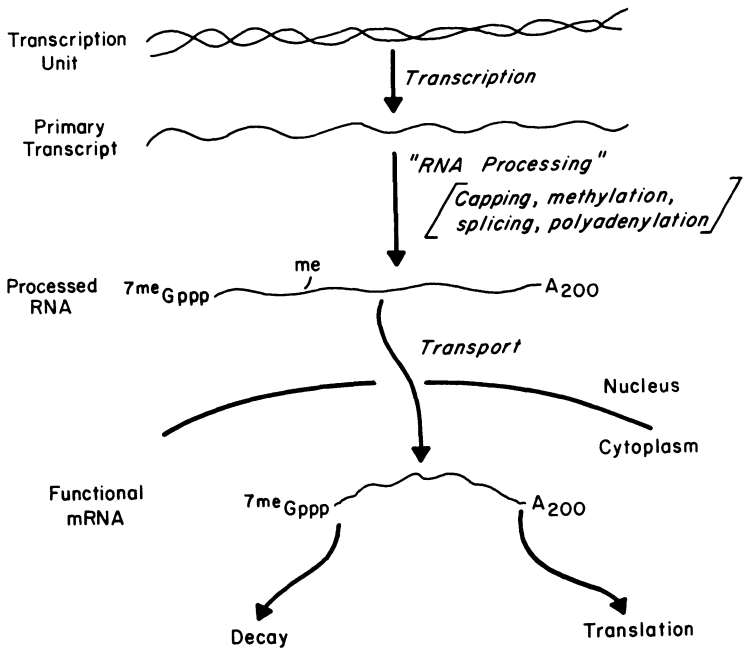

FIGURE 1. Schematic diagram of the pathway of formation of a eukaryotic mRNA.

3.1. Transcription

The primary level at which expression of a given gene may be controlled is at the transcriptional event; thus the study of the mechanisms of

transcription initiation and the possible controlling factors is of utmost importance. The details of transcription, including initiation, termination, and the definition of individual transcription units have been elucidated in some considerable detail now (see 6-10 for reviews).

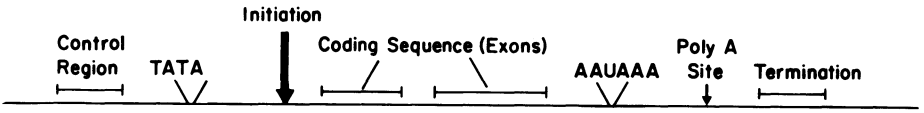

FIGURE 2. Schematic diagram of the components of a typical transcription unit.

Transcription initiates at the sequence encoding the mRNA 5' terminus, the so-called "cap site", as first shown for the adenovirus major late transcription unit (11). There is no evidence for initiation upstream from this site and therefore there appears to be no RNA processing to generate the mature 5' end of the mRNA. This is also entirely consistent with the results of in vitro transcription that yield initiations only at the cap site (12-14). Sequences that are necessary for transcription have now been mapped for a variety of genes through the in vivo and in vitro assay of genes that have been altered by deletion and site-directed mutagenesis procedures. In addition to the highly conserved TATA sequence that appears to be essential for accurate transcription initiation, there is a sequence, referred to in Figure 2 as the upstream control region, that is essential for the full transcriptional activity of the gene (see 7,8 for reviews). Our recent experiments have focused on the sequences that are essential for transcription in one of the early viral transcription units, the E2A gene. E2A promoter deletion mutants were constructed using plasmids containing the intact transcription unit and 288 nucleotides of sequence 5' to the transcription initiation site. Assay of these plasmids by transient transfection in human 293 cells, adenovirus-transformed human embryonic kidney cells that constitutively express the E1A function, gave the result that 79 nucleotides of upstream sequence are sufficient for full E2 expression (Imperiale & Nevins, submitted). Furthermore, full activity could be restored to a -28 mutant (a non-functional plasmid retaining 28 nucleotides of upstream sequence) by the insertion of sequences from -21 to -262 of the wild type

promoter. These results would suggest that the essential E2 promoter element maps between -21 and -79 relative to the transcriptional initiation site. In addition, it was found that this upstream control region could function in a reverse orientation as well as at a large distance from its normal position, characteristics that define the transcriptional enhancer elements (15).

3.1.1. Temporal control of early adenovirus transcription. The activity from the six early viral transcriptional units was determined during the course of an infection (16). The results indicated that there was differential activation such that E1A and E4 were activated prior to E2 and that there was negative control of the activity of the E4 and E2 transcriptional units. Subsequent experiments indicated that the negative control of the E4 transcription unit was mediated by the product of the E2 gene, the 72kd DNA-binding protein (17). Recent experiments using a cell-free transcription system have demonstrated that the purified 72kd protein can specifically inhibit transcription from the E4 promoter (18). Yet to be determined, is the mechanism of inhibition by this protein. Certainly the simplest model would be one in which the protein recognized a specific sequence at the E4 promoter and by binding to this sequence prevented the entry of RNA polymerase in a manner analogous to the inhibition of early SV40 transcription by T antigen (19). However, in view of the lack of evidence for binding of the protein to double-stranded DNA (20), it is also possible that the protein inhibits the action of the RNA polymerase directly, possibly through a transcriptional factor that is used specifically for E4 transcription.

3.1.2. Control of initiation. The specific activation of certain genes at a given time is of fundamental importance to cell differentiation, hormonal responses, as well as cancer. The study of the activation of the early adenovirus genes has provided some insight into the process of positive control of transcription initiation. Through the use of viral mutants, it was previously demonstrated that the early transcription units were activated as a result of the action of the product of the E1A gene (21,22). It was subsequently shown that the E1A action was at the level of transcription (3) although there also appears to be a post-transcriptional role for E1A as well (23).

The mechanism for E1A action appears to be indirect, involving a regulatory component of the host cell. Such a cellular component appears to

be a negative-acting factor as it was found that the requirement for E1A function could be overcome by inhibiting cellular protein synthesis (3,23) leading to the suggestion that the function of the E1A protein was to negate the action of a negatively-acting cellular regulatory factor. Further insight has been provided by the finding that the E1A gene product also induces at least one cellular gene, the gene encoding the major 70kd heat shock protein (4,24). These results suggest a common mechanism of control of the heat shock gene and the early adenovirus genes, a conclusion that has been supported by our recent experiments that suggest the presence of an E1A-like activity in certain cell lines (Imperiale et al., submitted). That is, various human cell lines, in particular those derived from human tumors, have the capacity to support early viral transcription in the absence of E1A, albeit at a level well below that which is found in the presence of E1A. Of primary interest is the fact that in those cell lines that support early viral transcription there is also a measureable level of transcription of the heat shock gene. Thus, there does in fact appear to be common control of these genes on a cellular basis.

Further evidence suggesting an indirect role for the E1A protein in gene activation has been provided by the finding that the herpesvirus immediate-early gene product can effectively activate early adenovirus transcription, acting in place of the E1A gene (5,25). Most importantly, the immediate-early mediated activation appears to be more efficient than E1A-mediated activation. This finding argues very strongly against a mechanism involving sequence-specific recognition by the protein given the fact that a heterologous factor is as efficient or more so than the homologous factor.

3.1.3. Termination control. The output of a transcription unit can also be affected by a variation in the termination of transcription. The expression of distal genes in a complex transcription unit can be prevented by transcriptional termination. Such is the case for the late adenovirus transcription unit. During the late stage of viral infection, transcription initiating at the late promoter proceeds to the end of the genome (26). However, during early infection transcription terminates in the middle of the genome (27-29); as a result, the L4 and L5 regions are not transcribed and thus not expressed (see Figure 3).

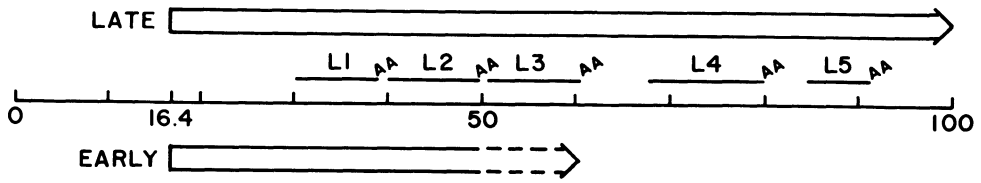

FIGURE 3. Transcriptional termination from the late adenovirus promoter during early and late infection.

The mechanism for such a change in termination is not clear at the moment. One possibility could involve an influence from transcription deriving from the E2 promoter which generates transcripts from the opposite strand in the vicinity of the region where termination occurs from the late transcription unit. Perhaps this opposing transcriptional activity prevents polymerase movement; this would then be relieved late in infection when E2 transcription is greatly reduced. Alternatively, there could be a strong termination site that is utilized early but is subject to some form of anti-termination late in infection.

Such a phenomenon has clear implications for the control of complex cellular transcriptional units, including that encoding the immunoglobulin heavy chain genes (30-32). That is, any transcription unit that contains multiple coding segments [i.e. multiple poly(A) sites] can potentially be regulated by transcriptional termination. Of course, an equally possible mechanism for control would be an alteration in poly(A) site selection as will be discussed in more detail in a later section (sect. 3.2.1).

In actuality, such control need not be confined to complex transcriptional units. Termination prior to reaching a poly(A) addition site in a simple transcription unit can certainly prevent the expression of the gene. In this case, the control would be an "on-off" regulation producing an effect equivalent to control of transcription initiation. This phenomenon does indeed occur during expression of the late adenovirus transcription unit (33,34). Late in infection, as many as 80% of the initiated transcripts terminate prematurely, generating transcripts of lengths of about 500

nucleotides. Whether this is a regulatory event, possibly mediated by a viral gene product, or just the inefficiency of the system, is not yet clear.

3.2. RNA Processing

Much of the complexity of formation of a eukaryotic mRNA resides in the fact that the primary transcript is not the mRNA but instead must first undergo a variety of "processing" steps before emerging from the nucleus as a functional RNA. Many of the details of these steps were first established with adenovirus. Most notably, the discovery of spliced mRNAs(35,36) was made with the late RNAs of adenovirus as was the finding that the generation of a poly(A) addition site involved an RNA chain cleavage (37).

3.2.1. Poly(A) addition. With the exception of the histone mRNAs, probably all eukaryotic mRNAs possess, at least initially, a poly(A) segment at the 3' terminus that is added as a post-transcriptional event. The details of the structure of poly(A), its mode of formation and its presence on a variety of mRNAs have been well documented in other reviews (38-40) and will not be dealt with here. Our experiments have focused on the function of the poly(A) segment and the mechanism of addition of poly(A) to nuclear RNA precursors. Experiments initially performed using adenovirus established that the formation of the RNA 3' terminus that serves as the poly(A) substrate occurred by an endonucleolytic cleavage; RNA polymerase does not terminate at the sequence encoding the mRNA 3' terminus but rather at some point downstream (37). This result has now been extended to several other transcriptional units (41,42) including at least one cellular gene, that encoding the mouse beta-globin mRNA (43).

The formation of a poly(A) site in the nuclear RNA precursor and subsequent polyadenylation is a very early event in mRNA biogenesis. A kinetic analysis of pulse-labeled nuclear RNA as shown in Figure 4 yielded the result that poly(A) addition preceded splicing, such that the initial polyadenylated nuclear RNA was an unspliced molecule (37,44). It should be pointed out, however, that such an analysis would not reveal if small splices were made. This in fact does appear to be the case as the splicing of the first leader segment to the second segment likely occurs while the transcript is still nascent (44a). Nevertheless, splicing largely takes place on polyadenylated molecules and poly(A) addition has been found to occur very early (44b). In fact, poly(A) addition at one of the five possible sites in the late adenovirus transcription unit occurs during

transcription prior to the termination by RNA polymerase (37).

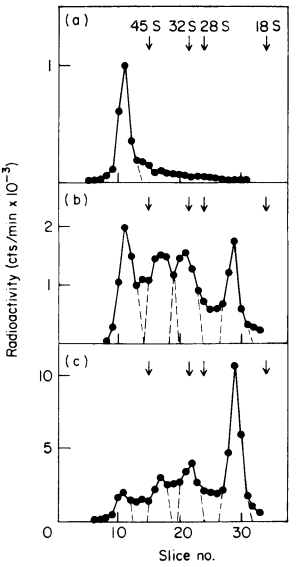

FIGURE 4. Kinetics of processing of poly(A)+ nuclear RNA that is the precursor to the fiber mRNA. Nuclear poly(A)+ RNA labeled for 5 min (a), 25 min (b) or 80 min (c) was prepared from late adenovirus infected cells, fractionated by gel electrophoresis and assayed for L5 sequences (fiber mRNA) by hybridization. Reproduced from (44) with permission.

Given this fact, it would appear that the selection of a given poly(A) site amongst several possibilities is an active process and not random. Otherwise, one might expect a polar effect in which the first site was favored and such is not the case. Obviously such a system provides the opportunity for regulation to exist whereby different poly(A) sites could be used in response to different circumstances. This actually does seem to be the case depending on when the late adenovirus transcription unit is expressed, either early or late in infection. Late in infection the L1 poly(A) site is utilized about half as frequently as the L2 site and at about 30% of the frequency of the L3 site (37). Early in infection, however, the L1 site is used at least twice as frequently as the L2 or L3 sites (27). Such a mechanism would also have profound implications for the expression of the immunoglobulin heavy chain genes. For instance,

the mu heavy chain transcription unit could be constant with respect to transcriptional termination, regardless of the state of differentiation of the lymphocyte. The production of an mRNA encoding the secreted form of the protein or the membrane-bound form could then be dictated by the choice of the appropriate poly(A) site.

3.2.2. Function of poly(A). Given the early time of addition of poly(A) to nuclear RNA, one might imagine that poly(A) was involved in some aspect of RNA processing. Alternatively, poly(A) could be involved in the transport of sequences from the nucleus to the cytoplasm; indeed, this possibility was suggested from the data of early experiments employing the drug cordycepin (45-47). However, it would appear that possibly the only function for the poly(A) segment may be in the cytoplasmic stabilization of mRNA. First, prevention of poly(A) addition to adenovirus nuclear RNA (using cordycepin) had no effect on the subsequent splicing of the RNA (48). Second, when measured with very short labeling times, there appears to be almost normal transport of this RNA to the cytoplasm (49). Once in the cytoplasm, this poly(A)$^-$ RNA does not accumulate but instead turns over rapidly. This result is consistent with other findings that measured the stability of mRNAs with and without their poly(A) (50-54). The conclusion from these experiments was that an RNA that was normally very stable upon injection into frog oocytes or mammalian cells was unstable if the poly(A) was first removed. Stability could be restored by the *in vitro* addition of poly(A).

3.2.3. Splicing. As indicated above, the existence of spliced eukaryotic mRNAs was first discovered through an electron microscopic (35,36) and biochemical (55) analysis of the structure of the late adenovirus mRNAs as well as the early viral mRNA (56). In fact, all of the adenovirus RNAs, with the exception of the mRNA encoding the virion protein IX (57) are spliced. Although the mechanism of accurate splicing as well as the definition of the factors involved is still not clear, recent advances in *in vitro* splicing systems should provide the means for such analysis (58-61).

Of course the initial discovery of splicing of the adenovirus RNAs together with RNA mapping experiments that demonstrated families of mRNAs with co-terminal 3' ends yielded the additional important aspect that most of the splicing pathways were alternate (62-66); that is, more than one mRNA could derive from a particular precursor RNA. This finding suggested that splicing could be regulatory in the sense of altering the products of

splicing in a controlled fashion. That such was indeed the case was demonstrated through an analysis of mRNAs produced from the late transcription unit during early or late infection. The L1 region is productively expressed both early and late in the lytic cycle. During late infection, three mRNAs are produced from a single 7.7 kb precursor RNA (27,28,62). However, early in infection a single mRNA, distinct in structure from the three late RNAs, is the product of the very same precursor (27,28,67).

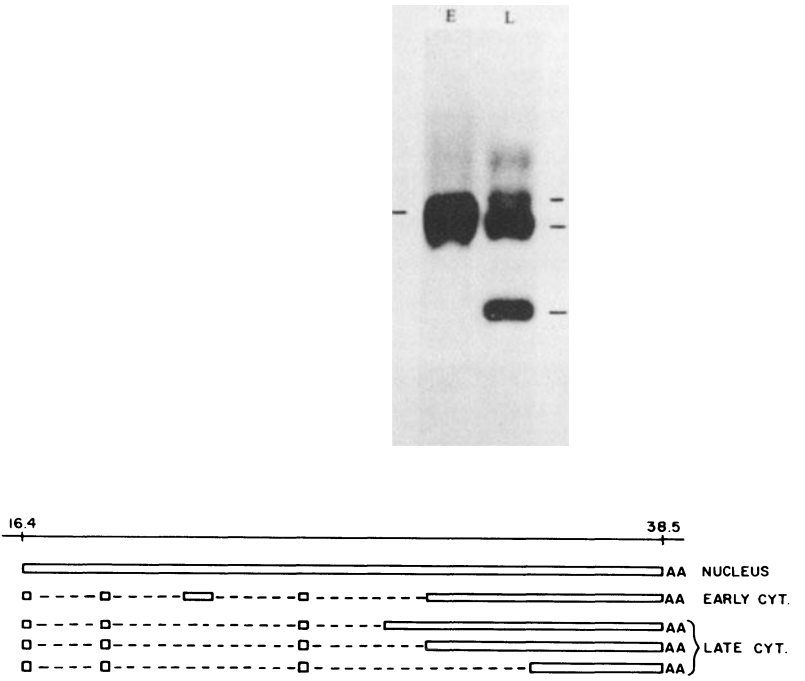

FIGURE 5. Splicing products of the adenovirus L1 region early and late in infection. Depicted is a Northern blot of poly(A)+ cytoplasmic RNA prepared from early (E) or late (L) adenovirus infected cells and hybridized with an L1-specific probe. The schematic depicts the structure of the unspliced common nuclear precursor RNA and the spliced mRNA products generated early or late in infection. Reproduced from (27) with permission.

Thus, alternative RNA splicing pathways can in fact be varied so as to achieve regulation of the final output of the gene. Presumably there must be viral specific factors produced early or late, or both, that direct the proper splicing pathway.

3.2.4. Other processing events. In addition to the events of polyadenylation and splicing of non-contiguous sequences, nuclear mRNA precursors are also modified by posttranscriptional methylation and 5' cap formation. The details of these events have been thoroughly reviewed elsewhere (68,69). Our experiments have dealt with two aspects of these modification events: the timing of cap formation and the relative conservation of internal methylation during RNA processing.

The kinetics of addition of 5' cap structure to RNA from the late adenovirus transcription unit was measured (70). Nuclear RNA labeled for only 20 minutes were found to be capped with no evidence for uncapped molecules. Furthermore, size fractionation of the labeled RNA revealed that even nascent RNA chains were quantitatively capped. It would thus appear that capping is a very early event, tightly coupled to transcription initiation.

Adenovirus late mRNAs, like many other eukaryotic mRNAs, possess posttranscriptional methylations at the N6 position of adenylate residues (71,72). That these internal methylations may be important in RNA processing was suggested by experiments that demonstrated that the methyl groups were conserved during RNA processing and transport to the cytoplasm (73). Furthermore, it was found that methylation occurred as an early event in RNA processing since unspliced precursor RNAs were methylated. Thus, methylation is selective in the sense that only those residues that are destined to become mRNA receive methylations. Of course, this result presents the intriguing possibility that the methyl groups themselves are directing the final splicing events. Such a conclusion will require mutants deficient in methylation or drugs that are selective; neither of these conditions are presently available.

3.3. Nuclear-Cytoplasmic Transport

That there is in fact a transport process that selectively delivers RNA from the nucleus to the cytoplasm is demonstrated by the observation that intron sequences do not leave the nucleus (26). However, the mechanism for selective transport, the factors involved in transport, as well as the question of whether transport plays a role in controlling gene expression remain unanswered. A suggestion that transport may be an event subject to control comes from experiments measuring cellular mRNA metabolism during adenovirus infection. During late viral infection, the only newly synthesized mRNA that reaches the cytoplasm is viral (74,75). The transcription of several specific HeLa genes was measured using cloned cDNAs

and found to be the same in uninfected and viral-infected cells (76) consistent with previous results that measured total, non-ribosomal, nuclear RNA synthesis (74). It would therefore appear that some event prior to cytoplasmic appearance but after transcription is affected by the virus. Whether this really is a direct and specific block in transport or possibly an indirect result due to the generation of improperly processed cellular transcripts, is yet to be determined.

3.4 Cytoplasmic mRNA Stability

Once a functional mRNA is generated in the nucleus and sent to the cytoplasm, it has two potential fates. The mRNA could either engage ribosomes and participate in directing protein synthesis or the mRNA could be degraded. Thus, any mechanism that affects mRNA turnover has a direct effect on the concentration of functional mRNA and therefore is a controlling factor in gene expression. Adenovirus mRNAs are subject to specific degradative processes in the cytoplasm that appear to operate to regulate mRNA levels. First of all, the early adenovirus mRNAs possess very short cytoplasmic half-lives, in the range of 20 minutes (77,78). Of importance is the fact that these half-lives are subject to change depending on certain circumstances in the cell. The most definitive case concerns the stability of these RNAs when produced in cells infected with an adenovirus temperature sensitive mutant, ts 125, possessing a lesion in the gene encoding the 72kd DNA binding protein. At the permissive temperature, the early RNAs decay with the normal 20 minute half-life but at the non-permissive temperature the half-lives of the RNAs increase five to tenfold (78). In addition, when early E1A and E1B mRNAs are synthesized in adenovirus-transformed cells, thus in the absence of the 72kd protein, they are stable. Introduction of the protein into these cells results in rapid degradation of the RNAs (Babich & Nevins, unpublished). It would therefore appear that the stability of a specific set of mRNAs is controlled by a specific viral protein.

That the protein is in fact directly involved in affecting mRNA stability is strongly supported by recent experiments that demonstrate an interaction between the protein and mRNA within the cell.

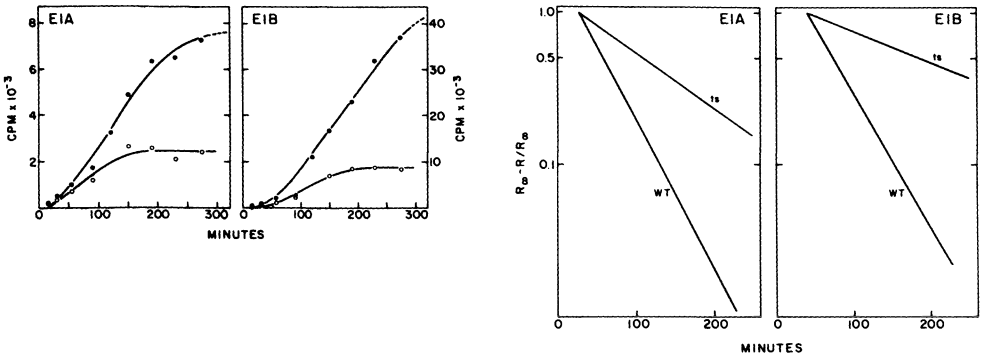

FIGURE 6. Stability of the E1A and E1B adenovirus mRNAs in wild type and ts125 infected cells. Depicted in the left panels are the kinetics of cytoplasmic accumulation of E1A and E1B mRNA in cells infected with wild type Ad5 (o) or ts125 (●) at 41°. In the right panels are decay curves generated from the accumulation data. For details, see ref. 68. Reproduced with permission.

By stabilizing ribonucleoprotein complexes with UV irradiation, it has been possible to identify the 72kd protein in poly(A)$^+$ RNA - protein complexes (Babich & Nevins, unpublished). Furthermore, it would appear that the protein is indeed associated primarily with viral mRNA, as predicted by the previous in vivo metabolic experiments. This system has great promise as it represents the only instance of the identification of a specific protein that is involved in affecting the stability of messenger RNA.

4. SUMMARY

One of the primary aims of present-day molecular biology is to achieve an understanding of the mechanisms by which eukaryotic genes are regulated. Certainly, a final understanding of cellular growth control and cellular differentiation to name just two examples, demands such an understanding. Furthermore, to begin to probe the questions of alterations of gene expression such as are likely taking place in malignant growth requires a complete knowledge of the normal course of expression of a gene. Through the analysis of gene expression from the nuclear DNA viruses, as discussed here, much of the framework of the pathway of mRNA biogenesis has been elucidated. What is left to be uncovered are the components of the system, the macromolecules involved in the various biochemical reactions.

Acknowledgements

This work has been supported by grants from the NIH (GM26765-05) and the ACS (MV141a). The author is the recipient of a Research Career Development Award (CA00666).

REFERENCES

1. Tooze, J. DNA Tumor Viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y. 1981.

2. Nevins, J. R. Meth. Enzymol. 65: 768-785. 1980.

3. Nevins, J. R. Cell 26: 213-220. 1981.

4. Nevins, J. R. Cell 29: 913-919. 1982.

5. Imperiale, M. J., Feldman, L. T. and Nevins, J. R. Cell 35: 127-136. 1983.

6. Nevins, J. R. and Chen-Kiang, S. Adv. Virus Res. 26: 1-35. 1981.

7. Breathnach, R. and Chambon, P. Ann. Rev. Biochem. 50: 349-383. 1981.

8. Shenk, T. Curr. Topics Microbiol. Immunol. 93: 25-46. 1981.

9. Nevins, J. R. Ann. Rev. Biochem. 52: 441-466. 1983.

10. Darnell, J. E. Nature 297: 365-371. 1982.

11. Ziff, E. B. and Evans, R. M. Cell 15: 1463-1476. 1978.

12. Weil, P. A., Luse, D. S., Segall, J. and Roeder, R. G. Cell 18: 469-484. 1979.

13. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. and Gefter, M. L. Proc. Nat'l. Acad. Sci. USA 77: 3855-3859. 1980.

14. Hagenbuchle, O. and Schibler, U. Proc. Nat'l. Acad. Sci. USA 78: 2283-2286. 1981.

15. Khoury, G. and Gruss, P. Cell 33: 313-314. 1983.

16. Nevins, J. R., Ginsberg, H. S., Blanchard, J.-M., Wilson, M. C. and Darnell, J. E. J. Virol. 32: 727-733. 1979.

17. Nevins, J. R. and Winkler, J. J. Proc. Nat'l. Acad. Sci. USA 77: 1893-1897. 1980.

18. Handa, H., Kingston, R. E. and Sharp, P. A. Nature 302: 545-547. 1983.

19. Tjian, R. Cell 26: 1-2. 1981.

20. Fowlkes, D. M., Lord, S. T., Linne, T., Pettersson, U. and Philipson, L. J. Mol. Biol. 132: 163-180. 1979.

21. Berk, A. J., Lee, F., Harrison, T., Williams, J. and Sharp, P. A. Cell 17: 935-944. 1979.

22. Jones, N. and Shenk, T. Proc. Nat'l. Acad. Sci. USA 76: 3665-3669. 1979.

23. Katze, M. G., Persson, H. and Philipson, L. Mol. Cell. Biol. 1: 807-813. 1981.

24. Kao, H.-T. and Nevins, J. R. Mol. Cell. Biol. 3: 2058-2065. 1983.

25. Feldman, L. T., Imperiale, M. J. and Nevins, J. R. Proc. Nat'l. Acad. Sci. USA 79: 4952-4956. 1982.

26. Fraser, N. W., Nevins, J. R., Ziff, E. B. and Darnell, J. E. J. Mol. Biol. 129: 643-656. 1979.

27. Nevins, J. R. and Wilson, M. C. Nature 290: 113-118. 1981.

28. Akusjarvi, G. and Persson, H. Nature 292: 420-426. 1981.

29. Shaw, A. R. and Ziff, E. B. Cell 22: 905-916. 1980.

30. Alt, F. W., Bothwell, A. L. M., Knapp, M., Siden, E., Mather, E., Koshland, M. and Baltimroe, D. Cell 20: 293-301. 1980.

31. Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. Cell 20: 313-319. 1980.

32. Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W. and Tonegawa, S. Cell 24: 353-365. 1981.

33. Evans, R., Weber, J., Ziff, E. and Darnell, J. E. Nature 278: 367-370. 1979.

34. Fraser, N. W., Sehgal, P. B. and Darnell, J. E. Proc. Nat'l. Acad. Sci. USA 76: 2571-2575. 1979.

35. Berget, S., Moore, C. and Sharp, P. A. Proc. Nat'l. Acad. Sci. USA 74: 3171-3175. 1977.

36. Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J. Cell 12: 1-8. 1977.

37. Nevins, J. R. and Darnell, J. E. Cell 15: 1477-1493. 1978.

38. Brawerman, G. Prog. Nucl. Acids Res. Mol. Biol. 17: 117-148. 1976.

39. Brawerman, G. CRC Crit. Rev. Biochem. 10: 1-38. 1981.

40. Nevins, J. R. In: Processing of RNA (Ed. D. Apiricon), CRC Press, Boca Raton, FL., 1983, in press.

41. Nevins, J. R., Blanchard, J.-M. and Darnell, J. R. J. Mol. Biol. 144: 377-386. 1980.

42. Ford, J. P. and Hsu, M.-T. J. Virol. 28: 795-801. 1978.

43. Hofer, E. and Darnell, J. E. Cell 23: 585-593. 1981.

44. Nevins, J. R. J. Mol. Biol. 130: 493-506. 1979.

44a. Manley, J. L., Sharp, P. A. and Gefter, M. L. J. Mol. Biol. 159: 581-599. 1982.

44b. Salditt-Georgieff, M., Harpold, M., Sawicki, S., Nevins, J. and Darnell, J. E. J. Cell Biol. 86: 844-848. 1980.

45. Darnell, J. E., Philipson, L., Wall, R. and Adesnik, M. Science 174: 507-510. 1974.

46. Philipson, L., Wall, R., Glickman, G. and Darnell, J. E. Proc. Nat'l. Acad. Sci. USA 68: 2806-2809. 1971.

47. Penman, S., Rosbash, M. and Penman, M. Proc. Nat'l. Acad. Sci. USA 67: 1878-1885. 1970.

48. Zeevi, M., Nevins, J. R. and Darnell, J. E. Cell 26: 39-46. 1981.

49. Zeevi, M., Nevins, J. R. and Darnell, J. E. Mol. Cell. Biol. 2: 517-525. 1982.

50. Huez, G., Bruck, C. and Cleuter, Y. Proc. Nat'l. Acad. Sci. USA 78: 908-911. 1981.

51. Huez, G., Marbaix, G., Gallwitz, D., Weinberg, E., Devos, R., Hubert, E. and Cleuter, Y. Nature 271: 572-573. 1978.

52. Huez, G., Marbaix, G., Hubert, E., Leclercq, M., Nudel, U., Soreq, H., Salomon, R., Lebleu, B., Revel, M., Littauer, U. Z. Proc. Nat'l. Acad. Sci. USA 71: 3143-3146. 1974.

53. Marbaix, G., Huez, G., Burny, A., Cleuter, Y., Hubert, E., Leclercq, M., Chantrenne, H., Soreq, H., Nudel, U. and Littauer, U. Z. Proc. Nat'l. Acad. Sci. USA 72: 3065-3067. 1975.

54. Nudel, U., Soreq, H., Littauer, U. Z., Marbaix, G., Huez, G., Leclercq, Hubert, E. and Chantrenne, H. Eur. J. Biochem. 64: 115-121. 1976.

55. Klessig, D. F. Cell 12: 9-21. 1977.

56. Kitchingman, G. R., Lai, S.-P. and Westphal, H. Proc. Nat'l. Acad. Sci. USA 74: 4392-4395. 1977.

57. Alestrom, P., Akusjarvi, G., Perricaudet, M., Mathews, M. B., Klessig, D. F. and Pettersson, U. Cell 19: 671-681. 1980.

58. Goldenberg, C. J. and Raskas, H. J. Proc. Nat'l. Acad. Sci. USA 78: 5430-5434. 1981.

59. Weingartner, B. and Keller, W. Proc. Nat'l. Acad. Sci. USA 78: 4092-4096. 1981.

60. Padgett, R. A., Hardy, S. F. and Sharp, P. A. Proc. Nat'l. Acad. Sci. USA 80: 5230-5234. 1983.

61. Hernandez, N. and Keller, W. Cell 35: 89-99. 1983.

62. Nevins, J. R. and Darnell, J. E. J. Virol. 25: 811-823. 1978.

63. McGrogan, M. and Raskas, H. J. Proc. Nat'l. Acad. Sci. USA 75: 625-629. 1978.

64. Chow, L. T. and Broker, T. R. Cell 15: 497-510. 1978.

65. Ziff, E. and Fraser, N. J. Virol. 25: 897-906. 1978.

66. Berk, A. J. and Sharp, P. A. Cell 14: 695-711. 1978.

67. Chow, L. T., Broker, T. R. and Lewis, J. B. J. Mol. Biol. 134: 265-303. 1979.

68. Banerjee, A. K. Microbiol. Rev. 44: 175-205. 1980.

69. Shatkin, A. J. Cell 9: 645-653. 1976.

70. Babich, A., Nevins, J. R. and Darnell, J. E. Nature 287: 246-248. 1980.

71. Moss, B. and Koczot, F. J. Virol. 17: 385-392. 1976.

72. Sommer, S., Salditt-Georgieff, M., Bachenheimer, S., Darnell, J. E., Furuichi, Y., Morgan, M. and Shatkin, A. J. Nucl. Acids. Res. 3: 749-765. 1976.

73. Chen-Kiang, S., Nevins, J. R. and Darnell, J. E. J. Mol. Biol. 135: 733-752. 1979.

74. Beltz, G. A. and Flint, S. J. J. Mol. Biol. 131: 353-373. 1979.

75. Castiglia, C. L. and Flint, S. J. Mol. Cell. Biol. 3: 662-671. 1983.

76. Babich, A., Feldman, L. T., Nevins, J. R., Darnell, J. E. and Weinberger, C. Mol. Cell. Biol. 3: 1212-1221. 1983.

77. Wilson, M. C. and Darnell, J. E. J. Mol. Biol. 148: 231-251. 1981.

78. Babich, A. and Nevins, J. R. Cell 26: 371-379. 1981.

2

THE INTERACTION BETWEEN VIRAL MESSENGER RNA AND EUKARYOTIC INITIATION FACTOR 2, A PROTEIN INVOLVED IN TRANSLATIONAL CONTROL

RAYMOND KAEMPFER

ABSTRACT

The messenger RNA species from Mengovirus or satellite tobacco necrosis virus (STNV) rank among the most efficient templates for translation known. The molecular basis for their outstanding ability to initiate translation, at the expense of weaker mRNA species, is shown to reside to a large extent in a very high affinity for a eukaryotic initiation factor, eIF-2. This protein is not only responsible for binding of Met-$tRNA_f$ at initiation of translation but, in addition, it interacts with mRNA. The binding of eIF-2 to Mengovirus or STNV RNA occurs at highly specific sequences. Nucleotide mapping analysis is used to show that eIF-2 recognizes, by itself, the sequences that constitute the ribosome binding sites in these viral mRNA molecules. Upon binding, eIF-2 induces a conformational change in the STNV RNA structure that may increase its accessibility to Met-$tRNA_f$ and the 40S ribosomal subunit. Studies of translational competition between mRNA species, and of translational inhibition by double-stranded RNA or monovalent anions support the concept that this interaction between mRNA and eIF-2 indeed occurs during protein synthesis and is critical to translational control. They reveal a direct correlation between the affinity of an mRNA for eIF-2 and its ability to compete in translation. Apparently, eIF-2 recognizes in different mRNA species a common conformation existing around the initiation codon, but differing in subtle ways that determine individual binding affinities.

INTRODUCTION

Viral messenger RNA is often characterized by its ability to act as a highly efficient template for translation. Indeed, the lytic replication of DNA or RNA viruses is usually associated with preferential translation of viral mRNA, at the expense of host mRNA. Thus, the genomic RNA of some picornaviruses, such as Mengovirus or satellite tobacco necrosis virus

Y. Becker (ed.), VIRAL MESSENGER RNA.

(STNV), rank among the most efficient mRNA species known (1).

The molecular basis for this efficiency must be sought in the exquisite ability of such viral mRNA species to initiate polypeptide chain synthesis. The initiation step is usually the rate-limiting one in translation (1). Within the initiation process, a crucial stage for control is the attachment of an mRNA molecule to a 40 S ribosomal subunit. Indeed, most cases of translational control concern this step (1). Once mRNA has entered a 40 S initiation complex, formation of the complete encoded polypeptide chain is generally assured. The functional properties of mRNA and the translation components involved in the events leading up to this complex are reviewed more fully elsewhere (1). Here, we examine molecular properties of STNV and Mengovirus RNA, in particular their very high affinity for a central component of the cell's protein-synthesizing machinery, eukaryotic initiation factor 2 (eIF-2), that may underlie their outstanding translation efficiency.

Evidence will be reviewed in support of the concept that eIF-2 recognizes in viral mRNA molecules the sequence and conformation that constitute the ribosome binding site. The ability of a given mRNA species to compete in translation is shown to correlate directly with its affinity for eIF-2. These properties impart on the interaction between mRNA and eIF-2 a critical role in differential gene expression at the level of translation. The implications of these findings for translational control (1) and for viral pathogenesis (2) have been reviewed elsewhere; the protein-nucleic acid interactions are analyzed here.

THE MESSENGER RNA-BINDING ACTIVITY OF EUKARYOTIC INITIATION FACTOR 2

Binding of mRNA to the 40 S ribosomal subunit cannot take place unless methionyl-tRNA$_f$ (Met-tRNA$_f$) is first bound (3, 4). This means that the recognition and binding of Met-tRNA$_f$ are an integral part of the mRNA binding process. Met-tRNA$_f$ is bound with absolute specificity by the initiation factor eIF-2 (5). This binding requires GTP and leads to formation of a ternary complex, eIF-2/Met-tRNA$_f$/GTP, that subsequently binds to the 40 S ribosomal subunit (e.g., ref. 6). The unique property of providing Met-tRNA$_f$ already imparts on eIF-2 a crucial role in the binding of mRNA. While additional factors participate in stable binding of mRNA (e.g., refs. 7 and 8; see ref. 1 for review), none can act in the absence of eIF-2.

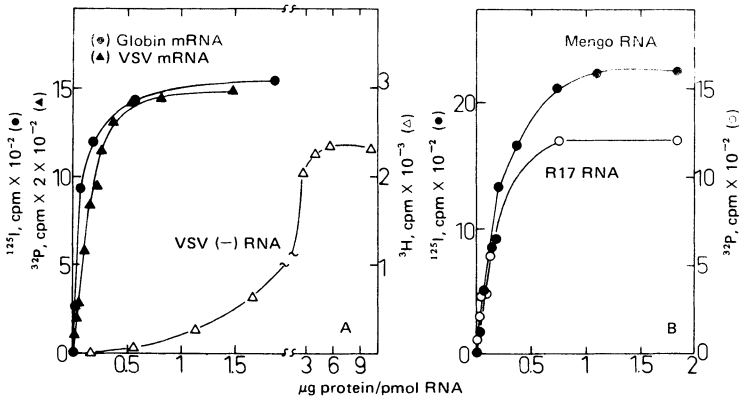

Fig. 1. First-order binding of mRNA to eIF-2. ^{125}I-labeled purified globin mRNA (0.031 pmol; 2300 cpm), ^{32}P-labeled VSV mRNA (3.6 pmol; 800 cpm), ^{3}H-labeled RNA extracted from VSV virions (0.21 pmol; 2800 cpm), ^{125}I-labeled Mengovirus RNA (0.0023 pmol; 2630 cpm), and ^{32}P-labeled R17 RNA (2.25 pmol; 1660 cpm) were incubated with increasing amounts of eIF-2, and RNA binding was assayed. Background without protein was subtracted; for all RNA species, this background was less than 3%. Note scale change in curve of VSV negative-strand RNA. Reproduced with permission (12).

In addition to binding Met-tRNA$_f$, eIF-2 itself can bind to mRNA (9-14). This binding is specific in that all mRNA species tested possess an effective binding site for eIF-2, including mRNA species lacking the 5'-terminal cap or 3'-terminal poly (A) moieties (12), while RNA species not serving as mRNA, such as tRNA (10-13, 15), ribosomal RNA (10, 16), or negative-strand viral RNA (12) bind far more weakly.

Figure 1 illustrates the extent of complex formation between highly purified eIF-2 and various labeled mRNA species, plotted as a function of the ratio of protein to mRNA molecules. Complex formation is assayed conveniently by the eIF-2-dependent retention of mRNA on nitrocellulose membrane filters (17). The saturation binding curves of globin mRNA, vesicular stomatitis virus (VSV) mRNA, Mengovirus RNA and R17 phage RNA all exhibit apparent first-order binding characteristics. As in the case of the lac repressor and its operator DNA (18), this property indicates that binding of one eIF-2 molecule to a molecule of mRNA is sufficient to cause retention of this mRNA molecule on the filter. These mRNA species, therefore, all possess an effective binding site for eIF-2. Neither the 7-methylguanosine 5'-phosphate moiety at the 5' end of mRNA, nor the poly (A) sequence at the 3' end seem to be required for binding, because

Mengovirus RNA (which lacks the cap) and R17 RNA or, as will be seen below, STNV RNA (which lack both cap and poly (A)) are bound efficiently.

If the interaction between eIF-2 and mRNA involves recognition of a site essential for translation, then RNA species lacking such a site should exhibit different binding properties. To test this point, the interaction between eIF-2 and negative-strand RNA extracted from VSV virions was studied. This RNA cannot serve as messenger template but is transcribed into complementary mRNA sequences during infection (19). As seen in Figure 1, the saturation binding curve for VSV negative-strand RNA has a pronounced sigmoid shape, typical of a multi-order interaction. At saturation, however, nearly all these RNA molecules are able to form complexes. It is unlikely that the higher molecular weight of VSV negative-strand RNA (3.5×10^6 M_r) (20) is responsible for the sigmoid shape, because binding of Mengovirus RNA (molecular weight 2.6×10^6) follows the shape observed for the far smaller globin mRNA or VSV mRNA species, and at plateau the extent of binding of Mengovirus RNA is particularly high. Hence, the sigmoid binding curve indicates that VSV negative-strand RNA possesses multiple, weak binding sites for eIF-2, but lacks the high-affinity binding site present in VSV mRNA and other mRNA species.

The fact that eIF-2 binds to RNA molecules in general, even though it clearly prefers mRNA, initially raised some question as to the specificity of its interaction with mRNA. However, a common property of proteins that recognize specific sites in nucleic acids, such as the lac repressor, is their tendency to bind with low affinity to non-specific sequences (21, 22). As will be seen below, in the case of eIF-2, specificity in its interaction with mRNA is now supported by both structural and functional evidence.

eIF-2 is thought to be composed of one copy each (23) of three non-identical subunits: α (32,000 M_r), β (35,000 M_r) and γ (55,000 M_r). The α-subunit is the substrate for various specific eIF-2 kinases that play a pivotal role in translational control (1); this subunit apparently binds GTP (10). The mRNA-binding property is a function of eIF-2 itself. Thus, eIF-2 is preferentially retained on mRNA-cellulose columns and upon elution is active in translation and in GTP-dependent binding of Met-$tRNA_f$ (12, 26). Binding of Met-$tRNA_f$ is inhibited competitively by mRNA (24, 11, 25, 14). Most convincingly, binding of mRNA to purified eIF-2 preparations can be inhibited completely by competing amounts of Met-$tRNA_f$,

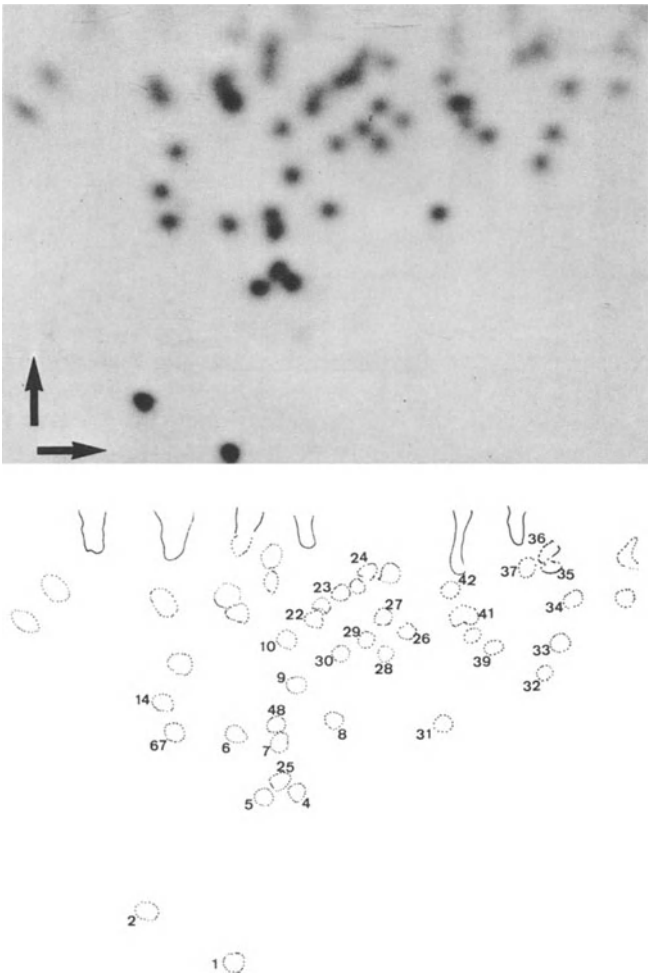

Fig. 2. Fingerprint of the genomic RNA of Mengovirus. ^{32}P-labeled Mengovirus RNA was digested to completion with RNase T_1, and the resulting oligonucleotides were resolved by two-dimensional polyacrylamide gel electrophoresis. The horizontal and vertical arrows indicate the direction of migration in the first and second dimensions, respectively. Numbers denote individual T_1 oligonucleotides. Reproduced with permission (30).

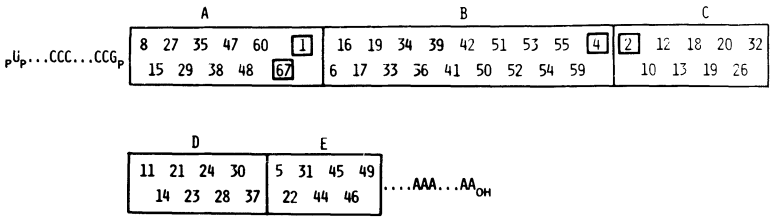

Fig. 3. Physical map of Mengovirus RNA. The major T_1 oligonucleotides shown in Fig. 2 were ordered relative to the poly (A) stretch at the 3'-terminus of the molecule. The order of individual nucleotides within a given region (A-E) has not been established. Boxed oligonucleotides 1, 2, 4 and 67 (see text) have been placed at arbitrary positions within their respective regions. Reproduced with permission (30).

provided GTP is present (15). Thus, mRNA and Met-tRNA$_f$ are mutually exclusive in their binding to eIF-2, suggesting that during initiation of translation, the interaction of a molecule of mRNA with eIF-2 located on the 40 S ribosomal subunit may cause displacement of the previously bound Met-tRNA$_f$ from this factor. We have proposed (15) that during initiation, three processes may actually occur in one step: binding of mRNA to eIF-2, displacement of Met-tRNA$_f$ from eIF-2, and base-pairing between mRNA and Met-tRNA$_f$.

SPECIFIC SEQUENCES IN VIRAL MESSENGER RNA ARE RECOGNIZED BY eIF-2

Globin mRNA molecules lacking the 3'-terminal poly (A) tail or an additional 90 nucleotides from the 3'-untranslated region bind to eIF-2 as tightly as native globin mRNA, with an apparent dissociation constant of 5×10^{-9} M at 150 mM KCl (13). On the other hand, cap analogs inhibit binding of both mRNA and Met-tRNA$_f$ to eIF-2 (24). Although this could suggest that the cap in mRNA interacts with eIF-2, the genomic RNA species from Mengovirus or STNV bind extremely well to eIF-2, in fact even better than globin mRNA, yet they do not carry a cap structure (24, 16, 27). This, and the observation that eIF-2 prefers native globin mRNA by five orders of magnitude over cap analogs led to the suggestion that binding of eIF-2 to mRNA occurs primarily at an internal sequence (24).

A relationship between binding site sequences for eIF-2 and ribosomes in Mengovirus RNA

Mengovirus RNA has a length of about 7,500 nucleotides and contains

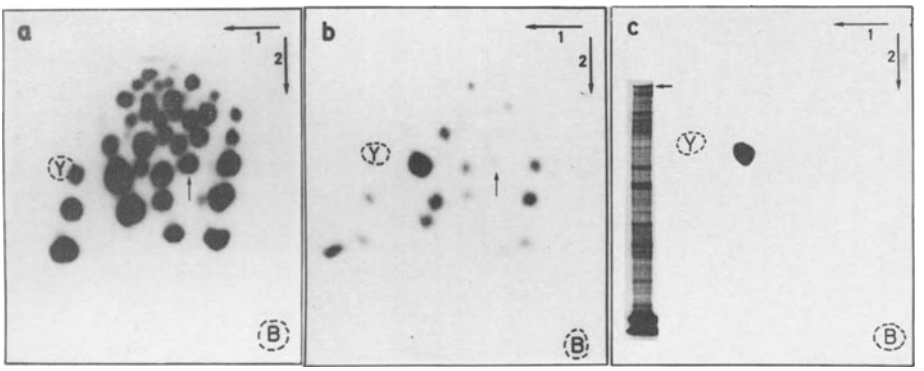

Fig. 4. Analysis of pancreatic ribonuclease digests of total (a), eIF-2 selected (b), and purified, intact (c) 5'-end-labeled STNV RNA by electrophoresis at pH 3.5 in the first dimension and homochromatography on a DEAE-thin-layer plate in the second dimension (see arrows), followed by autoradiography. (a) Total 5'-end-labeled STNV RNA. Arrow, position of pApGpCp. (b) eIF-2-selected RNA. Total 5'-end-labeled STNV RNA was incubated under conditions for RNA binding with increasing amounts of purified eIF-2; bound RNA was isolated by retention on 0.45-µm nitrocellulose filters and eluted with 0.1% $NaDodSO_4$, followed by ethanol precipitation. At saturating amounts of eIF-2, 70% of the labeled RNA could be bound; in the sample analyzed, 7.2% of the input label was bound by a limiting amount (0.3 µg) of eIF-2. Arrow, position of pApGpCp. (c) Intact 5'-end-labeled STNV RNA purified from the total kinase-treated RNA mixture by electrophoresis in a 4% polyacrylamide gel containing 7 M urea (arrow points to intact RNA). Marker dyes were methyl orange (Y) and xylene cyanol FF (B). Reproduced with permission (16).

a poly (C) tract located several hundred nucleotides from the 5' end (28). In the closely related foot-and-mouth disease virus RNA, the major initiation site for translation lies downstream from the poly (C) tract (29). When Mengovirus RNA was digested to completion with ribonuclease T_1 and the resulting oligonucleotides were separated by two-dimensional polyacrylamide gel electrophoresis, Dr. R. Perez-Bercoff (30) obtained the fingerprint shown in Fig. 2. The poly (C) tract migrates too slowly in the first dimension to be seen in the fingerprint, due to its length (81 nucleotides) and strong positive charge at pH 3.3 (31). Fig. 3 depicts a physical map of Mengovirus RNA, constructed by Dr. Perez-Bercoff (30). It shows the major T_1 oligonucleotides, ordered with respect to their distance from the 3'-terminus of the molecule. The order of oligonucleotides within a given region (demarcated by vertical lines) was not established. No major T_1 oligonucleotide could be assigned to the 5'-proximal region preceding the poly (C) tract.

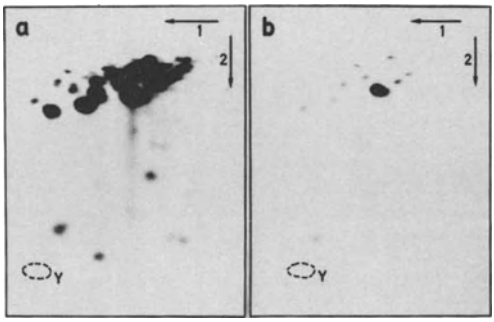

Fig. 5. Analysis of T_1 ribonuclease digests of total (a) and eIF-2 selected (b) 5'-end-labeled STNV RNA by electrophoresis at pH 3.5 in the first dimension and by homochromatography on a polyethyleneimine thin-layer plate in the second dimension (arrows). The intact RNA gave one spot comigrating with the major one in b (not shown). Y, methyl orange. Reproduced with permission (16).

Ribosomes were bound to Mengovirus RNA in extracts of L 929 or Krebs ascites cells, to generate 40 S or 80 S initiation complexes that were isolated by sucrose gradient centrifugation. The RNA sequences protected in these complexes were isolated after ribonuclease T_1 digestion. In either complex, four unique T_1 oligonucleotides were protected: these were oligonucleotides 1, 2, 4 and 67 (boxed in Fig. 3). Although the precise location of these oligonucleotides, each 15 to 28 nucleotides in length, within Mengovirus RNA is not yet known, it is seen from Fig. 3 that they fall into at least two widely separated domains (30). It is not yet clear if these domains form a single site held together by secondary or tertiary interactions in the RNA molecule, or two separate sites for initiation of translation. The recovery of two distinct N-terminal tryptic peptides upon in vitro translation of Mengovirus RNA (32) tends to support the latter interpretation.

Labeled, intact Mengovirus RNA was offered to eIF-2 and the sequences protected by the initiation factor against ribonuclease T_1 digestion were isolated by their retention on nitrocellulose membrane filters, in complex with eIF-2 (30). When the protected RNA fragments were digested to completion with ribonuclease T_1 and fingerprinted, three specific oligonucleotides were recovered, and these were oligonucleotides 1, 2 and 4. Most likely, oligonucleotide 67 also interacts with eIF-2, but too weakly to survive the isolation procedure (30). This finding demonstrates high

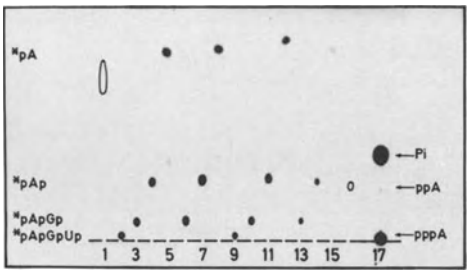

Fig. 6. Identification of the major products of pancreatic and T_1 ribonuclease digestion. The major spots in Figs. 4b and c and 5b, as well as in the ribonuclease T_1 digest of intact RNA, were eluted with triethylammonium bicarbonate (pH 9.0), digested as described below, and analyzed by chromatography on a polyethyleneimine thin-layer plate in 4 M urea/1 M HCOOH made to pH 4.3 with pyridine, followed by autoradiography. References were: optical density markers 5'-AMP (lane 1; overloading resulted in decreased mobility) and 5'-ADP (lane 16) and [γ-^{32}P]ATP (lane 17). Other lanes: major pancreatic ribonuclease spot of eIF-2-selected total RNA (Fig. 4b) (lanes 2-5) or of intact RNA (Fig. 4c) (lanes 9-12) after no additional digestion (lanes 2 and 9) or digestion with T_1 ribonuclease (lanes 3 and 10), T_2 ribonuclease (lanes 4 and 11), or P_1 ribonuclease (lanes 5 and 12); major T_1 ribonuclease spot of eIF-2-selected total RNA (Fig. 5b) (lanes 6-8) or intact RNA (lanes 13-15), after no additional digestion (lanes 6 and 13) or digestion with T_2 ribonuclease (lanes 7 and 14) or P_1 ribonuclease (lanes 8 and 15). The spots in lanes 1-15 are identified on the left; the spots in lanes 16 and 17 are identified on the right. Reproduced with permission (16).

sequence specificity in the interaction between eIF-2 and Mengovirus RNA and shows that out of a very large sequence, eIF-2 binds, by itself, precisely to those nucleotides in Mengovirus RNA that are protected by ribosomes. The virtual identity of the binding sites in Mengovirus RNA for ribosomes on one hand, and for eIF-2 on the other, suggests strongly that the binding of ribosomes to Mengovirus RNA is guided directly by eIF-2.

Sequence and conformation analysis of the binding site for eIF-2 in STNV RNA

A more detailed study was made of the interaction between STNV RNA and eIF-2. STNV RNA is particularly suitable for analyzing the binding site for eIF-2, because it is an efficient mRNA for translation (33, 34), has a known sequence of 1,239 nucleotides encoding only a single protein (35), lacks poly (A) and possesses an unmodified 5' end (36, 37) that can be labeled in vitro with polynucleotide kinase. As seen in the gel of Fig. 4c (inset), RNA isolated from STNV virions migrates, after 5'-end labeling, as a heterogeneous collection of fragments with only a minor amount of label

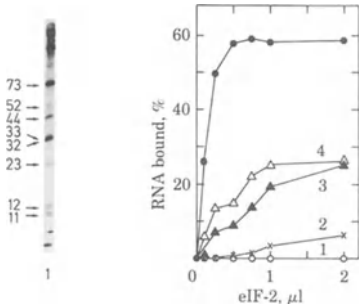

Fig. 7. eIF-2 saturation binding curves of intact STNV RNA and individual 5'-terminal fragments. Intact STNV RNA (●; 1,018 cpm) and fragments - 32-nucleotide (curve 1; 293 cpm), 44-nucleotide (curve 2; 283 cpm), 73-nucleotide (curve 3; 483 cpm) and 115-nucleotide (curve 4; 201 cpm) - isolated from a T_1 partial digest of intact, 5'-labeled STNV RNA (lane 1 on left; numbers denote nucleotide length) were incubated in the presence of the indicated amounts of eIF-2 (0.15 µg/µl), and the percentage of radioactivity retained on nitrocellulose filters was determined. Reproduced with permission (16).

in fully intact viral RNA (arrow). When intact RNA is excised from the gel and digested with pancreatic ribonuclease, a single spot is observed in the two-dimensional fingerprint (Fig. 4c). By contrast, the complete STNV RNA preparation is quite heterogeneous and contains at least 30 different 5' ends (Fig. 4a), attesting to the presence of many fragments originating from internal regions of the viral RNA molecule. When this RNA was offered to eIF-2 and RNA bound to the initiation factor was isolated and fingerprinted, one major spot was observed, migrating precisely as the 5' end of intact viral RNA (Fig. 4b). Likewise, the ribonuclease T_1 fingerprint of the 5' end-labeled, complete STNV RNA preparation displayed a complex collection of spots (Fig. 5a), while eIF-2-selected STNV RNA yielded primarily one spot (Fig. 5b), again migrating as the single spot generated from intact STNV RNA (not shown). The sequences of the material in the spots obtained from eIF-2-selected RNA were analyzed in Fig. 6 by additional digestion with T_1, T_2 and P_1 ribonucleases. The major spot generated upon pancreatic ribonuclease digestion (Fig. 4b) was identified as pApGpUp (Fig. 6, lanes 2-5) (the only other possibility, pApGpCp, is not compatible with Fig. 1b); the major spot in the T_1 ribonuclease pattern (Fig. 5b) was identified as pApGp (Fig. 6, lanes 6-8). The 5' end of isolated, intact STNV RNA yielded identical patterns (lanes 9-12 and 13-15, respectively). Indeed, STNV RNA is known to start with the sequence

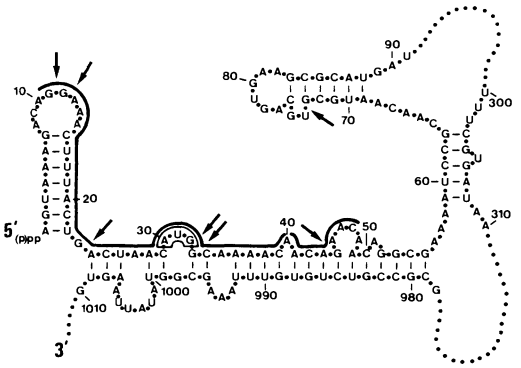

Fig. 8. Secondary structure model for the 5' end of STNV RNA. The model (35) depicts stable secondary interactions. Heavy line, nucleotides protected by 40 S ribosomal subunits against nucleases (39); arrows, prominent sites of ribonuclease T_1 cleavage (Fig. 7, lane 1). For eIF-2 binding site, see text. Reproduced with permission (16).

5'(p)ppApGpUp.. (37). These results show that eIF-2 recognizes and binds selectively to those STNV RNA fragments starting with the 5' end of the intact RNA (16).

To map the eIF-2 binding site more exactly, intact 5' end-labeled STNV RNA was isolated and digested partially with ribonuclease T_1, to generate a nested set of labeled RNA fragments, all containing the 5' end of intact RNA and extending to various points within the molecule. Prominent in the cleavage pattern were fragments 11, 12, 23, 32, 33, 44, 73 and 115 nucleotides long (Fig. 7, lane 1). Cleavage products of these lengths can be predicted from a model of the secondary structure of STNV RNA (35), as shown in Fig. 8. Arrows denote G residues sensitive to nuclease attack. Note that the 32-nucleotide long fragment terminates with the AUG initiation codon for coat protein synthesis (38). Fragments of discrete size were excised from the gel, and their ability to bind eIF-2 was studied. As seen in Fig. 7, eIF-2 does not bind to the 32-nucleotide long fragment (curve 1), but it does bind distinctly to the 44-nucleotide fragment (curve 2) or larger ones (curves 3 and 4). This places the 3'-proximal boundary of the eIF-2 binding site at or near nucleotide 44.

Fig. 7 also shows that the affinity of eIF-2 for short 5'-terminal STNV RNA fragments increases with their length and is greatest for fully intact viral RNA (upper curve). This finding suggests that the structural features leading to recognition by eIF-2 are not only localized in the

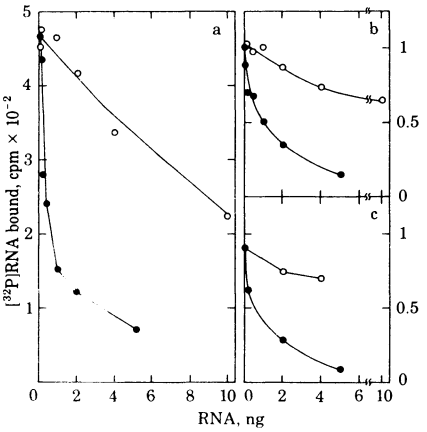

Fig. 9. Competition binding behavior of intact (a) STNV RNA and the 73-nucleotide (b) and 115-nucleotide (c) 5'-terminal fragments. Purified, intact STNV RNA (1,290 cpm) and 115-nucleotide (250 cpm) and 73-nucleotide (303 cpm) fragments isolated from a digest such as that of Fig. 7 (lane 1) were incubated with a limiting amount of eIF-2, and the indicated amounts of >90% intact, unlabeled STNV RNA (●) or E. coli rRNA (o). The amount of radioactivity retained on nitrocellulose filters was determined. Reproduced with permission (16).

5'-terminal region but are also determined, or strengthened, by the overall conformation of the viral RNA molecule. Nevertheless, binding of eIF-2 to short 5'-terminal STNV RNA fragments occurs with the same specificity as to intact RNA. This can be seen from Fig. 9, a competition experiment in which binding of intact 5' end-labeled STNV RNA, or 5'-terminal fragments, to a limiting amount of eIF-2 was studied in the presence of increasing amounts of unlabeled, intact STNV RNA, or Escherichia coli ribosomal RNA. eIF-2 binds with at least 20-fold greater affinity to STNV RNA than to ribosomal RNA (a), and it is clear that this differential competition is preserved for the 73- and 115-nucleotide fragments (b and c). Even binding of the 44-nucleotide fragment was still specific: 58% of the binding was inhibited by 5 ng of intact STNV RNA, but only 11% was inhibited by 5 ng of E. coli ribosomal RNA (16).

In the experiments shown in Fig. 10, purified, intact 5' end-labeled STNV RNA was partially digested with various nucleases. eIF-2 was then added and binding was allowed to occur in the presence of these nucleases. Lane 1 depicts the gel migration pattern of all fragments present after

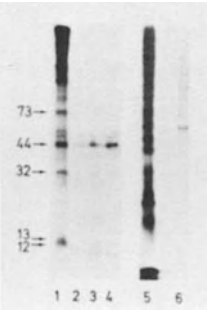

Fig. 10. Effect of eIF-2 on nuclease sensitivity of STNV RNA. Purified, intact STNV RNA was digested with T_1 (lanes 1-4) or P_1 ribonuclease (lanes 5 and 6) for 42 min at 0°C. For control (lanes 1 and 5), an aliquot of each digest was analyzed on a 12% polyacrylamide gel in 7 M urea. To other aliquots, increasing amounts of eIF-2 were added after 30 min of digestion, and after a further 12 min at 0°C the material retained on nitrocellulose filters was isolated; the filters were treated immediately with buffer containing 0.1% $NaDodSO_4$ and 0.5% diethylpyrocarbonate before extraction and analysis of RNA. Reproduced with permission (16).

digestion with ribonuclease T_1 for 42 min. When eIF-2 in increasing amounts was present during the last 12 min of digestion and the RNA bound to eIF-2 was isolated, it yielded the patterns shown in lanes 2-4. The major species recovered was 44 nucleotides long, while intact RNA was absent, in spite of the fact that it was the predominant species in the digest without eIF-2 (lane 1). This result was surprising because, as we have seen in Fig. 7, eIF-2 binds far more readily to intact STNV RNA than to the 44-nucleotide 5'-terminal fragment. eIF-2 itself is devoid of detectable nuclease activity (16). Hence, one would have predicted that intact RNA should predominate in lanes 2-4, rather than the 44-nucleotide fragment that is actually observed. The most plausible explanation of this discrepancy is that binding of eIF-2 to intact STNV RNA induces a conformational change in this RNA that greatly facilitates cleavage by ribonuclease T_1 at position 44.

Binding of eIF-2 to intact STNV RNA likewise increases the sensitivity of the RNA to cleavage by ribonuclease P_1 near position 60 (Fig. 10, lanes 5 and 6), attesting to a conformational change induced also at that point by the binding of the initiation factor molecule.

These results show that the binding site for eIF-2 extends in the 3'-terminal direction to a point between nucleotides 32 and 44 (see Fig. 8). On the 5'-terminal side, the binding of eIF-2 protects positions 11, 12,

23 and 32 against digestion (Fig. 10), placing the boundary at or before position 10. Since the G residues at positions 2 and 7 are hydrogen-bonded and thus resistant to nuclease attack (38), it is not certain if the eIF-2 binding site extends to the physical 5' end. The striking aspect of the eIF-2 binding site is, however, that it overlaps virtually completely with the binding site for 40 S ribosomal subunits (nucleotides 10-47) established in wheat germ extract (39) and depicted by the line in Fig. 8. Thus, eIF-2 by itself recognizes and binds virtually the same nucleotide sequence that is bound by 40 S ribosomal subunits carrying eIF-2, Met-tRNA$_f$ and all other components needed for initiation of translation. This finding extends and strongly reinforces the results with Mengovirus RNA and points to a critical role for eIF-2 in the recognition of mRNA by ribosomes.

VIRAL AND CELLULAR mRNA COMPETE FOR eIF-2 IN DIRECT BINDING AS THEY DO IN TRANSLATION

To study the functional implications of the interaction between mRNA and eIF-2, we analyzed translational competition between different species of mRNA, choosing to work with the mRNA-dependent reticulocyte lysate because it allows the precise quantitation of each mRNA species present during translation and is capable of repeated initiation with an efficiency approaching that of the intact cell (40, 41). The addition of increasing amounts of rabbit globin mRNA to this system generates conditions of increasing mRNA competition pressure that considerably magnifies even small differences in competing ability between individual mRNA species. Such conditions lead to selective synthesis of β-globin, at the expense of translation of α-globin mRNA (42, 43). Addition of an excess of eIF-2 restored the α/β globin synthetic ratio to the molar ratio of their respective mRNAs in the reaction, yet did not change the overall number of initiations of translation (42). Hence, eIF-2 relieves the competition between α- and β-globin mRNA during translation. Moreover, addition of eIF-2 relieved the selective inhibition of translation of α-globin mRNA by monovalent anions (e.g., Cl^-) (42). Such ions inhibit initiation principally by affecting the binding of mRNA to 40 S subunits carrying Met-tRNA$_f$ (44). The relieving activity of eIF-2 thus suggested that the anions may inhibit the interaction between eIF-2 and mRNA during translation. Indeed, Cl^- or OAc^- ions inhibit the direct binding of globin mRNA to eIF-2 in a manner that closely parallels their inhibition of globin

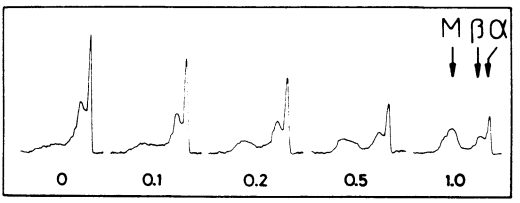

Fig. 11. Cellulose acetate electrophoresis analysis of products synthesized during simultaneous translation of globin mRNA and Mengovirus RNA. Translation mixtures contained 1.1 µg of globin mRNA and the indicated µg amounts of Mengovirus RNA. Densitometer scans of the autoradiogram of ^{35}S-labeled products are shown. α, α-globin; β, β-globin; M, Mengovirus RNA-directed products of translation. Reproduced with permission (27).

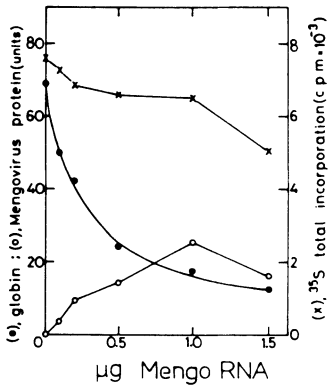

Fig. 12. Translational competition between globin mRNA and Mengovirus RNA. Areas under the curves of densitometer scans (Fig. 11) are plotted in arbitrary units as total amounts of globin (•) and of Mengovirus RNA-directed products (o). Values obtained with 1.5 µg of Mengovirus RNA are from the same experiment. Total ^{35}S-methionine incorporation into protein (X). Reproduced with permission (27).

mRNA translation (42). Furthermore, binding of isolated α-globin mRNA to eIF-2 displays greater sensitivity to anionic inhibition than does binding of β-globin mRNA (42). These results support the concept that a direct interaction between mRNA and eIF-2 occurs at initiation of translation and suggest a relationship between the affinity of an mRNA for eIF-2 on one hand, and its ability to compete in translation on the other (42, 45).

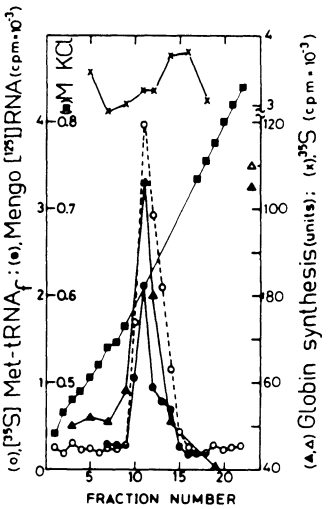

Fig. 13. Co-purification of the activity that relieves translational competition with eIF-2. For purification of eIF-2, see ref. 27. The gradient portion of the phosphocellulose column is shown. Aliquots were assayed for GTP-dependent binding of ^{35}S-labeled Met-tRNA$_f$ (o) or binding of ^{125}I-labeled Mengo RNA (●). Aliquots were added to translation mixtures containing 1 μg of globin mRNA and 0.5 μg of Mengovirus RNA. Incorporation of ^{35}S-methionine into total protein (X) and total amount of globin formed (▲) are shown. Triangles on right indicate amount of globin synthesized in reaction mixtures lacking Mengovirus RNA, incubated with (▲) and without (△) material from tube 11. Reproduced with permission (27).

More quantitative evidence for this concept came from a study of the competition between globin mRNA and Mengovirus RNA (27). In conditions where the total number of initiations remains constant, the addition of increasing amounts of Mengovirus RNA leads to a progressive decrease in globin mRNA translation, accompanied by increasing synthesis of viral protein (Fig. 11). From a plot of the integrated translation yields (Fig. 12), it can be calculated that half-maximal inhibition of globin mRNA translation occurs when 35 molecules of globin mRNA are present for every molecule of Mengovirus RNA. Assuming that equal proportions of these RNA molecules are translationally active, this means that a molecule of Mengovirus RNA competes 35-fold more strongly in translation than does (on average) a molecule of globin mRNA.

Does this competition involve eIF-2? Indeed, in conditions where globin synthesis is greatly depressed by the presence of Mengovirus RNA,

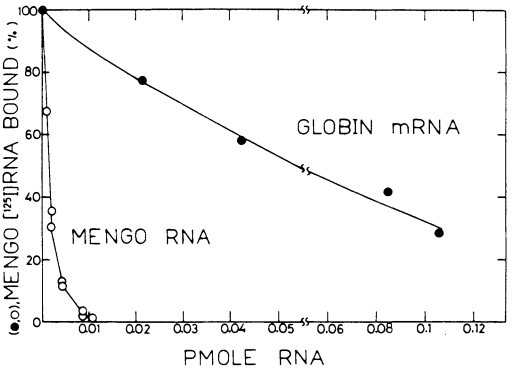

Fig. 14. Competition between globin mRNA and Mengovirus RNA in direct binding to eIF-2. See text. Reproduced with permission (27).

the addition of eIF-2 does not stimulate overall translation, yet restores globin synthesis to the level seen in the absence of competing Mengovirus RNA (Fig. 13). Globin synthesis in controls lacking Mengovirus RNA is not stimulated by the addition of eIF-2 (Fig. 13, triangles on right). Hence, addition of eIF-2 allows the more weakly competing, but more numerous, globin mRNA molecules to initiate translation at the expense of the more strongly competing, but less numerous, viral RNA molecules. Initiation factor eIF-2, therefore, relieves the mRNA competition.

The fact that eIF-2 acts to shift translation in favor of globin synthesis shows clearly that globin mRNA and Mengovirus RNA compete for eIF-2, but does not eliminate the possibility that eIF-2 could act in a non-specific manner, as by increasing the pool of 40 S/Met-tRNA$_f$ complexes. The results of RNA binding experiments, however, show that Mengovirus RNA and globin mRNA compete directly for eIF-2 with an affinity ratio that matches exactly with that observed in translation competition experiments.

In the experiment of Fig. 14, the only macromolecules present are eIF-2 and mRNA. Binding of labeled Mengovirus RNA to a limiting amount of eIF-2 is studied in the presence of increasing amounts of unlabeled, competing RNA. Unlabeled Mengovirus RNA competes as expected, with the same affinity for eIF-2 as the labeled viral RNA. By contrast, 30 times more molecules of globin mRNA must be present before binding of labeled Mengovirus RNA is reduced by one-half. The reciprocal experiment, in which the label was in globin mRNA, yielded the same result (27). Thus, a molecule

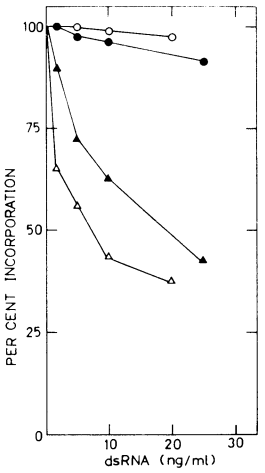

Fig. 15. Differential sensitivity of initiation of translation of Mengovirus RNA and globin mRNA to inhibition by dsRNA. Translation mixtures containing 1.5 µg of globin mRNA (Δ, ▲) or Mengovirus RNA (o, ●) were incubated for 45 min in the presence of *N*-formyl-[^{35}S]Met-tRNA$_f$ (▲, ●) or ^{35}S-methionine (Δ, o) in the presence of the indicated concentrations of P. chrysogenum dsRNA. Radioactivity in total protein was determined. Reproduced with permission (25).

of Mengovirus RNA binds to eIF-2 30-fold more strongly than (on average) a molecule of globin mRNA. Additional and independent evidence for the high affinity of Mengovirus RNA for eIF-2 is furnished by the observation (27) that binding of this RNA to eIF-2 is resistant to inhibition by monovalent anions at concentrations that completely inhibit the binding of globin mRNA. The high affinity of Mengovirus RNA for eIF-2 is not related simply to nucleotide length, for as we have seen in Fig. 1, VSV negative-strand RNA, which is even longer, binds only weakly and non-specifically to eIF-2.

VIRAL mRNA AND DOUBLE-STRANDED RNA BIND DISTINCTLY YET COMPETITIVELY TO eIF-2

Further evidence for a preferential interaction between Mengovirus RNA and eIF-2 during translation came, unexpectedly, from a study of the inhibition of translation by double-stranded RNA (dsRNA) (25). DsRNA blocks initiation of translation in reticulocyte lysates by causing the inactivation of eIF-2 (9, 46). In the presence of dsRNA, the α-subunit of eIF-2 becomes phosphorylated (46, 47). The consequent failure of Met-tRNA$_f$ to bind to the 40 S ribosomal subunit — a necessary prerequisite for the

binding of mRNA — should then be sufficient to explain why initiation of protein synthesis becomes blocked. This view would predict that dsRNA should equally inhibit the translation of all mRNA species, as each depends on prior binding of Met-tRNA$_f$.

We were surprised to discover that this is not the case (25). As seen in Fig. 15 (open triangles), the translation of globin mRNA in an mRNA-dependent reticulocyte lysate is already inhibited by as little as 1 ng/ml of dsRNA, and maximal inhibition is observed at 20 ng/ml. The residual synthesis of protein at that concentration is due to the fact that inhibition by dsRNA is manifested only after a 5-10 min lag period of synthesis at control rates (25). By contrast, the translation of Mengovirus RNA continues virtually unabated under these conditions (Fig. 15, open circles). This continued translation of Mengovirus RNA, we could show, is totally dependent on continued initiation of translation (25). In the experiment of Fig. 15, the number of initiation events was measured directly by the use of *N*-formyl-[^{35}S]Met-tRNA$_f$ as donor of labeled methionine (filled symbols). It is seen that the number of initiation events directed by Mengovirus RNA is not significantly affected at dsRNA concentrations that severely inhibit incorporation directed by globin mRNA, and that transfer of label from *N*-formyl-Met-tRNA$_f$ into protein closely reflects the incorporation of labeled methionine. Total cellular mRNA behaves as globin mRNA in terms of its sensitivity to dsRNA (25). Thus, in contrast to cellular mRNA, initiation of translation of Mengovirus RNA is resistant to inhibition by dsRNA.

What is the mechanism of this unexpected resistance? We could show that translation of Mengovirus RNA was not maintained by putative traces of eIF-2 left over after the bulk of this factor had been inactivated by dsRNA, for as soon as globin mRNA translation had become inhibited, the late addition of Mengovirus RNA completely failed to elicit translation (25). This meant that Mengovirus RNA was somehow able to prevent the establishment of translational inhibition, yet in principle was just as sensitive as globin mRNA to inhibition once established. We then discovered that Mengovirus RNA is able to inhibit competitively the dsRNA-dependent phosphorylation of eIF-2, while globin mRNA does not exhibit this property (25). Since we had already found in 1974 that eIF-2 binds to dsRNA (9), these results suggested that mRNA and dsRNA may directly compete for eIF-2.

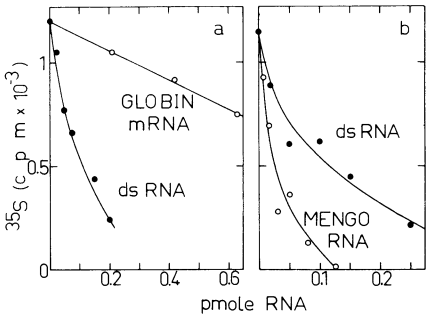

Fig. 16. Inhibitory effect of dsRNA, globin mRNA, and Mengovirus RNA on binding of Met-tRNA$_f$ to eIF-2. The assay contained constant amounts of ^{35}S-Met-tRNA$_f$ and purified eIF-2, and the indicated amounts of dsRNA, globin mRNA, or Mengovirus RNA. Control without eIF-2 [100 cpm in (a) and 80 cpm in (b)] was subtracted. Reproduced with permission (25).

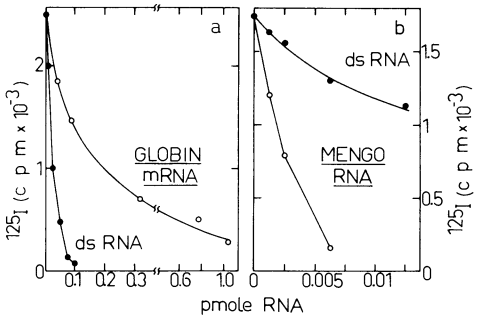

Fig. 17. Direct binding competition of dsRNA, globin mRNA, and Mengovirus RNA for eIF-2. RNA-binding assays in (a) contained ^{125}I-labeled globin mRNA (0.025 pmol;9,500 cpm) and 10.6 ng of eIF-2. In (b), they contained ^{125}I-labeled Mengovirus RNA (0.002 pmol; 4,470 cpm) and 0.28 ng of eIF-2. Unlabeled dsRNA, globin mRNA, or Mengovirus RNA was present in the indicated amounts. Control without eIF-2 [110 cpm in (a), 70 cpm in (b)] was subtracted. Reproduced with permission (25).

Figures 16 and 17 show that this is indeed the case. In Fig. 16, the ability of mRNA to inhibit competitively the binding of Met-tRNA$_f$ to eIF-2, already mentioned, is used to analyze the eIF-2/mRNA interaction. It is seen that dsRNA also exhibits this activity. On a molar basis, dsRNA is a more powerful inhibitor of eIF-2/Met-tRNA$_f$ binding than is globin mRNA (a),

but a weaker inhibitor than Mengovirus RNA (b). The experiment shows that eIF-2 binds dsRNA more strongly than globin mRNA, but more weakly than Mengovirus RNA. That mRNA and dsRNA compete directly for eIF-2 is shown in Fig. 17. Here, a limiting amount of eIF-2 was incubated with labeled globin mRNA (a) or Mengovirus RNA (b) and increasing amounts of unlabeled, competing RNA. When dsRNA is used as competing RNA, it is seen to bind to eIF-2 with about 6-fold higher affinity than globin mRNA (a). By contrast, a molecule of dsRNA binds to eIF-2 about 7 times more weakly than Mengovirus RNA (b). This and other experiments (25) led to a model in which binding of eIF-2 to dsRNA leads to the eventual inactivation of this factor and inhibition of translation, while binding of eIF-2 to Mengovirus RNA leads to initiation and a new round of protein synthesis.

Even though dsRNA and mRNA compete in their binding to eIF-2, the structural features recognized by eIF-2 in these RNA species are distinct. The nature of the interaction between dsRNA and eIF-2 was studied by analyzing the binding of eIF-2 to Penicillium chrysogenum dsRNA molecules labeled with ^{32}P at their 5' ends, and the ability of eIF-2 to protect such label against pancreatic ribonuclease digestion (48). The results revealed binding sites for eIF-2 at the 5' ends as well as throughout internal regions of the dsRNA molecule. Binding of eIF-2 is not restricted to specific sequences in dsRNA (12, 48). Our data support the interpretation that eIF-2 recognizes the A conformation in dsRNA, rather than sequence. Apparently, binding of eIF-2 at sites spaced 200 bp apart prevents relaxation of the intervening length of double helix, thereby stabilizing the dsRNA molecule against ribonuclease attack (48).

THE HIGH AFFINITY OF VIRAL MESSENGER RNA FOR eIF-2

An outstanding feature of translational control that remains to be clarified is the very wide range in initiation efficiency encountered among eukaryotic mRNA species. That this range is at least one hundredfold follows from the observations that Mengovirus RNA competes some 35-fold more effectively at initiation of translation than globin mRNA (Figs. 11 and 12), while albumin mRNA from liver, a weaker template than globin mRNA, competes about 4-fold better than haemopexin mRNA (49). Thus, the one hundredfold range may well be a conservative estimate.

A strong clue to this question is provided by our finding, reported here, that the ability of an mRNA species to compete in translation can

be correlated directly with its binding affinity for eIF-2. This is demonstrated most strikingly by the 30-fold higher affinity of Mengovirus RNA for eIF-2, as compared to globin mRNA (Fig. 14), and the 35-fold greater ability of a molecule of Mengovirus RNA to compete in translation (Fig. 12). Although similarly quantitative data are not available for α- and β-globin mRNA, we observed that the markedly greater ability of β-globin mRNA to compete in translation is coupled with a significantly higher affinity for eIF-2 (42; see above). Indeed, globin mRNA possesses a 10- to 15-fold greater affinity for eIF-2 than does phage R17 RNA (13), a ratio that fits remarkably well with the observation that this RNA is about 6% as efficient as globin mRNA as template in a reticulocyte cell-free system for translation (50). Studies with dsRNA (Figs. 15-17) provide independent evidence for a direct and preferential interaction between eIF-2 and Mengovirus RNA during translation. Finally, different monovalent anions inhibit the direct binding of globin mRNA to eIF-2 with characteristics that match virtually perfectly with those of their inhibition of globin mRNA translation in a whole lysate (42). Together, these results strongly support the concept that a direct interaction of mRNA with eIF-2 occurs during translation and is of key importance to translational control.

Although other initiation factors, notably eIF-4A and eIF-4F (51) have been shown to interact with mRNA, especially at the cap structure (51), and to influence translation rates (52, 53), their ability to act as targets of translational competition remains to be demonstrated conclusively (1). Moreover, sequence analysis of the type reported here for the binding site for eIF-2 in mRNA has not yet been reported for any other eukaryotic initiation factor, nor has specific interaction with an mRNA sequence, other than the cap, so far been demonstrated. As we have suggested elsewhere (1, 45), eIF-4A (8, 52) and/or cap-binding proteins (54, 55) may unwind secondary structure at the 5' end of mRNA and anchor it at the cap, while eIF-2 then acts to recognize and bind the mRNA sequence and conformation existing at the ribosome binding site. The conformational change induced around this site by binding of eIF-2 (see Fig. 10), moreover, may cause unfolding of the mRNA molecule, creating thereby space to accommodate the attached Met-tRNA$_f$/40 S ribosomal subunit.

The role of the 5'-leader sequence in determining initiation efficiency is difficult to determine on the basis of presently available data. This point is well illustrated by the five mRNA species of VSV. In vivo, these

mRNAs are apparently translated with identical efficiencies of initiation (56), yet they possess 5' leaders that are unrelated in sequence and vary in length from 10 to 41 nucleotides (57).

This variation in 5' ends in the face of identical initiation efficiency suggests that it may not be the leader sequence per se that determines initiation strength, but the interaction of this sequence with other, internal parts of the mRNA molecule. Since the coding sequence in an mRNA molecule is dictated by the individual protein encoded, and since the 3'-untranslated sequence is also highly variable (1), any stable interaction between 5' leader and internal sequences is only possible if the leader sequence is especially tailored to allow such a fit. If this view is correct, then an important determinant for the efficiency of initiation is the structure generated by this interaction around the initiation codon.

The fact that denaturation of eukaryotic mRNA does not lead to binding of ribosomes at internal sites in mRNA (58-60) is not in conflict with this concept. It merely indicates that structure in mRNA is not important for determining where initiation occurs. This point should be separated clearly from the question, how often initiation takes place. Two mRNA species can vary widely in initiation efficiency, even if, as is generally the case, initiation occurs at the first AUG codon in both. It is here that the contribution of structure may be of essence.

The absence of 5'-proximal sequences in both eIF-2 and ribosome-protected segments of Mengovirus RNA (Fig. 3) supports the concept that the RNA of this virus has evolved a highly efficient mechanism of initiation that bypasses the need for either a 5' end or a 5'-terminal cap structure. Apparently, this is because a structured site, highly favorable for initiation and possessing a very high affinity for eIF-2, is generated within this RNA molecule (Figs. 2, 3, 11-14). Indeed, the related encephalomyocarditis virus RNA possesses an unusual primary and secondary structure at the start of the coding region, and this highly structured site is resistant to nuclease treatment (61, 62). The existence of a structure possessing very high affinity for eIF-2 (and possibly for other components of initiation) would obviate the need for the additional stabilization imparted by binding at the cap. Clearly, the high affinity of Mengovirus RNA for eIF-2 could not be due to double-stranded RNA features, as Mengovirus RNA binds more strongly than authentic dsRNA to eIF-2 (Figs. 16, 17).

Yet, in STNV RNA, which is also an efficiently competing mRNA species lacking a 5'-terminal cap (33, 34, 36, 37), the site that binds eIF-2 and ribosomes lies close to the 5' end (Fig. 8). High efficiency of initiation may in this case result from a suitable structure at the initiation site, coupled with proximity to the 5' end. As the affinity properties of the structure generated around the initiation codon become less favorable, proximity to the 5' end may start to provide a contribution. For even less favorable initiation sites, the additional contribution of the 5'-terminal cap becomes increasingly important, as evidenced for all cellular and many viral mRNA species. Apparently, eIF-2 recognizes in different mRNA species a common conformation existing around the initiation codon, but differing in subtle ways that are important for determining individual binding affinity.

The accessibility of the AUG initiation codon to factors and ribosomes may well be influenced by secondary structure. Diminished accessibility has been suggested for α-globin mRNA, as compared to β-globin mRNA (63), but the generality of this concept as a determinant of initiation efficiency remains to be documented. Since eIF-2 is able to unfold the 5'-terminal conformation of the STNV RNA molecule around the ribosome binding site (Figs. 8, 10), it may in fact act to make the AUG initiation codon more accessible to the Met-$tRNA_f$ molecule lodged on the attached 40 S ribosomal subunit.

I thank Mrs. E. Herskovics for typing the manuscript. Research of the author was supported by the National Academy of Sciences of Israel, the Gesellschaft fuer Strahlen- und Umweltforschung (Muenchen), and the National Council for Research and Development of Israel.

REFERENCES

1. Kaempfer, R. Regulation of eukaryotic translation. In: Comprehensive Virology. Vol. XIX (Eds. H. Fraenkel-Conrat and R.R. Wagner), Plenum Publishing Corp., New York, 1984 (in press).
2. Kaempfer, R., Rosen, H., Di Segni, G. and Knoller, S. Structural feature of picornavirus RNA involved in pathogenesis: A very high affinity binding site for a messenger RNA-recognizing protein. In: Developments in Molecular Virology. Vol. 4 "Mechanisms of Viral Pathogenesis" (Eds. A. Kohn and P. Fuchs), Martinus Nijhoff Publishing, The Hague-Boston, 1984, pp. 180-200.
3. Schreier, M.H. and Staehelin, T. Initiation of eukaryotic protein synthesis: (Met-$tRNA_f$.40 S ribosome) initiation complex catalysed by purified initiation factors in the absence of mRNA. Nature New Biol. 242:35-38, 1973.

4. Darnbrough, C.H., Legon, S., Hunt, T., and Jackson, R.J. Initiation of protein synthesis: evidence for messenger RNA independent binding of methionyl-transfer RNA to the 40 S ribosomal subunit. J. Mol. Biol. 76:379-403, 1973.
5. Chen, Y., Woodley, C., Bose, K. and Gupta, N.K. Protein synthesis in rabbit reticulocytes: characteristics of a Met-tRNA$_f^{Met}$ binding factor. Biochem. Biophys. Res. Commun. 48:1-9, 1972.
6. Levin, D.H., Kyner, D. and Acs, G. Protein initiation in eukaryotes: formation and function of a ternary complex composed of a partially purified ribosomal factor, methionyl transfer RNA, and guanosine triphosphate. Proc. Natl. Acad. Sci. USA 70:41-45, 1973.
7. Trachsel, H., Erni, B., Schreier, M. and Staehelin, T. Initiation of mammalian protein synthesis. The assembly of the initiation complex with purified initiation factors. J. Mol. Biol. 116:755-767, 1977.
8. Grifo, J.A., Tahara, S.M., Leis, J.P., Morgan, M.A., Shatkin, A.J. and Merrick, W.C. Characterization of eIF-4A, a protein involved in ATP-dependent binding of globin mRNA. J. Biol. Chem. 257:5246-5252, 1982.
9. Kaempfer, R. Identification and RNA binding properties of an initiation factor capable of relieving translational inhibition induced by heme deprivation or double-stranded RNA. Biochem. Biophys. Res. Commun. 61:591-597, 1974.
10. Barrieux, A. and Rosenfeld, M.G. Characterization of GTP-dependent Met-tRNA$_f$ binding protein. J. Biol. Chem. 252:3843-3847, 1977.
11. Barrieux, A. and Rosenfeld, M.G. mRNA-induced dissociation of initiation factor 2. J. Biol. Chem. 253:6311-6315, 1978.
12. Kaempfer, R., Hollender, R., Abrams, W.R. and Israeli, R. Specific binding of messenger RNA and methionyl-tRNA$_f^{Met}$ by the same initiation factor for eukaryotic protein synthesis. Proc. Natl. Acad. Sci. USA 75:209-213, 1978.
13. Kaempfer, R., Hollender, R., Soreq, H. and Nudel, U. Recognition of messenger RNA in eukaryotic protein synthesis: Equilibrium studies of the interaction between messenger RNA and the initiation factor that binds methionyl-tRNA$_f$. Eur. J. Biochem. 94:591-600, 1979.
14. Chaudhuri, A., Stringer, E.A., Valenzuela, D. and Maitra, U. Characterization of eIF-2 containing two polypeptide chains of M_r= 48,000 and 38,000. J. Biol. Chem. 256:3988-3994, 1981.
15. Rosen, H. and Kaempfer, R. Mutually exclusive binding of messenger RNA and initiator methionyl transfer RNA to eukaryotic initiation factor 2. Biochem. Biophys. Res. Commun. 91:449-455, 1979.
16. Kaempfer, R., Van Emmelo, J. and Fiers, W. Specific binding of eukaryotic initiation factor 2 to satellite tobacco necrosis virus RNA at a 5'-terminal sequence comprising the ribosome binding site. Proc. Natl. Acad. Sci. USA 78:1542-1546, 1981.
17. Kaempfer, R. Binding of messenger RNA in initiation of eukaryotic translation. Methods Enzymol. 60:380-392, 1979.
18. Riggs, A., Suzuki, M. and Bourgeois, S. lac repressor-operator interaction. I. Equilibrium studies. J. Mol. Biol. 48:67-83, 1970.
19. Huang, A., Baltimore, D. and Stampfer, M. Ribonucleic acid synthesis of vesicular stomatitis virus. III. Multiple complementary messenger RNA molecules. Virology 42:946-957, 1970.
20. Rose, J. and Knipe, D. Nucleotide sequence complexities, molecular weights, and poly (A) content of the vesicular stomatitis virus mRNA species. J. Virol. 15:994-1003, 1975.
21. Lin, S. and Riggs, A.D. A comparison of lac repressor binding to operator and non-operator DNA. Biochem. Biophys. Res. Commun. 62: 704-710, 1975.

22. Von Hippel, P., Revzin, A., Gross, C.A. and Wang, A.C. Non-specific DNA binding of genome regulating proteins as biological control mechanism. The lac operon: equilibrium aspects. Proc. Natl. Acad. Sci. USA 71:4808-4812, 1974.
23. Lloyd, M.A., Osbourne, J.C., Safer, B., Powell, G.M. and Merrick, W.C. Characteristics of eukaryotic initiation factor 2 and its subunits. J. Biol. Chem. 255:1189-1193,1980.
24. Kaempfer, R., Rosen, H. and Israeli, R. Translational control: recognition of the 5' end and an internal sequence in eukaryotic mRNA by the initiation factor that binds methionyl-tRNA$_f^{Met}$. Proc. Natl. Acad. Sci. USA 75:650-654, 1978.
25. Rosen, H., Knoller, S. and Kaempfer, R. Messenger RNA specificity in the inhibition of eukaryotic translation by double-stranded RNA. Biochemistry 20:3011-3020,1981.
26. Kaempfer R. Purification of initiation factor eIF-2 by RNA-affinity chromatography. In: Methods in Enzymology, Vol. LX, Part G. (Eds. L. Grossman and K. Moldave),Academic Press, New York, 1979, pp.247-255.
27. Rosen, H., Di Segni, G. and Kaempfer, R. Translational control by messenger RNA competition for eukaryotic initiation factor 2. J. Biol. Chem. 257:946-952, 1982.
28. Perez-Bercoff, R. and Gander, M. The genomic RNA of mengovirus. I. Location of the poly (C) tract. Virology 80:426-429, 1977.
29. Sangar, D.V., Black, D.N., Rowlands, D.J., Harris, T.J.R. and Brown, F. Location of the initiation site for protein synthesis on foot-and-mouth disease virus RNA by in vitro translation of defined fragments of the RNA. J. Virol. 33:59-68, 1980.
30. Perez-Bercoff, R. and Kaempfer, R. Genomic RNA of Mengovirus: Recognition of common features by ribosomes and eukaryotic initiation factor 2. J. Virol. 41:30-41, 1982.
31. Perez-Bercoff, R. and Gander, M. In vitro translation of mengovirus RNA deprived of its terminally-linked ("capping"?) protein. FEBS Lett. 96:378-386, 1978.
32. Degener, A.M., Pagnotti, P., Facchini, J. and Perez-Bercoff, R. Genomic RNA of mengovirus. Translation of its two cistrons in lysates of interferon-treated cells. J. Virol. 45:889-894 , 1983.
33. Leung D.W., Gilbert, C.W., Smith, R.E., Sasavage, N.L. and Clark, J.M., Jr. Translation of satellite tobacco necrosis virus ribonucleic acid by an in vitro system from wheat germ. Biochemistry 15:4943-4950, 1976.
34. Herson, D., Schmidt, A., Seal, S., Marcus, A., and van Vloten-Doting, L. Competitive mRNA translation in an in vitro system from wheat germ. J. Biol. Chem. 254:8245-8249, 1979.
35. Ysebaert, M., van Emmelo, J. and Fiers, W. Total nucleotide sequence of a nearly full-size DNA copy of satellite tobacco necrosis virus RNA. J. Mol. Biol. 143:273-287, 1980.
36. Wimmer, E., Chang, A.Y., Clark, J.M. Jr., and Reichmann, M.E. Sequence studies of satellite tobacco necrosis virus RNA: isolation and characterization of a 5'-terminal trinucleotide. J. Mol. Biol. 38:59-73, 1968.
37. Horst, H., Fraenkel-Conrat, H., and Mandeles, S. Terminal heterogeneity at both ends of the satellite tobacco necrosis virus ribonucleic acid. Biochemistry 10:4748-4752, 1971.
38. Leung, D.W., Browning, D.S., Heckmann, J.E., RajBhandary, U.L. and Clark, J.M. Jr. Nucleotide sequence of the 5' terminus of satellite tobacco necrosis virus ribonucleic acid. Biochemistry 18:1361-1366, 1979.

39. Browning, K.S., Leung, D.W., and Clark, J.M. Jr. Protection of satellite tobacco necrosis virus ribonucleic acid by wheat germ 40S and 80 S ribosomes. Biochemistry 19:2276-2282, 1980.
40. Pelham, H.R.B. and Jackson, R.J. An efficient mRNA-dependent translation system from reticulocyte lysates. Eur. J. Biochem. 67:247-256, 1976.
41. Jackson, R.J. The cytoplasmic control of protein synthesis. In: Protein Biosynthesis in Eukaryotes (Ed. R. Perez-Bercoff), Plenum Publishing Corp., New York, 1982, pp. 363-418.
42. Di Segni, G., Rosen, H. and Kaempfer, R. Competition between α- and β-globin messenger ribonucleic acids for eukaryotic initiation factor 2. Biochemistry 18:2847-2854, 1979.
43. Lodish, H.F. Alpha and beta globin mRNA: Different amounts and rates of initiation of translation. J. Biol. Chem. 246:7131-7138, 1971.
44. Weber, L.A., Hickey, E.D., Maroney, P.A. and Baglioni, C. Inhibition of protein synthesis by Cl^-. J. Biol. Chem. 252:4007-4010, 1977.
45. Kaempfer, R., Rosen, H. and Di Segni, G. A direct correlation between the affinity of a given mRNA for eukaryotic initiation factor 2 and its ability to compete in translation. In: Mechanisms of Protein Synthesis (Ed. E. Bermek), Springer Verlag, Heidelberg, 1984 (in press).
46. Clemens, M.J., Safer, B., Merrick, W.C., Andersen, W.F. and London, I.M. Inhibition of protein synthesis in rabbit reticulocyte lysates by double stranded RNA and oxidized glutathione: indirect mode of action on polypeptide chain initiation. Proc. Natl. Acad. Sci. USA 72:1286-1290, 1975.
47. Farrell, P.J., Balkow, K., Hunt, T., Jackson, R.J. and Trachsel, H. Phosphorylation of initiation factor eIF-2 and the control of reticulocyte protein synthesis. Cell 11:187-200,1977.
48. Itamar, D., Gonsky, R., Lebendiker, M. and Kaempfer, R. The nature of the interaction of eukaryotic initiation factor 2 with double-stranded RNA. Submitted for publication.
49. Kaempfer, R. and Konijn, A.M. Translational competition by mRNA species encoding albumin, ferritin, haemopexin and globin. Eur. J. Biochem. 131:545-550, 1983.
50. Schreier, M., Staehelin, T., Gesteland, R. and Spahr, P. Translation of bacteriophage R17 and Qβ RNA in a mammalian cell-free system. J. Mol. Biol. 75:575-578, 1973.
51. Grifo, J.A., Tahara, S.M., Morgan, M.A., Shatkin, A.J. and Merrick, W.C. New initiation factor activity required for globin mRNA translation. J. Biol. Chem. 258:5804-5810, 1983.
52. Ray, B.K., Brendler, T.G., Adya, S., Daniels-McQueen, S., Miller, J.K., Hershey, J.W.B., Grifo, J.A., Merrick, W.C. and Thach, R.E. Role of mRNA competition in regulating translation: Further characterization of mRNA discriminatory factors. Proc. Natl. Acad. Sci. USA 80:663-667, 1983.
53. Parets-Soler, A., Reibel, L. and Schapira, G. Differential stimulation of α- and β-globin mRNA translation by M_r 50,000 and 28,000 polypeptide-containing fractions isolated from reticulocyte polysomes. FEBS Lett. 136:259-264, 1981.
54. Sonenberg, N. ATP/Mg^{2+}-dependent cross-linking of cap binding proteins to the 5' end of eukaryotic mRNA. Nucleic Acids Res. 9:1643-1656,1981.
55. Lee, K.A.W., Guertin, D. and Sonenberg, N. mRNA secondary structure as a determinant in cap recognition and initiation complex formation. J. Biol. Chem. 258:707-710, 1983.

56. Villareal, L.P., Breindl, M. and Holland, J.J. Determination of molar ratios of vesicular stomatitis virus induced RNA species in BHK_{21} cells. Biochemistry 15:1663-1667, 1976.
57. Rose, J.K. Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. Cell 19:415-421, 1980.
58. Kozak, M. Influence of mRNA secondary structure on binding and migration of 40 S ribosomal subunits. Cell 19:79-90, 1980.
59. Collins, P., Fuller, F., Marcus, P., Hightower, L. and Ball, L.A. Synthesis and processing of Sindbis virus nonstructural proteins in vitro. Virology 118:363-379, 1982.
60. Zagorska, L., Chroboczek, J., Klita, S. and Szafranski, P. Effect of secondary structure of mRNA on the formation of initiation complexes with prokaryotic and eukaryotic ribosomes. Eur. J. Biochem. 122:265-269, 1982.
61. Smith, A.E. Control of translation of animal virus messenger RNA. In: Control Processes in Virus Multiplication (Eds. D.C. Burke and W.C. Russell),25th Symp. Soc. Gen. Microbiol., Cambridge Univ. Press, Cambridge, pp. 183-223, 1975.
62. Porter, A.G., Frisby, D.P., Carey, N.H. and Fellner, P. Nucleotide sequence studies on picornavirus RNAs. In: In Vitro Transcription and Translation of Viral Genomes (Eds. A.L. Haenni and G. Beaud), INSERM, Paris, pp. 169-176, 1975.
63. Pavlakis, G.N., Lockard, R.E., Vamvakopoulos, N., Rieser, L., RajBhandary, V.L. and Vournakis, J.N. Secondary structure of mouse and rabbit α- and β-globin mRNAs. Differential accessibility of α and β initiator AUG codons towards nucleases. Cell 19:91-102, 1980.

Phage mRNA

3

BACTERIOPHAGE T7 mRNAs

F. WILLIAM STUDIER

During infection, T7 DNA is transcribed first by the host RNA polymerase and then by newly made T7 RNA polymerase. T7 RNA polymerase has a stringent specificity for its own promoters, which are not found in host DNA. Therefore, production of the new T7 RNA polymerase, together with inactivation of the host RNA polymerase, switches all transcription from host DNA to T7 DNA. An initial wave of transcription proceeds down T7 DNA from left to right, taking about 40% of the latent period to reach the right end of the DNA. This wave may be coupled to entry of the DNA into the cell, and the mode of entry may be an important factor in controlling gene expression. Primary transcripts are cut at specific sites by a host enzyme, RNase III, to generate the mRNAs observed in the cell. The RNase III cleavages leave relatively stable base-paired structures at the 3' end of most T7 mRNAs, which may be at least partly responsible for the unusual stability of T7 mRNAs relative to typical host mRNAs. Differences in translational efficiency among the different T7 mRNAs are also important in regulating gene expression during infection.

INTRODUCTION

T7 is a virulent bacteriophage that infects Escherichia coli. It rapidly takes over the metabolism of the host cell and produces as many as 250 new phage particles in as little as 13 min at 37 °C. In order to generate high levels of T7 gene products in such a short time, the T7 mRNAs must be efficiently synthesized and translated. This review summarizes our current understanding of the strategies T7 employs to accomplish this.

Y. Becker (ed.), VIRAL MESSENGER RNA.

GENETIC ORGANIZATION OF T7 DNA

T7 contains a single molecule of linear, double-stranded DNA that is almost 40,000 base pairs long. The nucleotide sequence of T7 DNA has been completely determined, and the coding sequences for all of the known and potential T7 proteins have been located in the nucleotide sequence (1-3). Fifty genes are closely packed but essentially nonoverlapping in the DNA, and more than 90% of the nucleotide sequence is used to code for proteins. The arrangement of these 50 genes in T7 DNA is shown in Figure 1. In addition, there may be as many as five genes having a coding sequence that overlaps one of the 50 closely packed genes in a different reading frame. T7 genes are numbered according to their relative positions in the DNA, the numbers increasing from left to right; for historical reasons, the integers 1 to 19 are used, as well as various decimal numbers from 0.3 to 19.5.

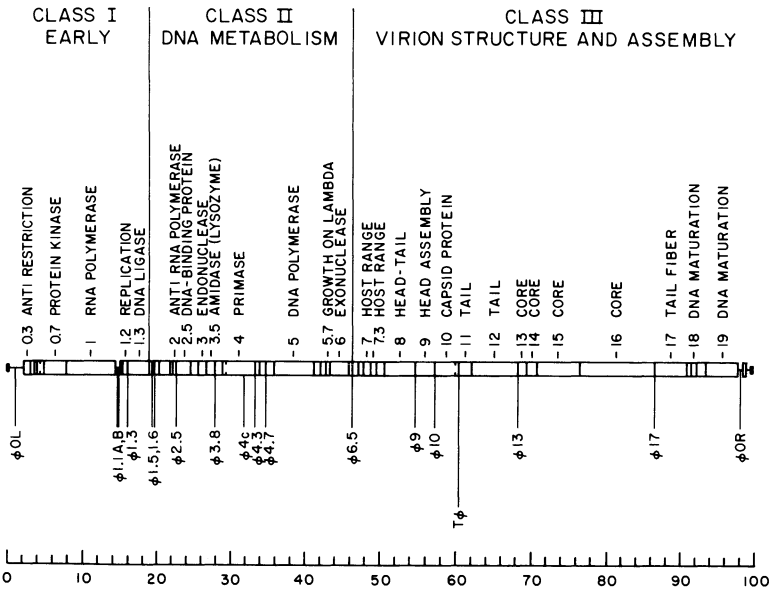

FIGURE 1. Genetic and physical map of T7 DNA (1,3). The positions of the terminal repetition (filled boxes), the T7 genes (open boxes), the promoters for T7 RNA polymerase (ø), and the terminator for T7 RNA polymerase (Tø) are drawn to scale according to their locations in the nucleotide sequence. The gene number and function are indicated for some of the genes.

T7 genes with related functions tend to cluster in the DNA, and three groups of genes that are expressed co-ordinately during infection have been recognized (4). The class I, or early genes are at the left end of the DNA and occupy almost one-fifth of the DNA. The class II, or DNA metabolizing genes come next, and occupy about one-fourth of the DNA. The class III genes, those which specify structural or assembly proteins of the phage particle, occupy the right half of the DNA. Among the 50 closely packed genes, there are nine class I genes, numbered 0.3 to 1.3; 21 class II genes, numbered 1.4 to 6.3; and 20 class III genes, numbered 6.5 to 19.5.

EXPRESSION OF T7 GENES

The T7 genes are expressed in order of their position in T7 DNA, as illustrated by the pattern of protein synthesis during infection (Figure 2). The early genes are the first to be expressed, and their synthesis shuts off along with the host proteins, or in some cases along with class II proteins. The class II, or DNA metabolizing genes, are the next to be expressed, and their synthesis shuts off near the middle of the latent period. The class III, or structural and assembly genes, are the last to be expressed, and their synthesis continues throughout the latent period.

TRANSCRIPTION OF T7 DNA

The primary control of gene expression during T7 infection is at the level of transcription. All transcription of T7 DNA is from left to right, the same direction as the order of expression of the T7 genes (4,6). T7 DNA is transcribed first by the RNA polymerase of the host cell, and then by T7 RNA polymerase, which is made in the early stages of infection (7). The overall transcription of T7 DNA is summarized in Figure 3.

Transcription by the host RNA polymerase

T7 DNA contains three strong promoters for *E. coli* RNA polymerase, located about 125 base pairs apart near the left end of the DNA (8,9). These promoters, designated A1, A2 and A3, appear to be strong relative to most promoters in host DNA, presumably so that the entering T7 DNA will be recognized and transcribed actively by the host RNA polymerase.

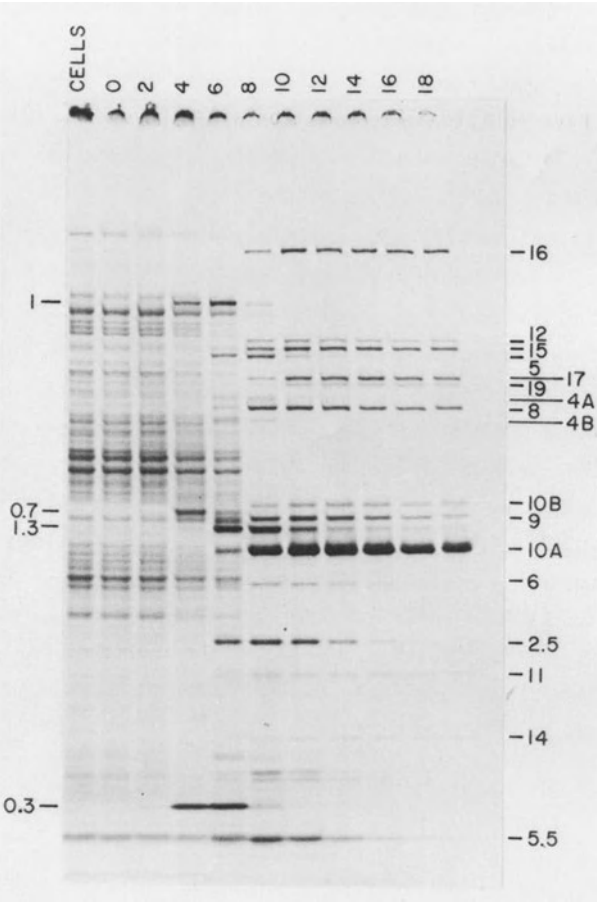

FIGURE 2. Time course of protein synthesis during T7 infection (1,3,4). A culture of E. coli growing in minimal medium at 30°C was infected with about 15 infective T7 particles per cell, and samples were pulsed for 2 min with [^{35}S]methionine immediately before and at 2-min intervals after infection. The cells were collected by centrifugation, lysed in buffer containing sodium dodecyl sulfate, and subjected to electrophoresis on a 10-20% polyacrylamide gradient gel having a 5% stacking gel, followed by autoradiography, essentially as described (5). The origin of electrophoresis is at the top of the patterns. The time at the beginning of each 2-min pulse is given above each lane; lysis of the culture would normally begin about 25 min after infection under these conditions. The gene numbers of prominent T7 proteins are indicated to the sides of the patterns, early proteins to the left, late proteins to the right (4).

Most transcripts initiated at A1, A2 and A3 appear to terminate at a relatively strong but not completely efficient transcription termination site, designated TE, at the end of the early region (4,10). A number of additional, weaker promoters for E. coli RNA polymerase have been identified, as have additional termination sites beyond TE (9), but no significant role for these additional promoters and terminators has yet been identified.

Transcription by T7 RNA polymerase

Gene 1, which lies in the middle of the early region, specifies the T7 RNA polymerase. A single-chain enzyme of molecular weight about 100,000, T7 RNA polymerase transcribes T7 DNA very actively but is essentially inert on host DNA (7). The basis for this great specificity of T7 RNA polymerase for its own DNA is the interaction between T7 RNA polymerase and its promoters. Seventeen such promoters are distributed across the T7 DNA molecule (Figures 1 and 3). Six of these promoters have identical nucleotide sequences for 23 continuous base pairs, positions -17 to +6 relative to the first nucleotide of the RNA made from the promoter (1, and references therein). The remaining 11 promoters have sequences that are identical to these sequences in 16 to 21 of the 23 base pairs. T7 RNA polymerase apparently interacts with a promoter sequence large enough that it is unlikely to be found by chance in any foreign DNA.

The øOL promoter, near the left end of T7 DNA, is not known to transcribe any mRNAs during infection, but is thought to be associated with an origin of replication (11). The ø1.1A and B promoters, just to the right of gene 1, and the øOR promoter, near the right end of the DNA, are also associated with origins of replication, but they direct the transcription of mRNAs as well (1,12). The 16 promoters that direct the transcription of mRNAs are divided into class II and class III promoters. Transcription initiated at class II promoters crosses class II genes and continues on into the class III region, but class III promoters direct the transcription of class III genes only. The six class III promoters are the strongest promoters (13,14), and are also the ones that are completely identical over 23 continuous base pairs.

T7 RNA polymerase recognizes only one transcription termination signal in T7 DNA, designated Tø, which is located just past gene 10, the gene for the major capsid protein of T7 (Figure 3). T7 RNA polymerase does not stop at TE, the transcription signal for E. coli RNA polymerase at the end of the early region, but proceeds through it into the late region (10). Termination at Tø is approximately 90% efficient (15). Completely efficient termination at Tø would be lethal to T7, since genes 11 and 12, which specify structural proteins of the T7 tail, are transcribed only by readthrough of Tø (see Figure 3). The arrangement of promoters relative to Tø may be part of a transcriptional strategy that ensures production of large amounts of mRNA for genes 9 and 10, whose proteins are needed in large amounts during infection: the 10 class II promoters and three of the class III promoters all direct transcripts toward Tø.

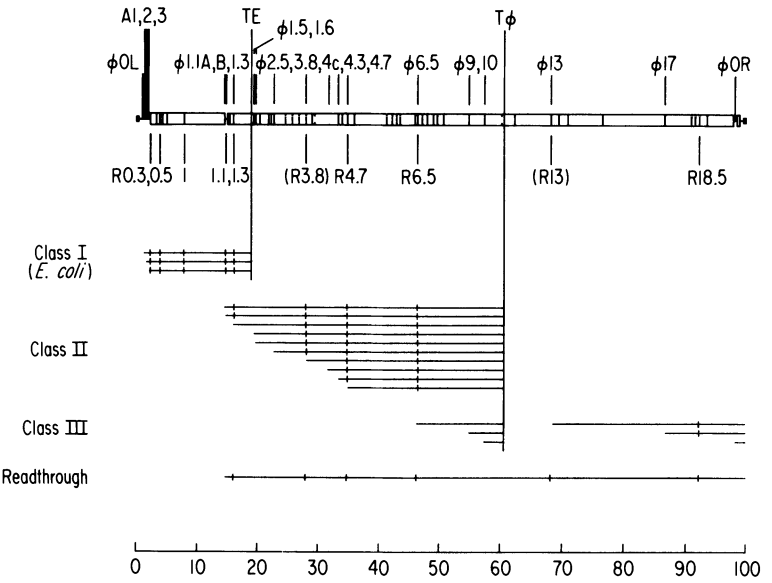

FIGURE 3. Transcription pattern of T7 DNA (1,3). The T7 genes are represented by open boxes; the positions of promoters and terminators for E. coli and T7 RNA polymerase are given above the genes, the positions of RNase III cleavage sites below. The primary transcript from each promoter is represented by a horizontal line, and sites of RNase III cleavage by short vertical lines. Not all RNAs are cut at the R3.8 and R13 RNase III cleavage sites, which is indicated by the parentheses. RNAs produced by readthrough of Tø are also represented.

Switch of transcription from host to T7 DNA

T7 DNA must be transcribed by the host RNA polymerase in order to produce T7 RNA polymerase, but once T7 RNA polymerase is made, the host RNA polymerase is completely dispensable. In fact, inactivation of the host RNA polymerase is a distinct advantage to T7 because it eliminates competition from continuing production of host mRNAs. The great selectivity of T7 RNA polymerase for its own promoters then directs all transcription in the cell to T7 DNA.

T7 specifies at least two proteins that can inactivate E. coli RNA polymerase. The first is a protein kinase, the product of gene 0.7 (5,16-18). This enzyme phosphorylates perhaps 20-40 host proteins, among them the two largest subunits of E. coli RNA polymerase (19). The action of the T7 protein kinase during infection appears by itself to be sufficient to shut off virtually all transcription due to E. coli RNA polymerase. The second protein that inactivates host RNA polymerase is the product of gene 2. A small protein of 63 amino acids, the gene 2 protein binds to E. coli RNA polymerase holoenzyme and prevents it from binding to DNA and initiating RNA chains (20). The gene 0.7 protein appears to be dispensable on the usual laboratory hosts, perhaps because the gene 2 protein can replace it. The gene 2 protein, on the other hand, appears to be needed at some later stage of infection, perhaps during the packaging process (21); it is dispensable in some hosts but not others.

The specificity of T7 RNA polymerase for its own DNA is the central controlling feature in T7 infection. It allows a simple switch of all transcription in the cell from host DNA to T7 DNA. Because it is also required for initiation of replication by the T7 replication complex, T7 RNA polymerase probably plays a similar role in switching all replication in the cell to T7 DNA (1).

TRANSCRIPTION MAY BE COUPLED TO ENTRY OF DNA INTO THE CELL

Close analysis of the kinetics of transcription of T7 DNA during infection shows that the initial transcription of both early and late regions proceeds from left to right, requiring about 40% of the latent period to reach the right end of the DNA (14,22). It is not surprising that the early region should have such a polarity, because all three

strong promoters for E. coli RNA polymerase are located near the left end of T7 DNA. However, promoters for T7 RNA polymerase are distributed across the entire molecule, and the strongest promoters are near the right end. If the entire DNA molecule were accessible at the time T7 RNA polymerase first appeared, one would not expect to see the left to right polarity that is actually observed.

One set of experiments suggests that a large part of the T7 DNA molecule may not be accessible at the time T7 RNA polymerase first appears (22). Individual promoters from T7 DNA were cloned in the plasmid pBR322, and the time during infection at which transcription begins from the cloned promoter was compared to the time at which transcription begins at the same promoter in T7 DNA itself. Promoters cloned in plasmids were found to become active as soon as T7 RNA polymerase was made, regardless of where the promoter originated in T7 DNA, whereas the same promoter in T7 DNA itself was not utilized until a time characteristic of its position in the DNA, as much as 4 min later. Apparently a promoter in a plasmid is immediately accessible to T7 RNA polymerase whereas the promoters in T7 DNA become accessible sequentially from left to right over a period of about 4 min. A simple interpretation is that what controls accessibility of the promoters in T7 DNA is a gradual entry of the DNA into the cell.

The mode of entry of the DNA thus may be an important factor in controlling transcription during infection. As mentioned above, the class II promoters for T7 RNA polymerase appear to be weaker promoters, but they are utilized before the stronger class III promoters during infection. If accessibility of the promoters is controlled by entry of the DNA, this order of utilization follows naturally. Later in infection, as the rate of transcription falls off, the remaining transcription shifts from the class II to the stronger class III promoters. This shift may in part be responsible for the shut off of class II protein synthesis midway through the latent period (Figure 2). The factors responsible for the decrease in transcription and the shift to class III promoters are not fully understood, but the T7 lysozyme, which cuts a bond in the cell wall, appears to be involved (14).

If T7 DNA enters the cell gradually, what controls the rate of entry? One possibility is that the process of transcription itself draws the DNA into the cell. An observation supporting this idea is

that T7 DNA apparently does not enter the cell when infection occurs in the presence of rifampicin, an inhibitor of E. coli RNA polymerase (23). Perhaps in a normal infection transcription by E. coli RNA polymerase draws in the early region, and transcription by the newly made T7 RNA polymerase (starting at the three promoters to the right of gene 1 within the early region) draws in the late region. Although this idea seems attractive, the possibility remains that something else controls the rate of entry of the DNA, which in turn controls the rate of movement of the initial wave of transcription down the molecule.

RNase III CLEAVAGE OF T7 mRNAs

The primary transcripts produced from T7 DNA by both the E. coli and T7 RNA polymerases are cut at specific sites by a host enzyme, RNase III, to produce the mRNAs observed in the cell (8). In the uninfected host cell, RNase III has a major role in processing ribosomal RNAs and is presumably involved in processing other RNAs as well (24). Ten RNase III cleavage sites have been identified in T7 RNAs, five in the early region and five in the late (Figure 3). Cleavage is very efficient at eight of these sites but less efficient at the remaining two. The nucleotide sequence around the eight efficient cleavage sites can be arranged in a characteristic, base-paired structure (1), illustrated in Figure 4 for the first three cleavage sites. The point of cleavage is located asymmetrically in these structures so that the upstream RNA that results from the cleavage contains a well paired stem and loop structure at its 3' end.

RNase III cleavages are a prominent feature of T7 mRNA synthesis, but it has not been clear why these cleavages are made. Most T7 proteins seem to be made with about the same efficiency whether or not the RNAs have been cut by RNase III, although there are a few exceptions (25). A possible role for the RNase III cleavages is that they might be at least partly responsible for the unusual stability of T7 mRNAs relative to typical host mRNAs (26,27): perhaps the stable stem and loop structures at the 3' end of the T7 mRNAs serve as barriers to exonucleolytic degradation of the RNAs. Relatively stable stem and loop structures also remain at the 3' end of the RNAs that terminate at TE or Tø. Preliminary experiments (Dunn and Studier, unpublished) suggest

that these structures may indeed provide stability to mRNAs both in vivo and in a cell-free protein synthesizing system from E. coli. Accumulation of stable mRNAs is apparently part of the strategy by which T7 directs ribosomes to its own mRNAs.

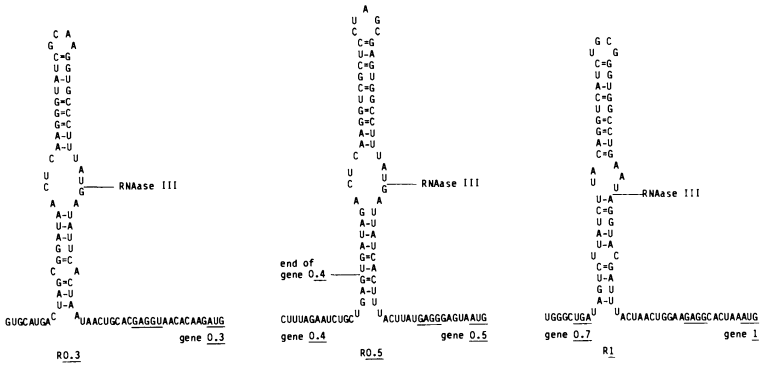

FIGURE 4. The first three RNase III cleavage sites in T7 early RNAs (1). Potential base-paired structures are indicated, as are the locations of the termination codons for the genes immediately preceding, and the ribosome-binding sequences and intitiation codons for the genes immediately following the cleavage sites. The point of RNase III cleavage is indicated at each site. Similar representations of all ten RNase III cleavage sites in T7 RNAs are given in reference 1.

TRANSLATION OF T7 mRNAs

The combination of multiple promoters, terminators and RNase III cleavage sites produces a rather large but well defined set of T7 mRNAs (Figure 3). Most of these mRNAs contain the coding sequences for more than one protein. Each coding sequence is headed by an obvious ribosome-binding and initiation site, even though in many cases part of this site is contained within the coding sequence of the preceding protein (1). Most chain-terminating mutations have no polar effects on the synthesis of T7 proteins (4), suggesting that synthesis of most T7 proteins is initiated independently. However, two cases have been identified where chain-terminating mutations do reduce the synthesis of the adjacent downstream protein (28, and unpublished), suggesting that translational activation may be needed for synthesis of some T7 proteins.

There is some indication that differences in efficiency of translation of the different classes of T7 mRNAs may have a role in the co-ordinate synthesis of the three classes of T7 proteins and the rapid shut off of synthesis of host proteins (Figure 2). It appears that the late mRNAs are able to outcompete the early mRNAs and any remaining host mRNAs for translation by ribosomes (29). The molecular interactions responsible for the differences in translational efficiency are not yet understood.

REFERENCES

1. Dunn JJ, Studier FW: Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. J. Mol. Biol. (166): 477-535, 1979.
2. Moffatt BA, Dunn JJ, Studier FW: Nucleotide sequence of the gene for T7 RNA polymerase. J. Mol. Biol., in press.
3. Studier FW, Dunn JJ: Organization and expression of bacteriophage T7 DNA. Cold Spring Harbor Symp. Quant. Biol. (47): 999-1007, 1983.
4. Studier FW: Bacteriophage T7. Science (176): 367-376. 1972.
5. Studier FW: Analysis of bacteriophage T7 early RNAs and proteins on slab gels. J. Mol. Biol. (79): 237-248, 1973.
6. Summers WC, Szybalski W: Totally asymmetric transcription of coliphage T7 in vivo: correlation with poly G binding sites. Virology (34): 9-16, 1968.
7. Chamberlin M, McGrath J, Waskell L: New RNA polymerase from Escherichia coli infected with bacteriophage T7. Nature (London) (228): 227-231, 1970.
8. Dunn JJ, Studier FW: T7 early RNAs are generated by site-specific cleavages. Proc. Natl.Acad. Sci., U.S.A. (70): 1559-1563, 1973.
9. Minkley EG, Pribnow D: Transcription of the early region of bacteriophage T7: selective initiation with dinucleotides. J. Mol. Biol. (77): 255-277, 1973.
10. Dunn JJ, Studier FW: The transcription termination site at the end of the early region of bacteriophage T7 DNA. Nucl. Acids Res. (8): 2119-2132, 1980.
11. Dunn JJ, Studier FW: Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. J. Mol. Biol. (148): 303-330, 1981.
12. Saito H, Tabor S, Tamanoi F, Richardson CC: Nucleotide sequence of the primary origin of bacteriophage T7 DNA replication: relationship to adjacent genes and regulatory elements. Proc. Natl. Acad. Sci. U.S.A. (77): 3917-3921, 1980.
13. Golomb M, Chamberlin M: A preliminary map of the major transcription units read by T7 RNA polymerase on the T7 and T3 bacteriophage chromosomes. Proc. Natl.Acad. Sci. U.S.A. (71): 760-764, 1974.
14. McAllister WT, Wu H-L: Regulation of transcription of the late genes of bacteriophage T7. Proc. Natl. Acad. Sci. U.S.A. (75): 804-808, 1978.

15. Carter AD, Morris CE, McAllister WT: Revised transcription map of the late region of bacteriophage T7 DNA. J. Virol. (37): 636-642, 1981.
16. Rothman-Denes LB, Muthukrishnan S, Haselkorn R, Studier FW: A T7 gene function required for shut-off of host and early T7 transcription. In: Fox CF, Robinson WS (eds) Virus Research. Academic Press, New York, 1973, pp 227-239.
17. Brunovskis I, Summers WC: The process of infection with coliphage T7 VI. A phage gene controlling shutoff of host RNA synthesis. Virology (50): 322-327, 1972.
18. Rahmsdorf HJ, Pai SH, Ponta H, Herrlich P, Roskoski RJr, Schweiger M, Studier FW: Protein kinase induction in Escherichia coli by bacteriophage T7. Proc. Natl.Acad. Sci. U.S.A. (71): 586-589, 1974.
19. Zillig W, Fujiki H, Blum W, Janeković D, Schweiger M, Rahmsdorf H-J, Ponta H, Hirsch-Kauffmann M: In vivo and in vitro phosphorylation of DNA-dependent RNA polymerase of Escherichia coli by bacteriophage-T7-induced protein kinase. Proc. Natl. Acad. Sci. U.S.A. (72): 2506-2510, 1975.
20. Hesselbach BA, Nakada D: I protein: bacteriophage T7-coded inhibitor of Escherichia coli RNA polymerase. J. Virol. (24): 746-760, 1977.
21. LeClerc JE, Richardson CC: Gene 2 protein of bacteriophage T7: purification and requirement for packaging of T7 DNA in vitro. Proc. Natl. Acad. Sci. U.S.A. (76): 4852-4856, 1979.
22. McAllister WT, Morris C, Rosenberg AH, Studier FW: Utilization of bacteriophage T7 late promoters in recombinant plasmids during infection. J. Mol. Biol. (153): 527-544, 1981.
23. Zavriev SK, Shemyakin MF: RNA polymerase-dependent mechanism for the stepwise T7 phage DNA transport from the virion into E. coli. Nucl. Acids Res. (10): 1635-1652, 1982.
24. Dunn JJ, Studier FW: T7 early RNAs and Escherichia coli ribosomal RNAs are cut from large precursor RNAs in vivo by ribonuclease III. Proc. Natl.Acad. Sci. U.S.A. (70): 3296-3300, 1973.
25. Dunn JJ, Studier FW: Effect of RNAase III cleavage on translation of bacteriophage T7 messenger RNAs. J. Mol. Biol. (99): 487-499, 1975.
26. Summers WC: The process of infection with coliphage T7 IV. Stability of RNA in bacteriophage-infected cells. J. Mol. Biol. (51): 671-678, 1970.
27. Marrs BL, Yanofsky C: Host and bacteriophage specific messenger RNA degradation in T7-infected Escherichia coli. Nature New Biology (234): 168-170, 1971.
28. Saito H, Richardson CC: Processing of mRNA by ribonuclease III regulates expression of gene 1.2 of bacteriophage T7. Cell (27): 533-542, 1981.
29. Strome S, Young ET: Translational discrimination against bacteriophage T7 gene 0.3 messenger RNA. J. Mol. Biol. (136): 433-450, 1980.

4

ANALYSIS OF VIRAL TRANSCRIPTION-TERMINATION SIGNALS

ALIK HONIGMAN, HANA C. HYMAN and AMOS B. OPPENHEIM

ABSTRACT

A number of plasmids were designed for the isolation and study of transcription-termination signals. In this work we demonstrate the use of two plasmid vector systems for the study of various types of terminators derived from the major left operon of the bacteriophage λ.

INTRODUCTION

The chain of events leading from the primary genetic information located on the DNA, to the final product is regulated at many steps. Expression of the genetic information begins with the transcription of the gene or operon. This complicated process is regulated by a variety of genetic elements among which transcription terminators play an important role.

Transcription termination is viewed as a complex process. Following pausing of the RNA polymerase, the transcription complex is dissociated to yield mRNA, RNA polymerase, and the template DNA. This process is influenced by the nature of the sequence and in many cases by the function of auxiliary factors such as Rho and NusA proteins (1-16).

Many terminators have been shown to function efficiently in vitro in the absence of additional protein factors. The Rho-independent terminators show structural similarity: a G-C-rich region of dyad symmetry followed by a sequence of uridine residues. The stem and loop RNA structure is considered to be responsible for the pausing of the RNA polymerase which can lead to termination. The role of the uridine residues is thought to be involved in the release of the nascent RNA chain (1-8).

Other terminators require host protein factors for efficient termination. Two such factors are the Rho and NusA proteins of E. coli (9-16). Rho-dependent terminators do not share an obvious structural similarity among themselves or when compared to Rho-independent terminators. In some, an A-U rich dyad symmetry is present which may be preceded or followed by a specific "consensus" sequence (12, 13). In addition, an untranslated region of RNA may be required before the stem and loop.

Transcription-termination signals are known to be involved in regulation of gene expression. Regulation at termination sites is achieved by controlling the levels of transcriptional read-through. Some antitermination mechanisms involve coupling between transcription and translation, while others involve specific antitermination proteins.

One system, in which translation directly affects transcription termination is known as the attenuation control system. This control mechanism has been discovered in numerous biosynthetic amino-acid operons (17-19). Transcription of such an operon is controlled by the level of charged tRNAs, which regulates the termination at the beginning of the operon.

Another mechanism coupling translation to transcription termination was proposed to explain the polar effect of nonsense mutations. It has been suggested (2) that Rho interacts with termination signals only when they are exposed on naked RNA. Nonsense mutations lead to a block in the progression of the ribosomes, permitting the action of Rho.

The N and Q proteins (11, 20-23) are responsible for the antitermination of a series of terminators in each of the major operons of the bacteriophage λ. The N protein-regulated antitermination event requires the presence of a cis-acting site, called nut and a set of participating host factors (24-25).

A number of specialized plasmids have been constructed for the study of transcription control. These plasmids have been designed to facilitate the isolation and analysis of elements involved in control processes such as repression, promotion, termination and antitermination (26-33). In plasmids constructed for the study of termination signals, the terminator is placed between the promoter and the structural gene.

This chapter describes a study of various prokaryotic transcription terminators using several E. coli plasmids designed for the study of transcription termination.

RESULTS AND DISCUSSION

The effect of translation on the use of plasmid vectors in the study of transcription termination

One problem in any study involving genetic control is to distinguish between transcriptional and translational effects of gene expression. Plasmid construction must take into account the possibility that a new DNA insert could contain a genetic signal which may initiate the synthesis of a polypeptide. Translation starting within that suquence may interfere with the initiation of translation of the structural gene that is used to monitor transcriptional terminators.

When constructing the pKG plasmid vectors, McKenney et al (26) were aware of this problem and took precautions to overcome it by constructing translation-termination codons in all three possible reading frames after the insertion site and before the ribosome binding site where galK translation is initiated. The introduction of a DNA fragment upstream from the galK initiation region may alter the position at which upstream translation terminates relative to the galK translation-initiation signal and thus may change the expression of the galK gene (34). This phenomenon of inhibition of translation of galK could be misinterpreted when selecting for a relatively weak transcription terminator. The plasmids pKG1800, pKG1900 (26), and pKPG (this work) utilize the gal promoter and carry the ribosome binding site for the translation of the galE gene. Translation initiating within that region terminates at different distances from the galK initiation codon. The translation stop codons in pKG1800, pKG1900, and pKPG are 13, 198, and 118 base pairs (bp), respectively, upstream from the galK start. The data presented in Table 1 demonstrate that the position of translation-termination codons relative to the point of initiation may vary the expression of the galK gene. These results are in agreement with the findings of Schümperli et al (34). Supporting

evidence for the influence of upstream translation on galK was also demonstrated by studies on the effect of nonsense mutation suppressor hosts on galK expression (H. Hyman and A. Honigman, in preparation).

Table 1. Quantitative analysis of t_{-L1} termination activity

E. coli strains	SA1615 (Rho$^+$)		AD1600 (Rho$^-$)	
Plasmids	Galactokinase units	Relative galK expression	Galactokinase units	Relative galK expression
pKG1900	95	100	603	100
pKG1800	149	157	698	116
pKPG	102	107	841	139
pT_{L1}	22	23	857	142
pT_{L11}	33	35	555	92

The plasmids pKG1900 and PKG1800 were described by McKenney et al (26). Plasmid pKPG (this work) is similar to pKG1800 except for a BamHI DNA linker inserted into the SmaI cleavage site. The bacterial strains SA1615 and AD1600, a Rho mutant, were described by Das et al (35). The enzymatic assay for galactokinase was carried out according to the procedure described by Wilson and Hogness (36). An enzyme unit of galactokinase determines the nanomoles of galactose phosphorylated during one minute in one ml of cells at OD_{650} = 1.0. These values were not corrected for plasmid copy number. The relative values of galactokinase activity were similar in several independent experiments.

Characterization of a Rho-dependent terminator, t_{-L1}, of the major leftward operon of the bacteriophage λ

We have used the pKG plasmid system to study a Rho-dependent terminator of the bacteriophage λ. The terminator termed t_{-L1} (37), was shown to be sensitive to the λN antitermination factor (9, 11, 22). It has been suggested that transcription stops within a region containing an inverted repeat sequence located between the AluI and RsaI restriction sites (34707-34437 on the λ coordinate, 38). We cloned this 272 bp long DNA fragment into pKG1900 between the gal promoter and the galK gene and

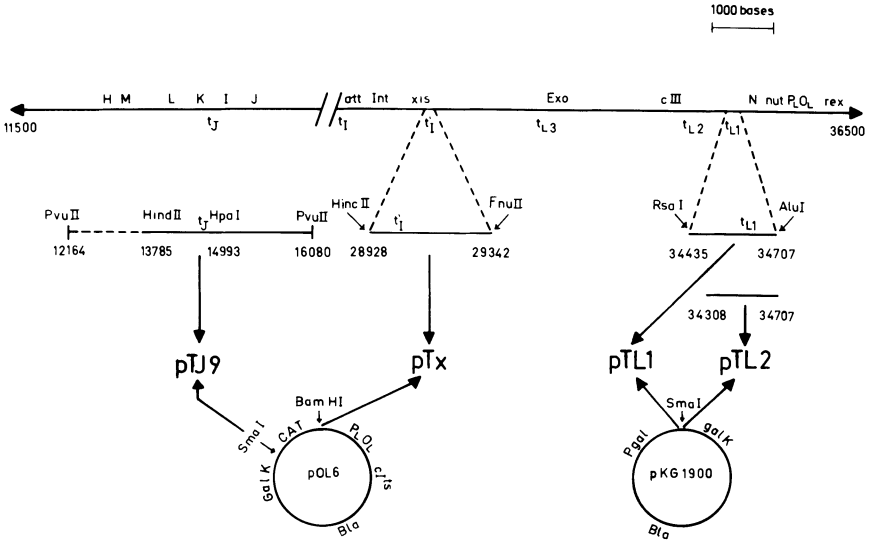

FIGURE 1. Construction of hybrid plasmid. The schematic map of bacteriophage λ indicates the DNA fragments used for cloning. The DNA fragments harboring the t_{L1} and t'_1 terminator are enlarged 5:1 relative to the λ map. The numbers on the schematic λ map and the DNA fragment relate to the position of the bases in the λ DNA sequence of Sanger et al (38).

tested its transcription-termination function. In the resulting plasmid (pT_{L1}; Fig. 1) the expression of galK was dramatically reduced. Table 1 shows that the level of galK was 20% of that expressed by the parental plasmid.

In order to test the effect of the E.coli Rho factor on this terminator, the pT_{L1} plasmid was introduced into strain AD1600 carrying a rho temperature-sensitive mutation. Our results (Table 1) show that in the rho^- host the block in expression of the galK gene was overcome. In addition, we noted that the level of galK expression, in all plasmids tested, was higher in the rho^- background that in the isogenic rho^+ host. The reason for this effect is not clear to us.

In order to determine if there are specific sequence requirements downstream from the stem and loop structure for the Rho-mediated termination, 127 bp starting at the fourth base after the stem and loop were deleted (pT_{L11}). The expression of galK in the plasmid pT_{L11}, was similar to that of pT_{L1} in both, rho^+ and rho^- hosts (Table 1). It has been suggested that a DNA sequence (ACGATCAT) forms part of several rho-dependent terminators including the t_{L1} terminator (8, 37; Fig. 2). The deletion described above removed the last two bases (AT) of the consensus sequence replacing them with C and T. This replacement creates however, a different consensus sequence (ACGATCCT) without changing its position with respect to the secondary structure of t_{L1}. This minor change in base composition and the deletion of downstream sequences do not modify termination efficiency (Table 1).

Characterization of the N-sensitive $\lambda t'_1$ and N-resistant λt_J terminator

The pLOCK vector system (39) utilizes a hybrid operon carrying the O_L P_L operator promoter region of phage λ, the chloramphenicol-acetyl transferase gene (cat) and the galactokinase gene (galK). Transcription is negatively regulated by the thermolabile λcI857 repressor (Fig. 1) and is efficiently induced upon heating. This plasmid facilitates kinetic studies and eliminates artifacts that may be caused by measuring steady-state enzyme levels. Introduction of a transcription ter-

minator between the two genes should affect only the distal gene (galK). This feature provides an independent measurement for the rate of initiation of transcription during the induction period (Cat enzyme levels) and for the efficiency of the termination signal (GalK enzyme levels). In addition the presence of the nut_L site near P_L permits us to study the effect of the antitermination protein, N, on a given terminator.

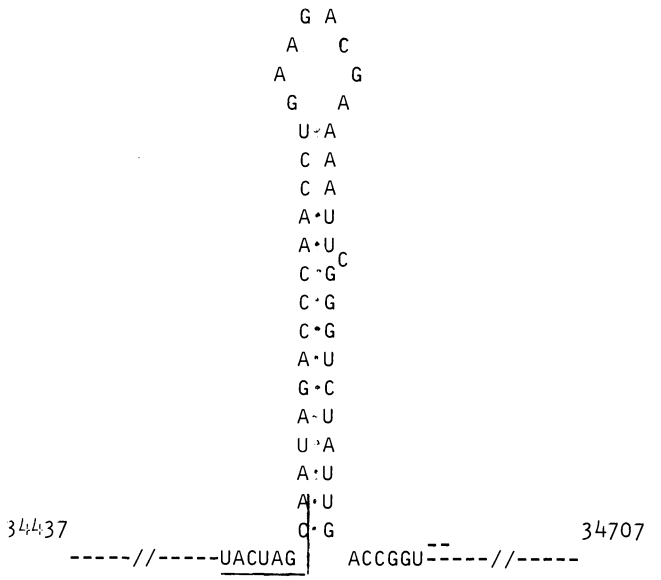

FIGURE 2. The stem and loop structure of the t_{L1} terminator. The numbers indicate the position of the sequence on the λ map. The underlined nucleotides correspond to the "consensus sequence".

We have used the pLOCK plasmid system for the isolation and analysis of two terminators t'_1 (40, 41, 42) and t_j (43, 44) from the major leftward operon of bacteriophage λ.

Immediately preceding the -35 region of the λp_I, the promoter for the CII-activated int gene transcription, lies a sequence containing a GC-rich stem and loop structure followed by a sequence of thymidine residues. Recently, it was demonstrated that this sequence, termed t'_1, functions as a terminator (40, 41).

A 100 bp DNA fragment (HinCII-FnuDII, 28940-29342) (38) carrying the putative t'_1 terminator was purified and inserted (using the BamHI lin-

kers) into the BamHI site of pOLCKR9, upstream to both cat and galK genes. Analysis of cat and galK expression following induction showed that t'_1 can act as an efficient terminator (pTX1, Fig.3). Termination also takes place in cells carrying the rho_{ts15} temperature-sensitive mutation (data not shown). The results presented in Figure 3 show that the lambda anti-termination function N can overcome the t'_1 transcription-termination barrier. This terminator was implicated to prevent unregulated transcription from reaching the xis and int gene and thus prevent excision of the λ prophage under repressed conditions (42).

The location of an N-resistant terminator, t_j, was predicted to be at the junction between the leftward transcribed "early" operon, and the "late" region of bacteriophage λ, in the vicinity of the gene J (43, 44).

The large BamHI DNA fragment (10.3 kbp) was purified and recleaved with PvuII generating 7 DNA fragments with blunt ends. These fragments were introduced into pOLCKR9 at the SmaI site between the cat and galK genes (see Fig. 1). The plasmids were introduced into E. coli strain C600 $galK^-$ and transformed cells were selected on MacConkey galactose plates containing ampicillin. Recombinant plasmids which fail to express the galK gene were purified from the "white" colonies ($galK^+$ clones form red colonies on MacConkey galactose plates) and introduced into E. coli C600 (λc_I857 $N^+ \triangle HI / \triangle BamHI$). This E. coli strain carries a cryptic λ prophage which is able to supply the N gene product upon thermal induction. Only one 3916 bp long PvuII fragment (12164-16080) abolished the expression of galK gene when introduced into either of the two E. coli strains. This plasmid, pTJ9, conferred chloramphenicol resistance to its host at 38°, indicating that the initiation of transcription at the p_L promoter at high temperature was not affected. Analysis of galactokinase and Cat activity showed that while the levels of the Cat enzyme were not changed relative to pOLCKR9 only very low levels of galactokinase are produced upon induction (Table 2). Similar results were obtained when a smaller PvuII-HinCII DNA fragment (13785-16080) was subcloned into pOLCKR9.

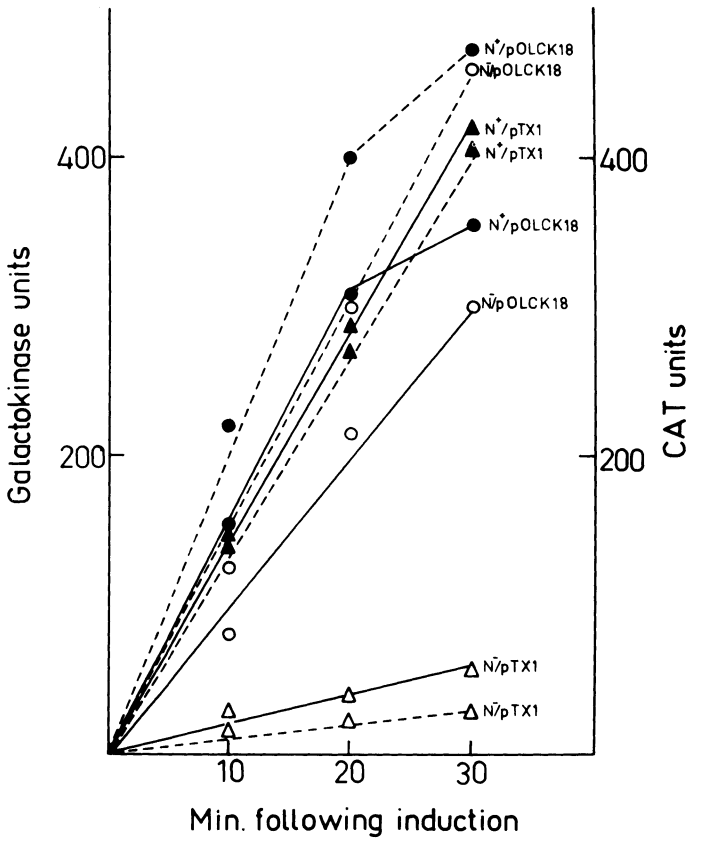

FIGURE 3. Analysis of t'_1 termination activity. The enzyme activity of either galactokinase (continuous line) or chloramphenicol acetyl transferase (broken line) was assayed either in N-producing cells W3102 (λc1857/ΔHI/ΔBamHI) or in W3102 (λc1857ΔHIΔ BamHIN_7N_{57}) in which the N function is inactivated by two amber mutations. For enzyme activity assays see legend to Table 2. The enzyme activity in the host cells (less than 10 units) was substracted.

Table 2. Termination efficiency of the N-unresponsive t_j transcription terminator

Bacterial clones	Enzyme activity following induction of the P_L promoter			
	Galactokinase units		Cat units	
	0	30 min[a]	0	30 min[a]
A12004/pTJ9	3	3.6	21	806
A2026/pTJ9	2	8.4	16	872
A2004/p0LCKR9	7	231	2	685
A2062/p0LCKR9	9	265	12	901

[a] Time after induction.

The E. coli strain A2062 is a derivative of SA500 (35) carrying the cryptic prophage λc1857N$^+$△HI△BamHI. The strain A2004 is isogenic derivative in which the N gene is inactivated by two amber mutations, N_7N_{53}. The enzymatic activities of galactokinase and chloramphenicol acetyl transferase were monitored following heat induction at 42°. The enzyme activity in the host cells was substracted from the values obtained for the plasmid-carrying clones. The assay and enzyme unit determination of galactokinase and chloramphenicol acetyl transferase are described in the legend to Table 1, and by Oppenheim et al (39).

CONCLUSIONS

Plasmid vectors specifically constructed for the isolation and study of transcription terminators were used to study three terminators of the major leftward operon of the bacteriophage λ: the t_{L1}, t'_1 and t_j.

Two locations for the t_{L1} terminator have been described so far. Das et al (45) localized a Rho-dependent terminator downstream from the BamHI cleavage site (34499 bp on the λ sequence; ref. 38) and could not detect any terminator between the p_L promoter and the nearest BamHI cleavage site. On the other hand Drahos and Szybalski (46) localized the t_{L1} terminator upstream to this BamHI cleavage site. We analysed the properties of the terminator described by Drahos and Szybalski and found that an 80% efficient Rho-dependent terminator is located upstream to

the BamHI cleavage site (between 34707-34310 on the λ coordinates, ref. 38). Moreover the sequences within the 127 bp located downstream from position 34310 had no effect on the Rho-dependent termination function of this terminator. In addition this deletion led to a change of one nucleotide, an A to C at position 7 in the consensus sequences, that did not affect the dependency of this terminator on Rho factor.

The two terminators, t'_1 and t_j differ in their sensitivity to the antitermination function of the λ N gene product. While the transcription barrier emposed by the t'_1 terminator is released by the N function, the t_j terminator is resistant to it. This distinction was made possible by using the pLOCK plasmids. Resistance to the λ N gene product can serve in classifying terminators into two functionally different groups.

We demonstrated that the position of the translation-termination signal with respect to the structural gene in pKG plasmids may have a strong influence on the expression of this gene. Thus changes in the relative position of such signals due to insertion of foreign DNA may be misinterpreted when looking for transcription terminators. The release of a block in the expression of the galK gene in the pLOCK system by the N protein can help to determine whether the lack of expression of galK is due to a translation effect or a transcriptional barrier.

We located a transcription barrier 26 bp upstream to the xis gene. This t'_1 terminator can be efficiently overcome when the λ N-antitermination function is present in the cell. The biological importance of this terminator is not clear. The possibility that a DNA sequence at this position may function as a transcription terminator was previously pointed out. It has been postulated that it may help in inhibiting transcription of the xis int region in the repressed λ prophage. Such induction would allow excision of the prophage without expression of the lytic functions of the bacteriophage (42).

The pLOCK vector system allowed us to clone and define t_j, a unique terminator that is resistant to the N-antitermination function. We localized the terminator in a DNA fragment of 3916 bp starting in the J gene and comprising the genes, I, K, L and M. Using the same system (results

not shown) we localized the t_j terminator in a DNA fragment spanning the junction region of genes J and I. These clones will allow us to look into the specific sequences composing the t_j terminator as well as providing a convenient system for isolation of mutations in the terminator. The mutations will assist in defining the components of this N unresponsive terminator. The analysis of N unresponsive terminators will add to our understanding of the function of the N gene product.

REFERENCES

1. Roberts, J.W. In: RNA polymerase (eds. Losick, R. and Chamberlin, M.), Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1976. pp. 247-271.
2. Adhya, S. and Gottesman, M. Annu. Rev. Biochem. 47:967-996, 1978.
3. Rosenberg, M. and Court, D. Annu. Rev. Genet. 13:319-353, 1979.
4. Greenblatt, J. Cell 24:8-9, 1981.
5. Platt, T. Cell 24:10-23, 1981.
6. Yanofsky, C. Nature (Lond.) 289:751-758, 1981.
7. Ward, D.F. and Gottesman, M.E. Science 216:946-951, 1982.
8. Holmes, W.M., Platt, T. and Rosenberg, M. Cell 32:1029-1032, 1983.
9. Roberts, J.W. Nature (Lond.) 224:1168-1178, 1969.
10. Greenblatt, J. and Li, J. J. Mol. Biol. 142:11-23, 1981.
11. Ward, D.F. and Gottesman, M.E. Nature (Lond.) 292:212-215, 1981.
12. Kupper, H., Sekiya, T., Rosenberg, M., Egan, J. and Landy, A. Nature (Lond.) 272:423-428, 1978.
13. Wu, A., Christie, G.E. and Platt, T. Proc. Natl. Acad. Sci. USA 78:2913-2917, 1981.
14. Lau, L.F., Roberts, J.W. and Wu, R. J. Biol. Chem. 258:9391-9397, 1983.
15. Morgan, D.W., Bear, D.G. and Hippel, P.H. J. Biol. Chem. 258:9553-9564, 1983.
16. Morgan, D.W., Bear, D.G. and Hippel, P.H. J. Biol. Chem. 258:9565-9574, 1983.
17. Lee, F. and Yanofsky, C. Proc. Natl. Acad. Sci. USA 74:4365-4359, 1977.
18. Kasai, T. Nature (Lond.) 249:523-527, 1974.
19. Zurawski, G., Brown, K., Killingly, O. and Yanofsky, C. Proc. Natl. Acad. Sci. USA 75:4271-4275, 1978.
20. Luzzati, D. J. Mol. Biol. 49:515, 1970.
21. Lozeron, H.A., Dahlberg, J.E. and Szybalski, W. Virology 71:262-267, 1976.
22. Gottesman, M.E., Adhya, S. and Das, A. J. Mol. Biol. 140:57-75, 1980.
23. Herskowiz, I. and Hagen, D. Annu. Rev. Genet. 14:399-445, 1980.
24. Salstrom, J.S. and Szybalski, W. J. Mol. Biol. 124:195-221, 1978.

25. Rosenberg, M., Court, D., Shamatake, H., Brady, C. and Wolff, D.L. Nature (Lond.) 272:414-423, 1978.
26. McKenney, K., Shimatake, H., Court, D., Schmeissner, U., Brady, C. and Rosenberg, M. In: Gene Amplification and Analysis, Vol. 2, (Eds. J.G. Chirikjian and T.S. Papas), Elsevier-North, Holland, Amsterdam, 1981, pp. 383-415.
27. Backman, K. and Ptashne, M. Cell 13:65-71, 1978.
28. Hedypeth, J., Ballivet, M. and Eisen, H. Molec. Gen. Genet. 163: 197-203, 1978.
29. Bernard, H., Remaut, E., Hershfield, M., Das, H., Helinski, D., Yanofsky, C. and Franklin, N. Gene 5:59-76, 1979.
30. Casadaban, M., Chou, I and Cohen, S. J. Bact. 143:971-980, 1980.
31. Remaut, E., Stanssens, P. and Fiers, W. Gene 15:81-93, 1981.
32. Honigman, A., Oppenheim, A.B., Hohn, B. and Hohn, T. Gene: 13:301-310, 1981.
33. Queen, C.J. Molec. Appl. Genet. 2:1-10, 1983.
34. Schümperli, D., McKenney, K., Sobieski, D.A. and Rosenberg, M. Cell 30:865-871, 1982.
35. Das, A., Court, D. and Adhya, S. Proc. Natl. Sci. USA 73:1959-1963, 1976.
36. Wilson, D. and Hogness, D.S. In: Methods in Enzymology, Vol. VIII, (Eds. Neufeld, E.F., Ginsburg, V.), Academic Press, New York, 1966, pp. 229-230.
37. Luk, K.C. and Szybalski, W. Gene 21:175-191, 1983.
38. Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. and Petersen, G.B. J. Mol. Biol. 162:729-773, 1982.
39. Oppenheim, A.B., Mahajna, J., Altuvia, S., Koby, S., Teff, D., Locker-Giladi, H., Hyman, H., and Honigma, A. In: Recombinant DNA Research and Viruses (Ed. Y. Becker), Developments in Molecular Virology, Martinus Nijhoff Publishing, Boston, The Hague, 1984 (in press).
40. Benedik, M., Mascarenhas, D. and Campbell, A. Virology 126:658-668, 1983.
41. Honigman, A. Virology 92:542-560, 1979.
42. Abraham, M., Mascarenhas, S.D., Fischer, R., Benedik, M., Campbell, A. and Echols, H. Proc. Natl. Acad. Sci. USA 77:2477-2481, 1980.
43. Gottesman, M., Adhya, S. and Das, A. J. Mol. Biol. 140:57-75, 1980.
44. Honigman, A. Gene 13:299-309, 1981.
45. Das, A., Gottesman, M.E., Wardwelli, J., Trisler, P. and Gottesman, S., Proc. Natl. Acad. Sci. USA 80:5530-5534, 1983.
46. Drahos, D. and Szybalski, W., Gene 16:261-274, 1981

mRNA of DNA viruses

5

HERPESVIRUS mRNA

EDWARD K. WAGNER, ROBERT H. COSTA, B. GAYATHRI DEVI, KENNETH G. DRAPER, RAYMOND J. FRINK, LINDA M. HALL, MARCIA K. RICE, AND WILLIAM STEINHART

SUMMARY

In the following pages, we have outlined our general methods for the precise characterization of herpes simplex virus (HSV) transcripts and the information they encode. We have shown that, despite the appearance of a complex transcription map, generally each viral transcript encodes a specific polypeptide. Thus, high resolution transcription mapping leads to a high resolution genetic map of the virus. Further, the very large number of individually acting promoters on the HSV genome provides an excellent source of data for comparative sequence analysis of one type of eucaryotic promoter. These promoters may also provide conveniently "engineerable" units for gene modification studies and for an experimental examination of the precise factors involved with the temporal control of HSV transcription.

INTRODUCTION

The replication cycle of HSV demonstrates that expression of specific classes of viral genes are temporally controlled (reviewed 1-3). Gene expression, as examined by the appearance of either specific proteins or the transcripts encoding them, can be divided into three general phases. The first group of genes (immediate-early or alpha) are expressed in the absence of de novo protein synthesis in the host. At least one immediate-early gene (that encoding

Y. Becker (ed.), VIRAL MESSENGER RNA.

ICP4) appears to be a major controlling factor in subsequent viral gene expression. Following the appearance of the immediate-early genes, genes encoding enzymes involved in viral DNA replication and certain alterations in the cell surface are expressed as early (beta) genes. Genes of both these groups are expressed as abundant transcripts in the absence or blockage of viral DNA synthesis. Early transcription continues at high rates following the onset of viral DNA replication. Structural proteins of the virion and many other proteins, most of which have as yet unknown function, are expressed as late (gamma) genes. Late transcripts can be divided into two general classes, depending on whether they can be detected at all in the absence of viral DNA synthesis. However, all require viral DNA synthesis to attain their maximum rate of transcription, or at least, maximum abundance on polyribosomes.

In this present review, we have briefly described results of our studies on the transcription of individual HSV genes. We briefly reviewed some of our procedures for such a study in the first volume of this series (4). HSV transcripts share many features with both cellular mRNA and the mRNAs expressed by other nuclear-replicating DNA viruses: they are capped, polyadenylated, and have (generally) a ca. 150-base leader between the cap and translation-initiation codon. Further, certain sequence features around individual transcription units are shared: HSV promoters contain "TATA" and often "CATC" boxes, and the sequence signal indicating a polyadenylation site ("AATAAAA") appears standard.

In spite of these similarities, there is one very distinct difference between the structure of HSV transcription units and those of other nuclear-replicating DNA viruses. Each HSV transcript appears to be controlled by its own promoter and encodes a specific polypeptide. Further, the high degree of splicing seen with most eucaryotic and viral mRNAs is not seen in HSV. Certain spliced transcripts do exist in HSV and other herpesviruses

(5-8), but these are in the minority (with HSV, at least). Thus, with HSV, a whole hierarchy of potential control points utilized in eucaryotic gene expression may be missing or rarely utilized. Despite this, the high density of gene packaging and relatively complex arrangement of partially overlapping transcripts is seen in HSV transcription as it is in other, smaller DNA viruses.

METHODOLOGY

We have described our procedures for handling specific regions of the HSV-1 genome as recombinant DNA fragments in pBR322 grown in LE392 (9, 10). Recently, we have begun to use single-stranded DNA clones grown in bacteriophage M13.

Procedures for isolation and characterization of specific HSV-1 transcripts have been outlined previously (1, 2, 4). Many variations of these methods are used by other laboratories with similar results.

Isolated mRNA can be translated *in vitro* using commercial rabbit reticulocyte lysate systems. We have used specific antibodies (polyclonal and monoclonal) to identify *in vitro* translation products with particular viral proteins.

We have used RNA (Northern) blots to determine the specific number and sizes of HSV transcripts homologous to any given region of the genome. We recently described the use of a method for doing RNA blots *in situ* by using vacuum-dried agarose strips as the supporting medium (11); such a procedure has the advantage of speed and excellent size resolution.

Radioactive DNA probes are prepared by nick-translation of cloned DNA fragments. Sizes of probes as small as 200-300 bases have been described (12, 13) for the precise localization of the 5' ends of overlapping mRNA species. Recently, the use of single-stranded DNA probes has allowed us to determine the direction of transcription and high resolution map location (ca. 100 bases) of mRNAs prior to S1

nuclease and exonuclease VII digestion of hybrids for precise location of transcripts (ca. 25 base resolution). Northern blots using RNA from infected cells in which viral DNA replication has been blocked allow the assignment of temporal class of a given transcript (4).

Procedures for S1 nuclease mapping of HSV-1 transcripts, the use of specific end-labeled DNA probes, and the use of exonuclease VII digestion of RNA:DNA hybrids to specifically locate given transcripts have been described in a number of papers from this group. Our most recent methodology is also described (12-14).

Nucleotide sequence analysis has been done by the method of Maxam and Gilbert (15). Procedures for the precise location of the 5' and 3' ends of specific mRNAs (ca. 5 base resolution) have been outlined by Frink et al. (12, 13, 16).

RESULTS

HSV transcription as a model for the group

Judging by the number of members, if nothing else, the herpesviruses must be considered a very successful group of nuclear replicating DNA viruses in vertebrates. Members infecting humans include HSV (types 1 and 2), herpes (varicella) zoster (VZV), cytomegalovirus (HCMV), and Epstein-Barr virus (EBV). Members of the herpesvirus group share general morphological similarities, large genome size, and certain complex features of the physical arrangement of their genomes. Many of these features have been reviewed in detail elsewhere (3). Some comparative structural features of herpesvirus genomes have been briefly reviewed by Becker (17) in the first volume of this series.

It is notable that all members of the herpesvirus group have complex biological interactions with their hosts. They have the ability to establish latent infections where the virus is refractory to host defense and yet periodically

re-erupt (18). Natural infection by certain herpesviruses appears to lead to formation of specific neoplasias. Certainly, specific genetic elements of all herpesviruses rigorously tested to date can be shown to induce morphological transformation in indicator cells (briefly reviewed for HSV by Galloway and McDougall, 19).

Features of gene expression control during the replication cycle of herpesviruses appear to be common among many members of the group. HSV (especially HSV-1) is the most extensively characterized of the herpesviruses and it is expected that the model it provides will be generally applicable to many other members of the group. In the case of the closely related HSV-2, we have collaborated with Galloway and colleagues to carry out a detailed examination of the homology between these virus types between 0.59-0.65 map units (Draper, Frink, Swain, Galloway, and Wagner, in preparation). We have used Northern and Southern blot cross-hybridization, as well as comparative sequence analysis. Generally, regions encoding proteins are homologous between the types, but DNA segments which are not contained in a translational frame are quite divergent. Further, several regions of encoding mRNA which are highly homologous still show great sequence divergence. These data, which are shown in Figure 1, suggest that significant changes in the pattern of expression of a given gene are quite possible, even between closely related herpesvirus subtypes, and generalizations should reflect this.

The HSV-1 transcript map

To date, we have carried out high resolution transcription mapping in the 23 kb region between 0.15 and 0.3, and in the 52 kb region between 0.4 and 0.75 on the HSV-1 genome. These data, along with data taken from several other laboratories, yield the map shown in Fig. 2.

Although many overlapping mRNA "families" are seen, most show no evidence of splices. There are, however, exceptions. These include the two 1.8 kb alpha mRNAs (0.86

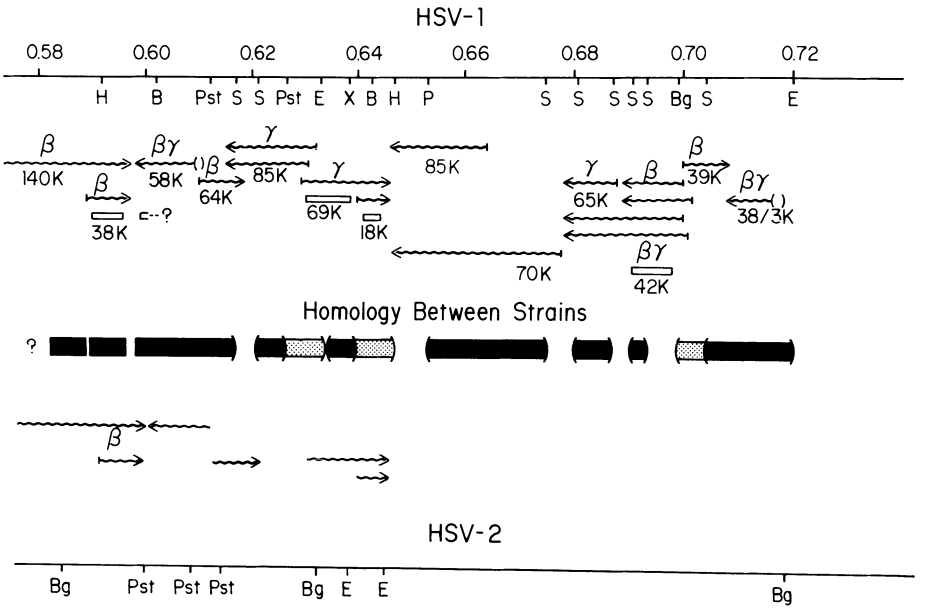

FIGURE 1

FIGURE 1. Summary of sequence homology between HSV-1 and HSV-2 in the region between 0.58-0.72. Transcripts mapped in the region for the two subtypes are indicated. Time of appearance and in vitro translation products are shown for type-1. Open boxes below transcripts indicate identified translational reading frames. Regions of homology based on comparative sequences and Southern blot hybridization analysis are shown: dark bars - high homology; stipeled bars - detectable homology; open areas - no detectable homology.

FIGURE 2 (opposite). High resolution map of specific HSV-1 mRNAs. The times of appearance of mRNAs are shown at the top along with the size of polypeptide products they translate in vitro (where known). The direction of transcription is indicated. Position of the transcripts encoded in brackets is based on high resolution Northern blot analysis (ca. 100 base resolution) and only partial S1 nuclease data. Therefore, the exact locations should be regarded as tentative. In addition to data from our laboratory, we have interpreted data from other laboratories (5, 20-23) for the mRNAs indicated with an asterik (*). Transcripts between 0.31-0.4 have been mapped by L. Holland, R. Sandri-Goldin, A. Goldin, J. Glorioso, and M. Levine (submitted).

and 0.96) characterized by Watson et al. (5). The transcription unit encoding gC (0.63-0.65) has several low abundance spliced members (6, 12). Several other mRNAs

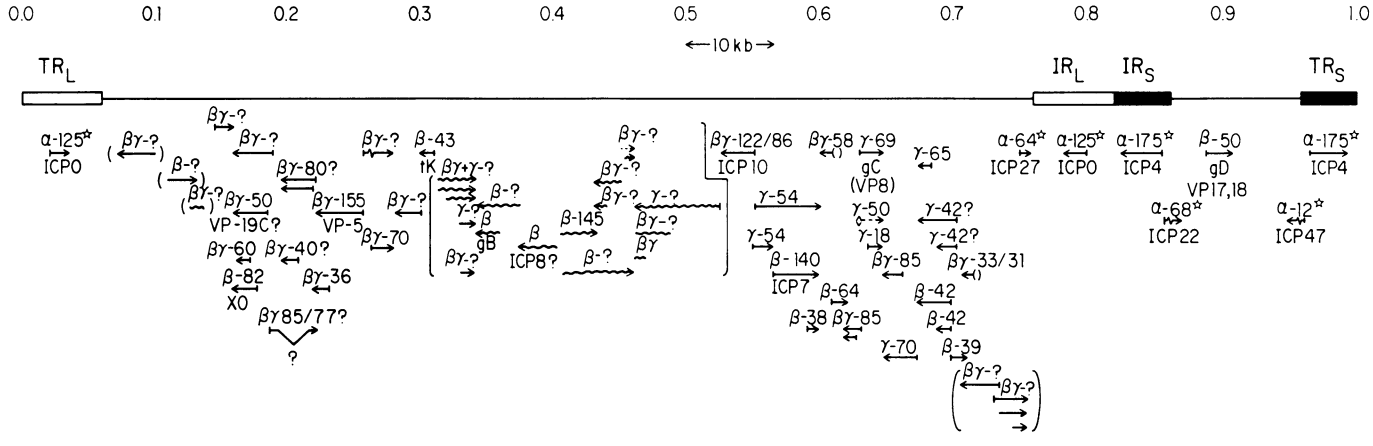
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
←10 kb→
TR_L
IR_L
IR_S
TR_S
α-125☆ ICP0
α-64☆ ICP27
α-125☆ ICP0
α-175☆ ICP4
α-175☆ ICP4
α-68☆ ICP22
α-12☆ ICP47
β-50 gD
VP17,18
βγ-122/86 ICP10
γ-69 gC (VP8)
γ-65
γ-54
γ-50
γ-18
γ-42?
γ-42?
γ-54
β-140 ICP7
β-64
β-38
βγ-85
β-42
β-42
γ-70
β-39
βγ-85
β-43
tK
βγ-80?
βγ-50 VP-19C?
βγ-155 VP-5
βγ-70
βγ-60
βγ-40?
β-82 XO
βγ-36
βγ85/77?
?
β-145
ICP8?
gB
β-?
βγ-33/31

FIGURE 2

appear to have short discontinuities near their 5' ends. These are indicated by brackets at those 5' ends in Fig. 2. To date, however, only one mRNA, a 2.8 kb one mapping between 0.185 and 0.225 map units, has a large intron. Here, we have determined it to be on the order of 4 kb (Costa, Draper, and Wagner, unpublished). We expect other exceptions to occur; but as yet, we have not found a region of the genome where transcripts are as extensively spliced as has been seen in some other viral systems.

Most HSV-1 mRNAs mapped do not show significant complementary overlaps with others, but there are exceptions. We (6) reported one minor 2.7 kb mRNA mapping in HindIII fragment L (0.592-0.647) which had a complementary overlap with a major beta mRNA, as well as with the mRNA encoding gC. We carefully analyzed one group of mRNAs in EcoRI fragment I (0.633-0.721) whose 5' ends showed complementary overlap and here we found that potential protein-coding frames did not overlap (11). Other individual cases will no doubt occur.

Nuclear forms of HSV mRNA

The relatively low frequency of spliced HSV mRNA species and the fact that transcriptional control sequences map contiguously to structural genes (see below) both lead to the prediction that HSV nuclear mRNA precursors will generally be close to the size of the mature product. We carried out some experiments to examine the properties of high molecular weight HSV nuclear RNA in general (Frink and Wagner, unpublished). We used guanidine-urea extraction of nuclear RNA (24) to examine the size distribution of nuclear RNA via Northern blot and S1 nuclease analysis. The mRNAs encoded by HindIII fragment K (0.527-0.592) were used as a standard. At both early and late times after infection, RNA transfer (Northern blots) showed that the poly(A+) nuclear RNA contained mRNA species identical to polyribosomal poly(A) mRNA (see Fig. 3, tracks 1-4). The picture with total nuclear RNA was complicated by the fact that late (but

not early) after infection, there was considerable heterodispersely migrating RNA of large size (not shown). S1 analysis of hybrids between ^{32}P-labeled HindIII fragment K DNA and poly(A) and total nuclear RNA, however, showed only the four major DNA fragments identified with polyribosomal (A+) mRNA extending 6.0, 4.3, 3.7, and 1.7 kb into this restriction fragment (Fig. 3, tracks 6 and 7). Similar conclusions were obtained with mRNA from BamHI fragment I (0.60-0.64). Thus, we concluded that no specific RNA species other than those seen on polyribosomes are generally present in readily detectable amounts in the nucleus.

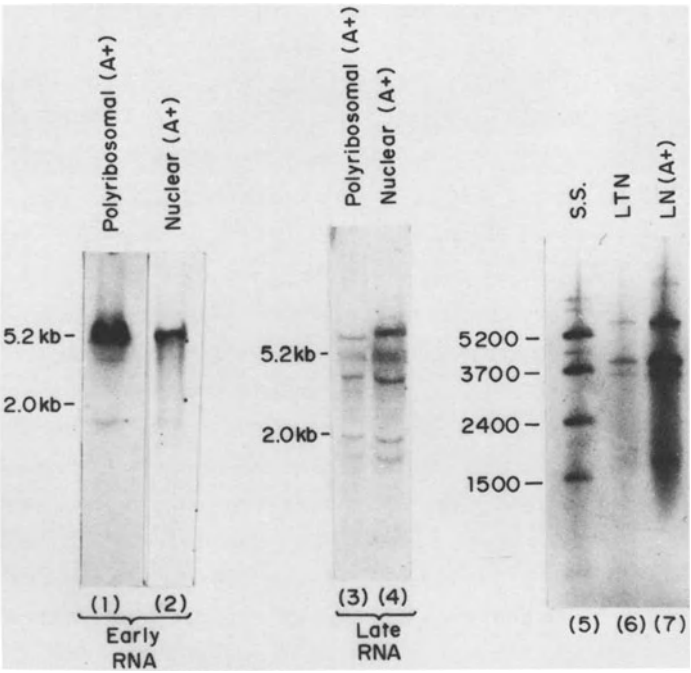

FIGURE 3. Nuclear forms of HSV RNA encoded by HindIII fragment K (0.527-0.592) are of the same general size as seen on polyribosomes. General methods for RNA transfer (Northern blots) and S1 nuclease mapping are discussed in the Methodology section. Track 5 (s.s.) shows size markers included as standards.

In a second set of experiments, we asked if any members of the overlapping mRNAs of HindIII fragment K were kinetic precursors to others. Very short duration ^{32}P-pulses were used, and nuclear poly(A) mRNA was isolated using preparative hybridization. This material was then subjected to size-fractionation. At the shortest pulses (ca. 5 min), we found that the proportion of 7 kb, 5.2 kb, and 1.9 kb mRNAs isolable using a specific HSV-1 DNA fragment were the same as for longer pulses. We concluded that no readily detectable precursor-product relationship exists between these mRNAs.

HSV transcripts encode identifiable viral proteins

The transcription map of Fig. 2 serves as a high resolution genetic map for HSV only if individual transcripts can be shown to encode specific viral proteins of known function. Such has been done for a number of such proteins by workers in this and other laboratories (1, 2, 4). As an example, consider the identification of the 6 kb late mRNA mapping between 0.25-0.29 as the transcript encoding the major HSV-1 capsid protein (VP5). We were able to tentatively make this assignment on the basis of a number of indirect criteria, including size of the *in vitro* translation product and correlation with intertypic recombinant mapping data (10). Such an identification should be considered tentative. We have recently made an absolute assignment by using a specific antiserum againt VP5 to precipitate the *in vitro* translation product of the 6 kb mRNA (14). A typical result is shown in Fig. 4. A final confirmation was done by comparing tryptic peptides of authentic and *in vitro* translated VP5.

The mRNA species characterized by workers in this lab that have been identified with viral proteins of known function are listed in Table 1. Although all assignments are considered to be reliable, those based solely on size of translation product and map position should be considered as tentative.

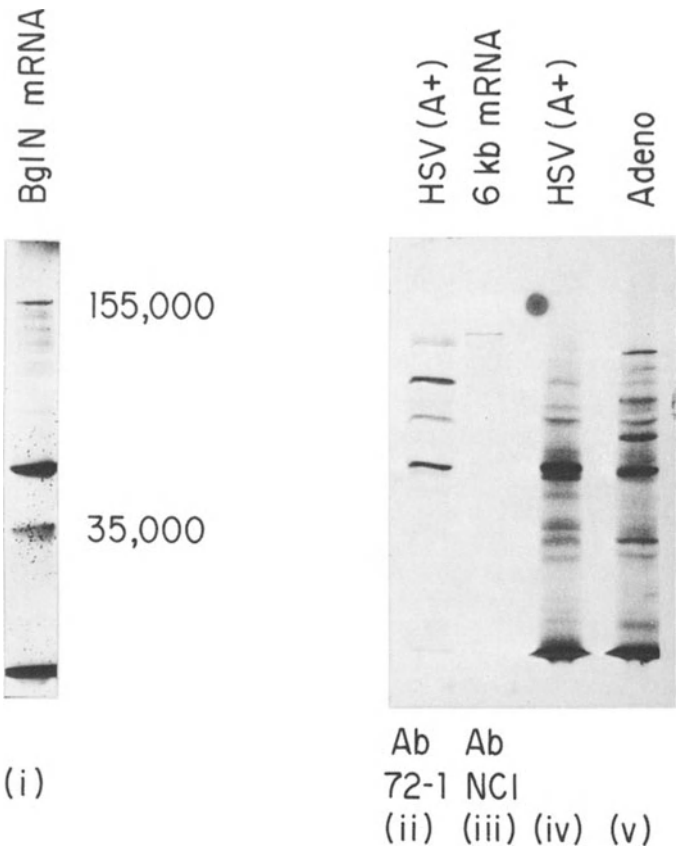

FIGURE 4. Immune precipitation of the major HSV-1 capsid protein (VP5) by specific antisera. The 155,000 d translation product of the 6 kb mRNA was reacted with antiserum NCI as described (14). Precipitation of total in vitro translation product of HSV poly(A) mRNA with immune rabbit serum (72-1) is shown for comparison.

The relative ease of making such assignments leads us to some confidence that the information encoded in the majority of transcripts seen in HSV-1 infection will have readily definable function. This will be of great use in studies on gene-controlling factors involved in HSV pathogenesis. For example, recently we collaborated with another group to determine that one or several HSV-1 genes involved in neurovirulence in mice can be mapped to a specific region of the genome (25). The patterns of HSV transcription we have characterized suggests that such functions will be

assignable to specific proteins and the transcripts encoding them will be identifiable. This leads to obvious approaches toward the study of the molecular mechanisms of HSV pathogenesis.

Sequence analysis of HSV transcription units

The complete nucleotide sequence of the transcription units around the gene encoding HSV-1 gC has been published by Frink et al. (12). A schematic representation of some important features of these data is shown in Fig. 5 and serves as a useful model for other HSV transcription units. We have briefly outlined some of the features of such units in the following sections.

Table 1. HSV-1 Transcripts Encoding Identified Proteins

Size (kb)	Map Position	Protein	Temporal Class	Reference
(Tentatively identified by correlation with map position and size translation product.)				
4.2	0.82-0.86 0.96-1.00	ICP4 Regulatory	I.E.	(26)
4.2	0.41-0.43	DNA Polymerase	Early	(4)
5.2	0.57-0.60	Ribonucleotide	Early	(2, 27)
1.5	0.59-0.60	Reductase	Early	(2, 27)
(Identified by immune precipitation.)				
2.5	0.16-0.175	Alkaline Exonuclease	Early	(13)
3.9	0.16-0.19	Capsid (vertex)(?) (Weak immune reaction)	Late	(13)
(Identified by immune precipitation and tryptic peptide comparison.)				
6	0.25-0.29	Capsid (major)	Late	(14)
2.7	0.63-0.65	glycoprotein C (gC)	Late	(12)

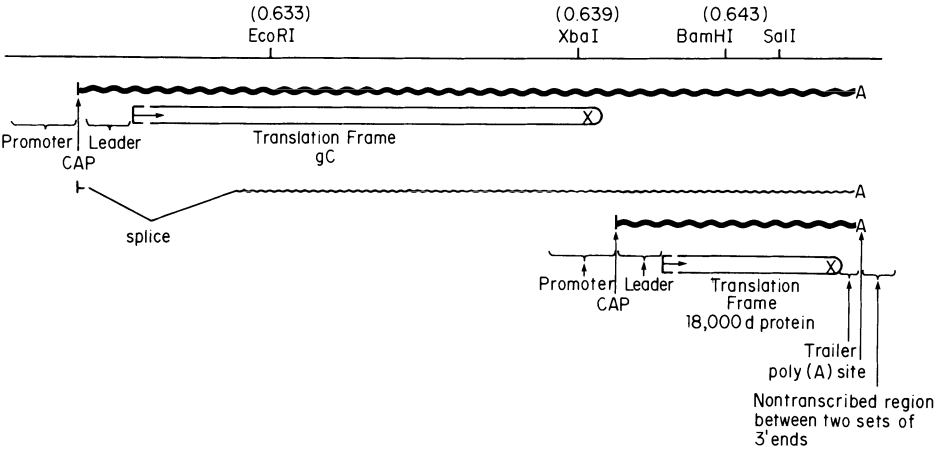

FIGURE 5. Transcription around the HSV-1 gC gene. Data are summarized elsewhere (12). Individual features of the DNA sequence encoding the gC transcript family are discussed in the text.

Promoters. The data available concerning the properties of different HSV promoters have been extensively reviewed (2). The best model for the promoters in general comes from the work of McKnight (28) with the HSV-1 thymidine kinase promoter. HSV promoters must be responsive to both cellular factors (RNA polymerase II, etc.) and viral control factors that regulate both the time of expression of given temporal classes of viral genes and, presumably, the amount of transcription taking place. On the basis of partial data from a number of other laboratories, it is reasonable to suggest that many, if not all, of the control sequences involved in early and late HSV promoter function will be found in the 250 or so bases directly upstream of the mRNA cap site. Actually, the size of many such promoters may be only 120 bases or so based on the modification studies of McKnight referred to above.

Early HSV genes are generally expressible at low levels in uninfected cells. We showed that, generally, early (beta) HSV-1 promoters are properly recognized by RNA polymerase complexes from uninfected cells (Manley systems, 16, 29). In contrast, we have found that the promoter for

the late 6 kb mRNA encoding VP5 is not efficiently recognized. Recently, we have prepared Manley-type extracts from HSV-1 infected cells taken 8 h after infection. Such a system shows increased activity as judged by total incorporation of label into product. Further, our data are consistent with initiation of transcription near the cap site for the 6 kb mRNA (Draper and Wagner, unpublished). Unfortunately, the system from infected cells initiates at a number of sites which do not correspond to recognizable late mRNA cap sites; therefore, specificity is suspect.

R. Costa, in my laboratory, has generated a transcription-expression marker by modification of pBR322. This marker contains a restriction site into which the VP-5 promoter has been placed. We have found that transfection of this marker into HeLa cells, followed by superinfection with HSV, leads to specific induction of transcription. Such a marker will hopefully be of value in assessing which viral factors are responsible for efficient transcription from late promoters.

We have carried out preliminary experiments with protein transfer blots to examine the specificity of ds-DNA binding to infected cell nuclear extracts (Steinhart and Wagner, unpublished). High salt (1.7 M) extracts of infected cell nuclei contain several protein species not seen in uninfected cell extracts that bind to DNA probes made by nick-translating BamHI fragment F' (0.268-0.272), the fragment containing the promoter region for the 6 kb late mRNA. Competition experiments with pBR322 DNA or HSV fragment SalI fragment N' (0.617-0.621) DNA (a region in HindIII fragment L which does not contain any recognizable promoter) suggest that this binding is sequence-specific (see Fig. 6). Such preliminary data give hope that protein factors involved in the regulation of specific HSV promoter function may be available for detailed experimental investigation.

Comparative sequence analysis of a number of early and late HSV promoters has demonstrated some common features and

some class-specific ones. A "TATA" box or its variant is generally (but not always) seen around 30 bases upstream of specific mRNA cap sites. In early (beta) promoters, a string of ca. 15 bases which is very "AC-rich" in no particular order is seen around 100 bases upstream of the cap site. Late promoters do not show this feature, but often have an "AT-rich" region this far upstream. These features are shown in Table 2.

Translation frames. The mRNAs that we have currently identified have leaders of around 150 bases between the cap and the probable initiation codon. Specific sequences in such leaders can be seen to diverge quite widely in comparative sequence studies between HSV-1 and HSV-2 DNA-encoding transcripts for highly homologous proteins. This conclusion is based on comparison of the sequence for the 1.2 kb early mRNA encoding the 38,000 protein mapping between 0.59-0.60 and the corresponding HSV-2 sequence (29; Galloway and Swain, personal communication).

Translation-initiation codons are often the canonical "Pu-ATGG" identified by Kozak (30). The high "G+C" content of HSV DNA is reflected in codon-use frequency, but such frequency is notably assymetric (2). High proline contents are common in the predicted amino acid contents of HSV proteins, but they can vary widely. The very high proline content predicted for HSV-1 gC may account for the discrepency between its calculated residue molecular weight and that actually observed (12).

Termination signals are generally repeated several times in phase downstream of the first signal seen in a given reading frame. Trailer sequences can vary from a length of ca. 10 bases to many thousands of bases between termination and polyadenylation signal.

Splices. As noted above, splicing is rare in HSV-1. In the case of the gC transcription unit, several splice signals appear to function at low frequency. We have not fully characterized the splice donor sequence, but it is very short (ca. 25-50 bases). Canonical splice acceptor

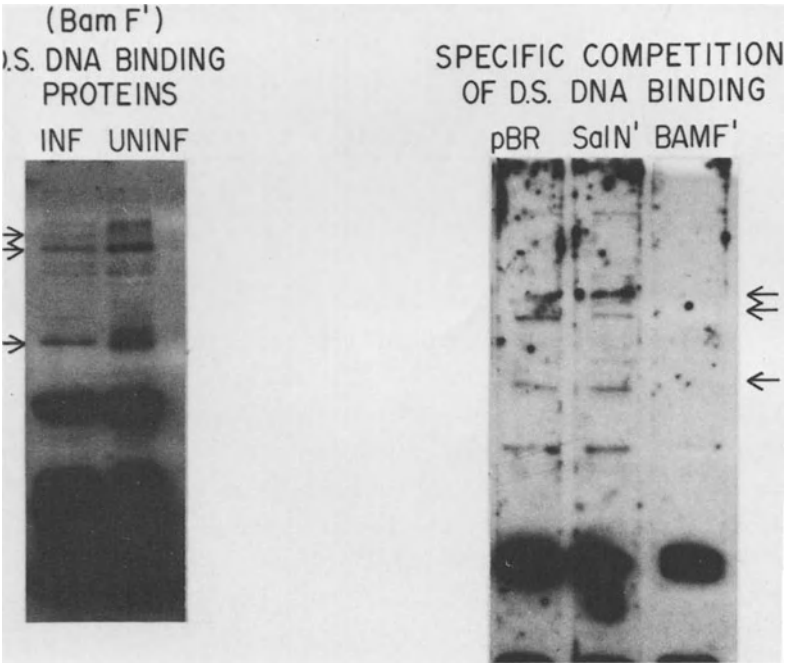

FIGURE 6. Specific competition of double-strand HSV-1 BamHI fragment F' (0.268-0.272) binding to protein transfer (Western) blots of high salt extracts of HSV-infected HeLa cell nuclei. Cells were isolated at 8 h after infection and nuclear proteins extracted with 1.7 M salt. These were fractionated on SDS-acrylamide gels by electrophoresis and transferred to nitrocellulose. Blots were incubated with ^{32}P-labeled (nick-translated) double-stranded BamHI fragment F' DNA. This labeled DNA was mixed with a large excess of unlabeled competing DNA as indicated.

signals are seen at positions corresponding to the located acceptors determined by high-resolution S1 mapping. These data are reviewed (12). We have not determined any biological function for the spliced mRNA species we have seen. Several alternate reading frames for translation could be utilized by them, and one species could give rise to a protein sharing the C terminal sequence of the gC protein. None of these proteins has been rigorously identified either by <u>in vitro</u> translation or in the infected cell.

Partially overlapping mRNAs. Several mRNAs sharing the same polyadenylation site is a common feature of the HSV transcription map. These overlapping mRNAs have independent 5' ends and appear to be under their own promoter control. Such promoter regions can lie within a translational reading frame of an upstream mRNA.

It is not at all clear whether the partial overlapping of mRNAs encoding distinct proteins is a reflection of closely related function. In a number of cases, we have found that there is no obvious immunological cross-reactivity between such proteins (14), yet the situation with the two identified ribonucleotide reductase proteins (26) suggests that the 140,000 d and 38,000 d proteins encoded by distinct translational reading frames do share an epitope. Further, we have found that the 1.9 kb mRNA underlying the 3' end of the 2.3 kb alkaline exonuclease mRNA may well share that mRNA's translational reading frame (13). Thus, further data are needed to make generalizations.

Polyadenylation sites. The polyadenylation signal for HSV mRNAs appears to be the canonical "AATAAA(A)" (31). Often, transcript groups or "families" encoded by both strands of the viral DNA have their polyadenylation signals close together. We have determined the nucleotide sequence between such juxtaposed polyadenylation signals in the region around 0.60 for both HSV-1 and HSV-2 (Draper, Frink, Galloway, Swain, and Wagner, in preparation). It was seen that the region in both virus types is characterized by a high "A+T" composition. However, there is no evidence of any sequence conservation between the two types. In fact, HSV-2 has an extra 70 or so bases in this region compared to HSV-1. However, as soon as the analogous translational reading frames are encountered, a good deal of homology is noted.

Table 2. Some Features of HSV Promoters.

Position of the feature in the sequence upstream of the cap site:	
Early mRNAs	
"AC" string	"TATA" box
Alkaline exonuclease (5' end at 0.175) (13)	
(-120)AGACCAACACCCACGGCC(-103)	(-26)TATAAATTA)-18)
Thymidine kinase (5' end at 0.315)(20)	
(-112)ACACAAACCCCGCCCA(-97)	(-27)ATATTAA(-21)
Ribonucleotide reductase (?) (5.2 kb 5' end at 0.565)(16)	
(-114)AAGGAACACACCCCC(-99)	(-27)ATAAAAA(-21)
Ribonucleotide reductase (?) (1.2 kb 5' end at 0.590)(29)	
(-117)ACCATTAGCCAATCCATGACCC(-97)	(-29)ATATAA(-24)
Unknown function 38,000 protein (1.5 kb 5' end at 0.699)(11)	
(-119)ACCCCTCACCCCACACA(-103)	(-42)ATAATA(-37)
Late mRNAs	
"AT" string	"TATA" box
VP5 mRNA (5' end at 0.265, revised position)(14)	
(-107)AATTTCTTCCTGCACGCTTTT(-87)	(-28)TATATAA(-22)
Glycoprotein C (5' end 0.630)(12)	
(-114)TATTTTTCAATAAAAGGCATTA(-93)	(-28)TATAAATT(-21)
Unknown function 18,000 d protein (5' end 0.640)(12)	
(-92)TTACCTTTTTAATATCTATATAGTTT(-67)	(-32)TATAAA(-27)

DISCUSSION

The data briefly described in this review of work carried out in my laboratory over the past few years indicate that we have a good understanding of the physical properties of individual HSV transcription units. Hopefully, such knowledge will be of great value in gaining a full description of the molecular biology of the herpesviruses in general. Deciphering the molecular

processes responsible for the complex biology of these viruses will not be a trivial task; however, gene packaging of the viruses appears to be readily interpretable.

Further, data presented allows the inference that temporal regulation of the expression of specific viral functions requires *trans* action of viral regulatory functions upon defined regions of the viral genome. The large number of such regulatory sequences and the fact that they can be readily located suggest potentially fruitful approaches for investigation of the actual molecular processes involved in this control.

ACKNOWLEDGMENTS

Research was supported by grants from the National Cancer Institute of the NIH (CA11861) and the American Cancer Society (MV139). Support from the California Cancer Research Coordinating Committee was also for a portion of the time covered in this review. R.C. was supported by training grant GH-07311. W.S. was on sabbatical leave from the Biology Department of Bowdoin College during the 1982-83 academic year. I thank J. Wagner for manuscript preparation.

REFERENCES

1. Wagner, E. *In*: Advances in Viral Oncology, Vol. III (Ed. G. Klein), Raven Press, New York, 1983, pp. 239-270.
2. Wagner, E. *In*: Comprehensive Virology: The Herpesviruses, Vol. III (Ed. B. Roizman), 1984, in press.
3. Spear, P. and Roizman, B. *In*: Molecular Biology of Tumor Viruses, 2nd ed., Part 2: DNA Tumor viruses (J. Tooze, ed.), Cold Spring Harbor, NY, 1980, pp. 615-746.
4. Wagner, E., Anderson, K., Costa, R., Devi, G., Gaylord, B., Holland, L., Stringer, J., and Tribble, L. Isolation and characterization of HSV-1 mRNA, *In*: Herpesvirus DNA, Vol. I, Developments In Molecular Virology (Ed. Y.

Becker), Martinus Nijhoff, B.V., The Hague, 1981, pp. 45-67.

5. Watson, R.J., Sullivan, M., and Vande Woude, G.F. J. Virol. 37:431-444, 1981.
6. Frink, R.J., Anderson, K.P., and Wagner, E.K. J. Virol. 39:559-572, 1981.
7. Stinski, M.F., Thomsen, D.R., Stenberg, R.M., and Goldstein, L.C. J. Virol. 46:1-14, 1983.
8. Van Santen, V., Cheung, A., Hummel, M., and Kieff, E. J. Virol. 46:424-433, 1983.
9. Anderson, K., Frink, R., Devi, G., Gaylord, B., Costa, R., and Wagner, E. J. Virol. 37:1011-1027, 1981.
10. Costa, R.H., Devi, B.G., Anderson, K.P., Gaylord, B.H., and Wagner, E.K. J. Virol. 38:483-496, 1981.
11. Hall, L.M., Draper, K.G., Frink, R.J., Costa, R.H., and Wagner, E.K. J. Virol. 43:594-607, 1982.
12. Frink, R.J., Eisenberg, R., Cohen, G., and Wagner, E.K. J. Virol. 45:634-647, 1983.
13. Costa, R., Draper, K., Banks, L., Powell, K., Cohen, G., Eisenberg, R., and Wagner, E. J. Virol. 48:591-603, 1983.
14. Costa, R., Cohen, G., Eisenberg, R., Long, D., and Wagner, E. J. Virol., 1984, submitted.
15. Maxam, A. and Gilbert, W. Methods Enzymol. 65:499-559, 1980.
16. Frink, R.J., Draper, K.G., and Wagner, E.K. Proc. Natl. Acad. Sci. USA 78:6139-6143, 1981.
17. Becker, Y. Current trends in herpesvirus DNA Research, In: Herpesvirus DNA. Developments in Molecular Virology, Vol. I (Ed. Y. Becker), Martinus Nijhoff, B.V., The Hague, 1981, pp. 1-10.
18. Stevens, J.G. In: Oncogenic Herpesviruses, Vol. II (Ed. F. Rapp), CRC Press, Boca Raton, Florida, 1980, pp. 1-17.
19. Galloway, D. and McDougall, J. Nature 302:21-24, 1983.
20. McKnight, S.L. Nuc. Acids Res. 8:5949-5964, 1980.

21. Watson, R.J., Weis, J.H., Salstrom, J.S., and Enquist, L.W. Science 218:381-384, 1982.
22. Mackem, S. and Roizman, B. J. Virol. 44:939-949, 1982.
23. Sharp, J.A., Wagner, M.J., and Summers, W.C. J. Virol. 45:10-17, 1983.
24. Berk, A.J., Lee, F., Harrison, T., Williams, J., and Sharp, P.A. Cell 17:935-944, 1979.
25. Thompson, R. Wagner, E., and Stevens, J.G. Virology, 1984, in press.
26. Anderson, K., Costa, R., Holland, L., and Wagner, E. J. Virol. 34:9-27, 1980.
27. Dutia, B.M. J. gen. Virol. 64:513-524, 1983.
28. McKnight, S.L., Cell 31:355-365, 1982.
29. Manley, J., Fire, A., Cano, A., Sharp, P., and Gefter, M. Proc. Natl. Acad. Sci. USA 77:3855-3859, 1980.
30. Draper, K.G., Frink, R.J., and Wagner, E.K. J. Virol. 43:1123-1128, 1982.
31. Kozak, M. Nuc. Acids. Res. 9:5233-5252, 1981.
32. Proudfoot, N. and Brownlee, G. Nature 263:211-214, 1976.

6

ORGANIZATION AND CONTROL OF THE mRNA OF THE HSV TK GENE

JAMES R. SMILEY

1. SUMMARY

The sequences necessary for the activation of HSV TK gene transcription by HSV immediate-early proteins are largely coincident with the upstream region of the TK promoter. The implications of this arrangement are discussed.

2. INTRODUCTION

Herpes simplex virus (HSV) encodes its own distinctive thymidine kinase (TK) activity (1-3). Because TK is a convenient selective marker, the HSV TK gene has been extensively used in gene transfer studies in mammalian cells, and in genetic studies of HSV. In most of these applications the primary emphasis has been on the function of other genes or sequences. However in this review, I will focus on the properties of the TK gene itself, limiting discussion to the sequences and factors which regulate its expression both in natural viral infections and in various more or less artifical expression systems.

3. MAP LOCATION AND STRUCTURE

TK is unique among HSV genes in that its map location was established using gene transfer techniques rather than the more classical approaches of genetic analysis and/or hybrid-selection translation. Following the initial demonstration by Wigler et al. (3) that the 3.5 kb BamHI Q fragment of HSV-1 DNA was sufficient to transform TK$^-$-deficient mouse cells to TK$^+$, several groups contributed to narrowing the boundaries of the minimal transforming sequence to a 1.7 kb region (4-6). McKnight (7), and Wagner et al. (8) then established the nucleotide sequence across this segment, identifying an open reading frame capable of

Y. Becker (ed.), VIRAL MESSENGER RNA.

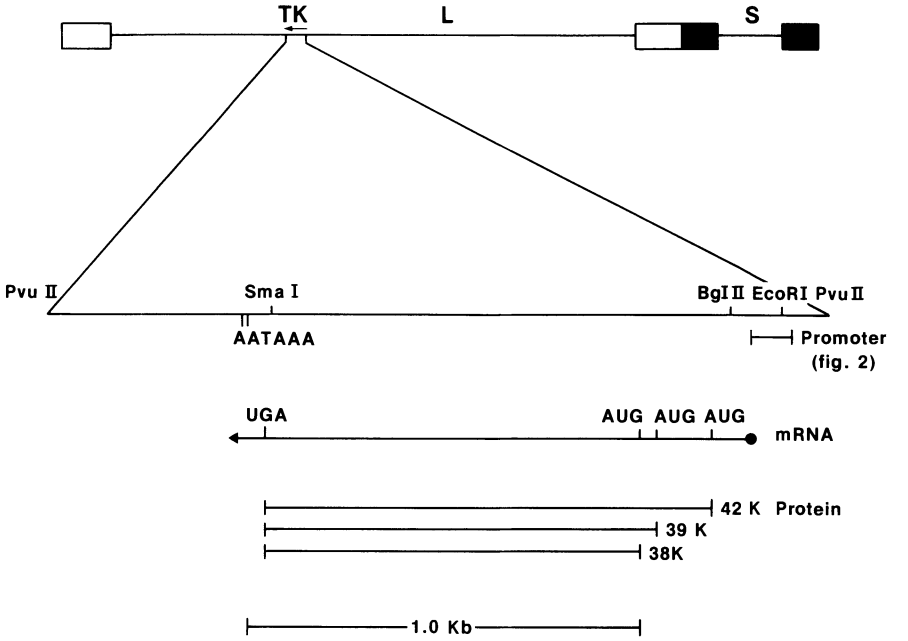

FIGURE 1. Location and structure of the HSV TK gene. The map location of the TK gene is indicated on the prototype HSV genome. The positions of the regions encoding the 42K, 39K and 38K polypeptides are diagrammed in the lower portion of the figure.

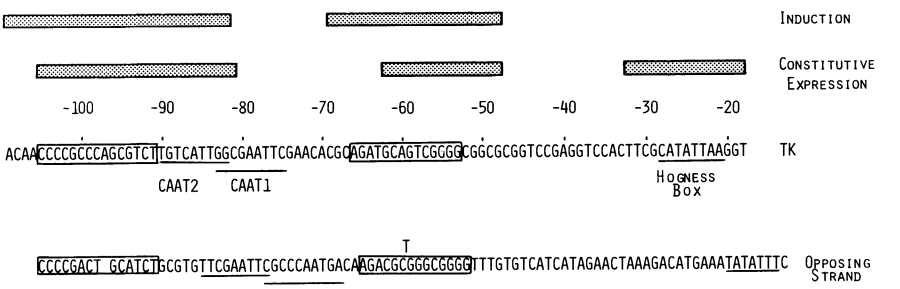

FIGURE 2. Sequence and structure of the TK regulatory region. The nucleotide sequence from -109 to -18 is given. The 15 base imperfect inverted repeats are boxed, while CAAT and Hogness sequences are underscored. The stipled boxes above the sequence indicate the areas needed for induction by ICP4, and for constitutive expression in Xenopus oocytes. The lower sequence is that of the opposing strand, beginning at residue -53 of the TK sequence.

encoding a protein of the appropriate size. The polarity of this coding region confirmed the 5' to 3' polarity of TK previously inferred from genetic analysis (9). Transcript mapping showed that the major TK mRNA is approximately 1.3 kb in length, and colinear with the DNA template (7, 8). The transcript initiates 109 nucleotides upstream from the first potential initiator codon, and ends about 70 nucleotides downstream from the UAG terminator codon. As detailed further below, the TK-promoter region contains signals common to many polymerase II promoter sites. The 3' end of the transcript is preceeded by two tandem AAUAAA sequences, known to form part of the polyadenylation signal for eukaryotic mRNAs (10, 11). These features are summarized in Figure 1.

The predicted size of the major TK polypeptide (39K, 376 amino acids) inferred from the DNA sequence agrees well with the value of 42K measured by gel electrophoresis. During lytic infections, small amounts of 39 and 38K TK polypeptides are also found in addition to the major 42K species (12). Marsden et al. (13) proposed that these smaller proteins result from internal translational initiations at the second and third AUG codon in the TK mRNA (see figure 1). This proposal has recently been confirmed using a mutant virus which is deleted for the first methionyl codon, and initiates transcription at residue +198 of the wild-type mRNA (Halpern and Smiley, unpublished). The mutant fails to synthesize any 42K protein, and produces dramatically elevated amounts of both the 39 and 38K species (Haar, Smiley, Marsden & Preston, unpublished). As the mutant induces 40% of the wild-type levels of enzyme activity, it follows that either or both of the 39 and 38K polypeptides retain appreciable catalyic activity. This result strongly suggests that the first AUG codon of the wild-type mRNA is in fact the preferred translational initiation site, and further demonstrates that the first 45 amino acid residues of the wild-type TK polypeptide are dispensable for catalytic activity. The function, if any, of the 39 and 38K polypeptides produced during wild-type infections remains unknown.

4. EXPRESSION OF HERPES SIMPLEX VIRAL GENES

Before concentrating on the control of TK expression, it is useful to very briefly review what is known about the control and structure of HSV genes in general. HSV genes are transcribed in the nucleus of infected cells by the host cell RNA polymerase II (14). Perhaps the most

interesting aspect of HSV gene expression is that not all viral genes are expressed simultaneously after infection; rather, different sets of genes are expressed in a well-ordered sequence with at least three (and likely more) discernable phases (15). The transition from one phase to the next depends on viral-coded products (15), and seems to be accomplished largely at the level of transcription. In contrast to most other viral and cellular transcripts, HSV transcripts are usually not spliced; perhaps as a consequence, HSV genes generally are each transcribed from their own promoter sites. Making the situation more complex, genes belonging to the same kinetic class are not always grouped together along the viral chromosome; instead genes of different classes are interspersed with one another. The problem then is to account for the existence of a limited number of kinetic classes of viral genes given this organization. A priori, the most obvious strategy to accomplish this would be to arrange that the promoter site of each gene respond to activators and inhibitors in a fashion defining its kinetic class. The major predictions of this type of model are first that the promoter region of each HSV gene defines its kinetic class, and second, that the regulatory responses of each promoter will be largely indifferent to the context in which it finds itself. As detailed below, considerable evidence supports this point of view.

Immediately following infection, the five immediate-early or α-viral genes are transcribed, each from a separate promoter site (16-18). These five α genes are the only ones transcribed in the absence of viral protein synthesis, suggesting that the protein product of one or more of them is necessary for the transcription of the remaining non-α viral genes. Consistent with this view, temperature-sensitive mutations in the α gene encoding the protein ICP4 block the appearance of all but α transcripts at the non-permissive temperature (19-21). More strikingly, temperature-shift experiments have shown that the ICP4 function is required continuously throughout infection to maintain ongoing transcription of non-α genes (21). Thus it seems that ICP4 somehow contributes to the activation of the non-α HSV genes. It is not yet known whether or not any of the other four gene products also contribute to this activation, although in at least one case it appears that ICP4 alone is capable of at least partial activation of non-α promoters (see below). The mechanism of action of ICP4 remains entirely

unclear, and will be discussed at length later in this review.

The α promoters themselves appear to be positively regulated, most likely by a viral protein present as part of the virus particle: this conclusion emerged from studies in which α promoters were fused to the TK structural gene. In the first of these experiments, the upstream region of the gene for ICP4 was studied (22). When an ICP4-TK hybrid gene was recombined into the intact HSV chromosome, TK was converted to an α gene, demonstrating that α control resides within the α upstream region. Surprisingly, when this same ICP4-TK hybrid gene was introduced into uninfected TK-cells, only comparatively low constitutive levels of expression were obtained. However this level was boosted dramatically following superinfection with HSV, with α kinetics. Studies with inhibitors and mutants have shown that this transactivation of the ICP4 promoter by infection does not depend on viral protein synthesis, or even on complete uncoating of the superinfecting virus particle (23, 24). These observations have lead to the suggestion that a virion component serves as the α inducer. The elements within the ICP4 upstream region that are sufficient for the α induction phenomenon have been mapped to upstream of -100 (23). Interestingly, these regulatory sequences are not needed for the constitutive level expression in uninfected cells: rather they are specifically required for induction by superinfecting virus. Analogous results have been obtained with another α upstream region (25, 26). Taken together, these data imply that the HSV virion contains one or more factors which activate α promoters, thus ensuring efficient transcription early in infection. This activator factor requires for its action upstream cis-acting sequences distinct from the constitutive promoter elements. In turn, at least one of the resulting α proteins, ICP4, is required for expression of the non-α HSV genes. However as will be outlined in greater detail below, α-mediated activation of non-α genes appears to require cis-acting sequences which are largely indistinguishable from the constitutive promoter of the target genes.

The non-α genes of HSV have classically been divided into two broad categories, β and γ (or early and late) (15). Transcription of both of these classes requires continuous ICP4 function (19-21). The defining differences between these classes are first that the maximal rate of expression of β genes is obtained earlier than the maximal rate of expression of γ genes, and second that the expression of γ genes is not

reduced by blocking viral DNA replication, while the expression of β genes is. The fact that the rate of β gene expression does not increase concomitantly with DNA replication implies either that β expression is actively turned off late in infection, or that replicating viral DNA is not a template for β transcription. It is not yet possible to distinguish between these possibilities. It seems likely that the classification of non-α genes into two categories is oversimplified, as originally emphasized by its formulators (15). At present, the clearest additional complication is that at least two classes of γ genes exist: those that are detectably expressed in the absence of DNA replication (eg. the gene encoding the major capsid protein, VP5) (27), and those that are not (eg. the gene encoding glycoprotein C) (28,29). One interpretation of this situation is that genes belonging to the former class of γ genes (the so-called β-γ class) are activated with β kinetics, by α proteins, and are not shut-off as DNA replication begins. By contrast, the latter class (true γ genes) may either directly require factors in addition to immediate-early gene products for their expression, or be expressable only from a replicating template.

5. EXPRESSION OF TK

5.1. General Considerations

TK is a β, or early, viral gene (30). As such, during lytic infection its transcription appears to require only α viral proteins in addition to the host transcription machinery. This was most convincingly demonstrated by Preston (20), using the ICP4 temperature-sensitive mutant tsK. At the non-permissive temperature during tsK infections, no detectable TK mRNA was made, demonstrating that ICP4 is required for TK mRNA synthesis. Following a shift down to the permissive temperature, TK mRNA accumulated rapidly, even when the temperature shift was carried out in the presence of cycloheximide. Since only α proteins were made at the nonpermissive temperature, and since no further protein synthesis was allowed after the shift down, the conclusion is that only α proteins are required for TK mRNA synthesis. This conclusion agrees with those of several other studies, which used inhibitors of protein synthesis (30, 31). This requirement for α function is remarkably tight: when infections are carried out in the presence of cycloheximide, no TK mRNA is detectable by RNA-driven hybridization to a TK DNA probe, setting an

upper limit for the leak-through level to about 0.1 RNA molecules per cell (32). It seems clear then that TK transcription in lytic infections is entirely dependent on α function. However this is not the case in various other expression systems where TK DNA is presented to cells out of context (discussed further below). Similar to the situation with other β genes, TK gene expression appears to be shut off as viral DNA replication begins. The evidence for this assertion is first that the rate of increase of TK enzymatic activity peaks before the peak of DNA synthesis, and second, that inhibiting viral DNA replication results in an overproduction of TK activity(30). Correlated with these observations made at the level of enzymatic activity, the steady state levels of TK mRNA peak relatively early during infections (33). Whether this apparent shut off is the result of a specific repression of TK transcription, or instead reflects a change in the conformation of DNA template as infection proceeds, remains unclear.

5.2. TK Expression in Uninfected Cells

5.2.1. General. Munyon and coworkers (2) were the first to demonstrate HSV TK expression in uninfected cells. Following infection of TK-deficient mouse cells with ultraviolet-inactivated HSV virions, cell lines that expressed the viral enzyme were recovered at a low frequency. Somewhat later it was shown by other workers that this biochemical "transformation" could also be achieved using either purified HSV DNA (34), or isolated HSV DNA fragments (3). In the initial experiments using virus as the transforming agent it remained possible that the expression of TK in the resulting transformants was dependent on α gene function; however the more recent experiments using purified restriction fragments have shown that fragments containing TK as the only complete HSV gene efficiently transform cells to TK^+ (4). In addition, the purified TK gene is also expressed in Xenopus oocytes. One interpretation of these findings is that the low levels of TK expressiion observed in transformed cells reflect a certain amount of "leaky" ICP4-independent expression, which also occurs during lytic infection. However, this view does not seem to be tenable, as TK mRNA cannot be detected in the absence of ICP4 function during lytic infections (20, 32): expression of TK is not "leaky". It therefore seems more likely that expression in uninfected cells results from a partial removal of the

necessity for α function, and so differs from expression during viral infection. For this reason it is interesting to review what little is known of the general requirements for TK expression in uninfected cells.

The first point to be made is that transcription in uninfected cells initiates at the same major site as in lytic viral infections (7, 8), implying that the TK promoter site is, at least under some conditions, recognized by the uninfected cells transcription machinery. Further, efficient expression requires the integrity of the TK promoter sequence. The second very general point is that TK genes resident in uninfected cells remain responsive to whatever changes α proteins induce. There are three distinct systems in which the purified TK gene is expressed in uninfected cells: in microinjected Xenopus oocytes, in acutely transfected mammalian cells, and in more or less stably transformed mammalian cell lines established by transfection or microinjection of DNA. In the former two situations, TK is apparently expressed from extrachromosomal copies of the gene, while in the latter case expression is from chromosomally integrated copies. In a provocative paper, Harland, Weintraub and McKnight (35) demonstrated that in order to be expressed extrachromosomally in Xenopus oocytes, the TK gene must be part of a circular DNA molecule. Although linear molecules were stable in the microinjected oocytes, they were not detectably transcribed. Apparently then, the conformation of the TK gene profoundly influences its expression in oocytes, suggesting that the activity of the TK promoter is sensitive to such alterations. Perhaps a circular conformation allows the molecule to become torsionally stressed, for example by supertwisting; conceivably this conformational alteration partially mimics the effects of ICP4. It is not yet known whether or not TK gene must also be circular in order to be expressed extrachromosomally in transfected mammalian cells, however in the absence of any evidence to the contrary, it seems reasonable to assume that this is so.

In contrast to the above situations where TK is expressed extrachromosomally, the TK gene is associated with host cell chromosomes in biochemically transformed cells (36, 37). This situation raises the possibility that expression of TK is influenced by the flanking cellular DNA sequences. There are strong reasons for believing that this is the case. Perhaps the most compelling evidence comes from studies on the effects of viral enhancer sequences on TK transformation. When cells are

exposed to pure TK DNA lacking exogenous enhancer sequences, transformants are isolated at a characteristic frequency. In some cases these transformed cells contain a single copy of the TK gene (37). This result indicates that, at some integration sites, a single TK gene can be expressed at a high enough level to result in a TK^+ colony. However, this transformation frequency is boosted 20 to 40 fold by placing an SV40 or retroviral enhancer sequence in close proximity to the TK gene (38). As the presence of the enhancer sequence does not appear to alter the integration frequency of TK DNA, the clear implication is that the viral enhancer sequence increases the fraction of TK insertions that result in a TK^+ phenotype. This in turn implies that most insertions of enhancer-less TK genes do not result in a TK^+ colony. One explanation is that expression of the TK gene from chromosomal sites in uninfected cells depends on its proximity to an active enhancer element. When the transforming DNA segment itself lacks an enhancer, only the subset of insertions that place the TK gene close to an active cellular enhancer result in TK^+ colonies. By contrast, when an enhancer element forms part of the transforming segment, expression becomes relatively independent of the insertion site, so that a much higher fraction of the insertion events yield a TK^+ colony. Although this scheme has yet to be proven, it seems plausible, and provides a possible explanation for the ICP4-independent expression of chromosomally integrated TK genes. That is, one can imagine that the TK promoter is activated by linkage to an adjacent enhancer, instead of by the action of viral regulators.

In summary, there are good reasons to suspect that the α-independent expression of TK observed in uninfected cells depends on factors which remove the α requirement. In the case of extrachromosomal genes, circularity seems to be essential; in the case of integrated genes, flanking enhancer sequences may be required.

5.2.2. The TK Promoter Sequence. Although very little is known about the external factors that contribute to TK expression in uninfected cells, a great deal is known about the cis-acting TK DNA sequences that are required for this expression. In fact, the constitutive TK promoter is probably the best characterized of all eukaryotic polymerase II promoters,due mainly to the intensive and elegant studies of McKnight and coworkers. Before reviewing the results of these studies, it is worthwhile to point out some of the more obvious features of the TK

promoter sequence. These features are summarized in Figure 2. Like most polymerase II promoters, the TK promoter contains a recognizable Hogness sequence "CATATTAA" beginning at -28 with respect to the transcription initiation site. Studies of other promoters have suggested that this sequence functions mainly to direct the initiation event to one or a few sites approximately 30 residues downstream (39, 40). In some systems this sequence does not appear to have a major influence on the overall levels of transcription initiation. The results obtained with the TK promoter support these views. Further upstream, between -90 and -72, the TK promoter contains two partially overlapping sequences similar to the "CAAT" homology which is found at similar positions in many but not all polII promoters (41). For some reason, only one of these two CAAT sequences is recognized as such in the literature (that which extends from -83 to -72). In the case of the β-globin gene, the "CAAT" sequence seems to play a major role in governing the overall level of transcription *in vivo*; however its role in TK promoter function remains controversial.

In order to delineate functional regions, McKnight and colleagues have systematically mutagenized the TK promoter, using two complementary strategies. In the first of these studies (42) two series of deletion mutations were constructed; one of which progressively removed 5'-flanking sequences, the other which removed 3'-flanking sequences. Each deletion endpoint was marked by a 10 base *BamH*1 linker. The transcriptional competence of each mutant gene was then assayed following microinjection of Xenopus oocytes, and in some cases by testing for its ability to transform mouse LtK^- cells to TK^+. The results demonstrated the following points:

1. Using the 5'-deletion mutants it was found that a mutant retaining only 109 5'-flanking residues supported wild-type levels of accurate initiation. By contrast, a deletion extending up to -95 resulted in a markedly reduced level of transcription; however the residual transcripts synthesized by the mutant still initiated at +1. This suggests that the 5' boundary of a region required for efficient initiation is located between -109 and -95. Further deletions extending through to -46 displayed the same general pattern.

2. 5' deletions extending to the Hogness sequence resulted in a further drop in the level of initiation, and had the additional effect of

eliminating the tight clustering of initiation events at +1. These mutants initiated at a multiplicity of sites.

3. 3'-deletion mutants extending upstream beyond +1, but not covering the Hogness sequence had little effect on initiation: transcripts still initiated 25-30 residues downstream of the Hogness sequence, at novel sites.

4. 3' deletions covering the Hogness sequence reduced the total level of transcription somewhat, but had the more striking effect of eliminating the tight clustering of initiation events.

5. 3' deletions extending further to -52 had relatively minor addition effects; however, deletions extending upstream of -52 dramatically reduced the total level of transcription.

Taken together, these results lead the authors to conclude that expression in uninfected cells depends on a minimum of two upstream regions. One, extending from about -32 to -16, including the Hogness sequence, acts to fix the transcription initiation site 25-30 nucleotides downstream, and plays only a minor role in governing the overall level of initiation. The other, extending from about -109 to -52, including both "CAAT" homologies, governs the overall level of transcription. These conclusions received further support from the analysis of a novel class of clustered point mutations, termed linker-scanning mutations, which were constructed by McKnight and Kingsbury (43). These mutations systematically replaced 10 base pair segments of the TK upstream region with synthetic BamH1 linkers, leaving the spacing of all other sequences unaltered. The main additional feature revealed by this analysis was that the previously identified upstream region extending from -109 to -52 was composed of at least two essential subregions separated by at least 18 bases of dispensable DNA. The two essential elements, termed the first and second distal regions, extend from about -62 to -48 and -105 to -81, respectively. The dispensable region separating them includes one of two "CAAT" homologies, (CAAT1) while the other (CAAT2) sequence apparently forms part of the second distal element (see Figure 2). Because the dispensable CAAT1 sequence was the only one recognized as such by McKnight and Kingsbury, these authors concluded that the CAAT sequence does not play a major role in TK expression . In my view, this conclusion is premature, as it remains possible that expression requires the integrity of one or the other of the two homologies. So far,

mutations that destroy CAAT2 without affecting CAAT1 have not been analyzed. Interestingly, linker-scanning mutation -84/-74, which destroys both CAAT homologies, is a strong down mutation. It therefore seems possible that there are three, rather than two, important sequences in the -105 to -48 region; the second distal element (-105 to -91), the CAAT cluster (-90 to -72) and the first distal element (-62 to -48). One reason for favouring this arrangement is the complex pattern of overlap between the TK promoter and the promoter for a diverging transcript (described below). Presumably, these residual uncertainties will be resolved by the analysis of point mutations in the TK promoter region now in progress (S. McKnight, personal communication).

An interesting point raised by McKnight and Kingsbury is that portions of the two distal upstream elements could potentially hydrogen bond: residues -103 to -98 in the second distal element (CCGCCC) are complementary to residues -55 to -49 in the first distal element (GGGCGG). In fact, close inspection of the sequences reveals that the homology between the second distal element and the complement of the first distal element is even more extensive than noted by McKnight and Kingsbury (see figure 2): allowing for a single base deletion in the first distal element, the sequences match at 11 out of 15 sites. That is, the homology between these sequences covers a substantial fraction of their functionally important residues. This fraction increases further if one chooses to separate the CAAT2 sequence from the second distal element. It is therefore conceivable that expression depends on pairing between two distal elements (43); alternatively, expression may involve the recognition of these two related, inverted, elements by a common factor. It is my prejudice that the latter explanation is more likely to prove correct.

Having identified three separate functional areas in the TK upstream region, McKnight (44) investigated the interactions between these signals, in two ways. The first approach was to determine the effects of combined mutations that separately inactivate two of the three functional elements. The results demonstrated that mutations in the first and second distal elements do not interact: the phenotypes of double mutants were similar to those of single mutants in which only one element was inactivated. By contrast, mutations in the second distal element interacted strongly with mutations in the proximal element, yielding a

marked decrease in expression by comparison to either of the single mutants. These results suggested that the first and second distal signals are involved in a common step of promoter recognition and perhaps comprise a single functional unit. The proximal element was suggested to be required for a different step of transcription initiation, and to comprise a separate functional unit. These ideas were tested further by examining the effects of altering the spacing between the three signals. The spacing between the proximal element and the two distal signals could be increased by up to thirty bases without effect, while increasing the separation by more than fifty bases abolished transcription. The spacing between the first and second distal signals was only somewhat less flexible: it could be contracted or expanded by ten residues without effect, while expansion by sixteen residues reduced, and expansion by thirty-six residues eliminated, the function of the upstream region. These results tend to support the idea that the proximal signal is functionally distinct from the distal signals, but also suggest that if the first and second distal signals actually do comprise a single functional region, then the size of its dispensable central area is not rigidly constrained. Based on all of the available evidence, McKnight (44) suggested that the distal signals together form a preferred RNA polymerase loading site, while the proximal signal (ie. the Hogness sequence) triggers polymerase to initiate 30 nucleotides downstream.

In summary, the results of mutational analysis of the TK promoter are consistent with the following picture: The promoter contains two functionally distinct domains. The first, corresponding to the Hogness sequence, serves to fix the site of the transcription initiation event, but does not play a major role in setting the level of transcription initiation. The second domain, extending from -105 to -48, appears to determine the overall level of transcription. The outer boundaries of this second region are occupied by imperfect inverted repeats of 15 bases; these inverted repeats are largely coincident with the two functionally important elements of the region as established by mutational analysis. Between these inverted sequences are two overlapping CAAT homologies, as well as other sequences. It is not yet known whether or not at least one intact CAAT sequence is required for promoter function, although mutations which destroy one but leave the other intact have no measurable effect, and a mutation which destroys

both impairs transcription. Sequences to the proximal side of the CAAT sequences appear to be dispensable, and the spacing between the functionally important portions of the region is not rigidly constrained. However, mutations in the functionally important ends of the region do not interact, suggesting that the region forms a single functional unit. One attractive possibility is that the region is an RNA polymerase II loading site.

One very striking feature of the organization of the TK promoter is the presence of inverted repeats which define the boundaries of, and also largely correspond to the functionally important elements within, the upstream region of the promoter. This feature leads one to wonder whether or not the TK upstream region also forms part of a diverging promoter. The evidence on this point is suggestive, but inconclusive at present. Read and Summers (45) have examined the *in vitro* transcription of the TK region, and found a total of 5 promoter sites recognized. One of these transcripts initiated approximately 180 nucleotides upstream of the TK initiation site, on the opposite DNA strand. More recent data from the same laboratory indicate that this same transcript is also found during lytic viral infections (W.C. Summers, personal communication). An interesting consequence of this arrangement is that the presumed upstream elements of this diverging promoter overlap with the upstream regions of the TK promoter in a striking fashion. Specifically, two CAAT homologies located at -86 to -70 with respect to the diverging transcript are coded by the complement of some of the residues specifying the TK CAAT homologies. As a consequence, the CAAT homologies of the diverging transcript are flanked by the same inverted sequences that define the ends of the TK upstream element. This suggests the possibility that the upstream regions of TK and the diverging transcript are coincident (diagrammed in Fig. 2). Although this suggestion remains to be proven, it at least raises the possibility that the presence of inverted repeats in the TK upstream region reflects the overlap of two diverging promoters, rather than indicating an inherent requirement for HSV promoter function.

5.2.3. *Transformation by Promoter Mutants.* Although an intact TK promoter sequence is required for efficient transformation of TK^- cells to TK^+, transformants do arise after exposing cells to TK genes bearing promoter mutations, albeit at a much reduced frequency. For example, 5'

deletion mutants lacking the portions of the upstream region beginning at about -105 transform cells at about 5% of the wild-type frequency, while mutants deleted to +56 transform about 0.1% as well as wild-type (42). In all of these cases, the resulting transformants contain approximately the same levels of viral TK activity as those isolated using the wild-type gene. Two related questions arise from these findings: how does TK come to be expressed in these cell lines, and what accounts for the reduced transforming ability of the mutant genes? Unfortunately, the data relating to these questions are incomplete, and scattered in the literature. Nevertheless, there are indications that at least part of the explanation involves the expression of TK from alternative promoter sites. As these alternative promoters are capable of providing only low levels of TK enzymatic activity, many copies of the mutant gene are required to produce the TK^+ phenotype.

The first indication that alternative promoters might contribute to TK expression in transformed cells came from studies of TK-related RNA present in cells transformed by wild-type TK gene. Both Roberts and Axel (46) and Ostrander, Vogel and Silverstein (47) found that some of these cell lines contained a novel 0.9 kb transcript, in addition to the bona fide 1.3 kb TK mRNA. As the 0.9 kb transcript accumulated to about the same level as the 1.3 kb RNA, Roberts and Axel proposed that TK DNA contains another relatively strong promoter site which is used in transformed cells, but not in lytic infection.

Roberts and Axel (46) introduced a single copy of a TK DNA fragment lacking the sequences upstream from -59 into TK^-, $APRT^-$ cells, by selecting for an unlinked APRT marker. These cells, which remained phenotypically TK^-, expressed normal amounts of the 0.9 kb RNA, and much smaller amounts of a 1.1 kb RNA. These data suggested that the synthesis of the 0.9 kb RNA does not depend on the TK promoter sequence, and that the 0.9 kb RNA does not code for fully active TK. The presence of lower levels of the 1.1 kb RNA in these cells further suggested the existence of an additional weak promoter site. Although the cells expressing these two RNAs were phenotypically TK, TK^+ variants could be selected at a low frequency. These variants were found to have amplified the transforming DNA segment, resulting in a considerably increased TK DNA copy number. Two classes of such amplified TK^+ variants were found: those that synthesized dramatically increased amounts of the 0.9 kb RNA, and those

that synthesized the 1.1 kb RNA in quantities comparable to the levels of the 1.3 kb RNA found in wild-type transformed cells. Based on these data, the authors proposed the following model: TK DNA contains three potential promoter sites which are used in transformed cells. The first corresponds to that identified by McKnight, which drives synthesis of the 1.3 kb transcript. This is the only promoter that functions at a detectable level during viral infections. The second promoter, comparable in strength to the bona fide TK promoter, drives the 0.9 kb RNA. Although the structure of the 0.9 kb mRNA remains to be directly determined, Roberts and Axel proposed that it is 3' coterminal with the normal 1.3 kb TK mRNA, and initiates at about +400 with respect to the 5' end of the 1.3 kb RNA. As this site lies within the TK coding sequence, the 0.9 kb RNA could encode only a truncated TK polypeptide lacking the first 120 amino terminal residues. This proposal is thus consistent with the observation that very high levels of this transcript appear to be required to convert cells to the TK^+ phenotype: the implication is that the resulting truncated protein retains only marginal activity. Although plausible, these ideas about the 0.9 kb RNA have yet to be directly tested. The third proposed promoter is the very weak one that drives the 1.1 kb RNA. This transcript was postulated to initiate at about +200 with respect to the 5' end of the 1.3 kb RNA, encoding a truncated polypeptide lacking the first 45 residues of the polypeptide. Thus the putative product of the 1.1 kb RNA would correspond to the 39K TK polypeptide made in small amounts during lytic viral infections (12). As only relatively low levels of the 1.1 kb RNA suffice to confer the TK^+ phenotype, Roberts and Axel proposed that its protein product retains close to full catalytic activity. This proposal is fully consistent with the finding that a mutant virus that initiates transcription at +198 induces 40% of the wild-type levels of TK activity during lytic infections (Haar, Smiley, Marsden and Preston, unpublished). Although Roberts and Axel did not directly establish the structure of the 1.1 kb RNA, recent work in my laboratory has shown that, in one cell line transformed by a functionally promoter-less TK fragment, a presumably analogous transcript initiates at +200 (Dennis and Smiley, submitted). This result therefore supports the scheme outlined above.

In summary, the data of Roberts and Axel suggest that in addition to the bona fide TK promoter, at least two other promoter sites are

potentially able to drive TK expression in transformed cells. One of these is a relatively strong promoter, which however results in an only marginally active, truncated TK polypeptide; the other is a much weaker promoter, which generates an active although truncated enzyme. When TK DNA is present at a low copy number, neither of these two sites provide a sufficiently high level of enzymatic activity to convert cells to TK^+. However, if the template is present in high numbers, then either promoter site is able to drive sufficient levels of activity. This model is capable of accounting for much of the available data on transformation of cells by TK genes bearing promoter mutations. It predicts that cells transformed by promoter mutants will contain many copies of the mutant gene, and will express TK from one of these two alternative promoters. Further, since the acquisition of many copies should be a rarer event than the acquisition of only a few, the model also accounts for the reduced transforming activity of mutant genes. Consistent with these predictions, both Zipser et al. (48) and El Kareh and Silverstein (personal communication) have found an inverse correlation to exist between the transforming activity of TK genes bearing promoter mutations and the transforming DNA copy number found in the resulting cell lines. El Kareh and Silverstein also demonstrated that cell lines established with a variety of mutant genes lacking the -105 to +56 region synthesized a 1.1 kb transcript. In addition, Dennis and Smiley (submitted) found a comparable transcript initiating at +200 in a high copy number cell line established by a mutant gene lacking sequences upstream of +56. These data all support the scheme outlined above. However, Smiley et al. (49) found that some cell lines transformed by impaired genes do not contain a large number of mutant gene copies. It therefore seems likely that at least some impaired templates can be expressed by alternative mechanisms, which do not involve a high template number. Perhaps some promoter defects are partially alleviated by insertion at specific chromosomal sites. It seems likely that a systematic study of cell lines transformed by a variety of TK promoter mutants would help to test this suggestion.

5.3. Transactivation of TK Expression by Proteins

Following the first demonstration by Munyon and colleagues (2) that the HSV TK gene could biochemically transform TK^- cells to TK^+, this same group of investigators asked whether or not the viral TK gene resident in

the transformed cells remained competent to respond to the viral-coded factors which activate its expression during lytic infection. They found that the TK activity in transformed cells was increased substantially by superinfection with a TK-deficient mutant HSV-1 (50). The results of these and other studies (51, 52, 22) established that this induction resembles the activation of TK expression from the viral chromosome in several ways. Induction occurs with β kinetics, and requires α protein synthesis; furthermore it is blocked by ts mutations in the ICP4 gene of the superinfecting virus. Interestingly, in some cases even silent resident TK genes are activated by superinfection (51, 52). Recently, it has been shown directly that induction occurs at the level of TK mRNA: the increase in TK enzymatic activity observed following infection is paralleled by a comparable increase in the amount of TK mRNA derived from the resident TK gene, and the induced transcript initiates at the same site used during normal viral infections (El Kareh and Silverstein, personal communication; Halpern and Smiley, unpublished). These data are all consistent with the idea that induction reflects α-mediated transcriptional activation of the resident TK gene, implying that the transfected TK DNA sequences retain cis-acting DNA sequences which are necessary for the respose to the trans-acting proteins. These views have received considerable support from the results of recent studies.

A number of groups have contibuted to the identification of the TK DNA sequences needed for the response to the inducing signal supplied by α proteins. Most of these studies have employed a similar experimental design: various mutant TK genes prepared *in vitro* were introduced into TK^- cells by transformation, and the resulting transformants were then scored for inducibility by superinfecting HSV. The results are largely consistent with one another, and when taken in combination suggest that induction requires the integrity of at least two sets of upstream sequences, extending from -109 to about -47. These regions are largely coincident with the two distal elements of the constitutive TK promoter (diagrammed in Figure 2). The data supporting these assignments are as follows:

1. Genes retaining only 109 5'-flanking bases remain fully inducible, while those deleted up to -95 respond poorly (El Kareh and Silverstein, personal communication; McKnight, personal communication). This finding is consistent with earlier published data showing that genes

retaining 200 or 180 5'-flanking bases are inducible, while those truncated to -60 and -6 are not (6, 48, 49).

2. Sequences extending from -80 to -70 are dispensable for induction while sequences located somewhere between -70 and -12 are required (49). Since an XhoI linker insertion at -47 abolished induction (48), some of the relevant sequences in the -70 to 12 region are close to or span -47.

3. Recently McKnight (personal communication) has shown that the -109 to -47 region, when placed upstream of the Hogness sequence of another gene, confers α inducibility on that gene. Thus the -109 to -47 region appears to be sufficient, as well as necessary, for induction. On the surface, this result seems inconsistent with those of Zipser et al. (48), who showed that an Xho I linker at -47 eliminated induction. One possible way to reconcile these two findings is to propose that residue -47 lies within, but close to the 3' boundary of, the important sequences. The construction of McKnight is bounded by a BamHl linker, starting at -47. Perhaps the BamHl linker is able to partially substitute for the natural residues downstream of -47, while an XhoI linker is not.

4. In addition to contributing to the mapping of the induction sequences, the data of Zipser et al. (48) also demonstrated that inducibility can be dissociated from constitutive promoter function. A linker insertion at -6 drastically reduced constitutive expression, but did not affect inducibility. Conversely, a linker insertion at -47 eliminated inducibility without markedly affecting constitutive expression.

The most striking feature of these findings is that the induction sequences seem to be largely coincident with the two distal elements of the TK promoter site. This is in contrast to the situation with HSV promoters, where the regulatory site is located upstream of all of the constitutive promoter elements. The data therefore provide very strong evidence that induction of TK occurs at the level of transcription initiation, rather than, for example, at the level of RNA processing or transport. Because the distal regions of the constitutive TK promoter seem to be involved in setting the overall level of transcription initiation, an attractive possibility is that α proteins activate TK by boosting the activity of this region. How might α proteins accomplish this? I can think of at least four distinct possibilities which are

consistent with the available data. In considering these possibilities I will assume that ICP4 alone among α proteins is sufficient for promoter activation. Although this seems to be the case for the HSV VP5 promoter (Smiley, Persson and Bacchetti, unpublished) it remains to be directly established for TK.

1. ICP4 might recognize and bind to specific sequences in the -109 to -47 region of the TK promoter. The bound protein might then facilitate polymerase loading, either by protein-protein interactions, or by inducing a favourable conformational change in the DNA. Although partially purified ICP4 does not appear to bind DNA, the protein present in crude extracts does (53), raising the possibility ICP4 binds to DNA in cooperation with other proteins. If ICP4 does act by binding to specific sites then one might expect to find conserved ICP4 recognition sequences in the promoter regions of all non-α HSV genes. Further, since ICP4, and its Pseudorabies virus counterpart, appears to also activate some adenovirus early genes, (54, S. Bachenheimer, personal communication) one would also expect to find similar sequences in some adenovirus promoters as well. It is not possible to test these predictions by comparative sequence analysis alone, for several reasons. Most importantly, it is not yet clear which of the sequences in the -109 to -47 region of the TK promoter are specifically involved with recognizing the signal. It is possible that induction requires the integrity of two classes of sequence in this region: those that are required for the proper functioning of the upstream region, and those that are specifically involved in the response of the region to the signal. The results of Zipser et al. support this view (48). Complicating matters further, these two classes of signal might overlap in a complex way. Second, because the spacing between the three elements of the constitutive TK promoter is relatively flexible, one might expect a similar flexibility to exist in the position of the putative ICP4 recognition sequence. Third, even if ICP4 binds to specific sites, it is not obvious that this would necessarily involve recognition of all, or even a majority of the bases in the binding site. Consequently, binding sites might show only patchy homology with one another. Nevertheless, inspection of several HSV and adenovirus promoters has revealed some interesting homologies with one or the other of the two distal elements of the TK promoter region. For example, residues -104 to -91 of the second distal region match residues -68 to

-55 in the promoter of the β HSV gene encoding a 38K polypeptide (55), at 11 of 14 positions. Residues -56 to -47 of the first distal region of TK match two overlapping sites in the promoter for the β-γ VP5 gene (27) (-47 to -35 and -41 to -28) at 10 of 12 positions. The first distal element (-56 to -46) is also homologous to positions -47 to -37 in the adenovirus type 2 E1 B promoter at 9 of 11 sites. I conclude that it is possible that ICP4 acts in a sequence-specific fashion.

2. ICP4 might interact with components of the cellular transcription machinery, altering its promoter preferences. The main differences between this possibility and the first are that ICP4 is not postulated to recognize specific DNA sequences, nor to bind directly to DNA. However, one might still expect the altered transcription machinery to show some sequence specificity in its altered promoter recognition spectrum.

3. ICP4 might act very indirectly, for example by altering the template conformation in a sequence independent fashion. This altered conformation might be required for the efficient use of HSV promoters by an unaltered host transcription machinery. This view is consistent with the effects of template conformation on TK expression in Xenopus oocytes (35). According to this model, the induction sequences identified by mutation would correspond to those promoter elements whose function is sensitive to changes in the template conformation.

4. ICP4 might act to overcome cellular negative control of HSV transcription. For example, a cellular repressor might block HSV transcription; perhaps ICP4 inactivates the repressor, allowing transcription to proceed. A similar model has been proposed to account for the activation of adenovirus early transcription by the immediate-early E1A gene products (56). In my view, this model is unlikely to be correct because of some of the effects of mutation on induction. For example, the XhoI linker insertion at -47 isolated by Zipser et al. abolishes induction without affecting constitutive expression. I can see no straightforward way of accomodating this result with the repressor model.

In summary, although the activation of TK expression by ICP4 requires the integrity of sequences largely overlapping the distal TK promoter elements, the precise mechanism by which ICP4 exerts its efects remains unknown. The solution to this problem is of general interest, and will likely hinge on studies of the biochemical properties of the

ICP4 protein, and its effects (if any) on *in vitro* transcription systems.

ACKNOWLEDGEMENTS

The work of the author's laboratory cited here was supported by the Medical Research Council of Canada and the National Cancer Institute of Canada. I thank Ms. J. Maljar for her rapid and efficient preparation of this manuscript.

REFERENCES

1. Dubbs D, Kit S: Mutant strains of herpes simplex deficient in thymidine kinase-inducing activity. Virology (22): 493-502, 1964.
2. Munyon W, Kraiselburd E, Davis D, Mann J: Transfer of thymidine kinase to thymidine kinaseless L cells by infection with ultraviolet-irradiated herpes simplex virus. J Virol (23): 234, 1971.
3. Wigler M, Silverstein S, Lee L, Pellicer A, Cheng Y, Axel R: Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell (11): 223-232, 1977.
4. Colbere-Garapin F, Chousterman S, Horodniceau F, Kourilsky P, Garapin A: Cloning of the active thymidine kinase gene of herpes simplex virus type 1 in *E. coli* K12. Proc Natl Acad Sci USA (76): 3755-3759, 1979.
5. McKnight S, Gavis E: Expression of the herpes thymidine kinase gene in Xenopus laevis oocytes: an assay for the study of deletion mutants constructed in vitro. Nucl Acids Res (8): 5931-5948, 1980.
6. Wilkie N, Clements J, Boll W, Mantei N, Lonsdale D, Weissman C: Hybrid plasmids containing an active thymidine kinase gene of herpes simplex virus - 1. Nucl Acids Res (7): 859-877, 1979.
7. McKnight S: The nucleotide sequence and transcript map of the herpes simplex virus thymidine kinase gene. Nucl Acids Res (8): 5949-5964, 1980.
8. Wagner M, Sharp J, Summers W: Nucleotide sequence of the thymidine kinase gene of herpes simplex virus type 1. Proc Natl Acad Sci USA (78): 1441-1445, 1981.
9. Smiley J, Wagner M, Summers W, Summers W: Genetic and physical evidence for the polarity of transcription of the thymidine kinase gene of herpes simplex virus. Virology (102): 83-93, 1980.
10. Proudfoot N, Brownlee G: Sequence at the 3' end of globin mRNA shows homology with immunoglobin light chain mRNA. Nature (252): 359-362, 1974.
11. Fitzgerald M, Shenk T: The sequence 5'-AAUAAA-3' forms part of the recognition site for polyadenylation of late SV40 mRNAs. Cell (24): 251-260, 1981.
12. Preston C, McGeoch D: Identification and mapping of two polypeptides encoded within the herpes simplex virus type 1 thymidine kinase gene sequences. J Virol (38): 593-605, 1981.
13. Marsden H, Haar L, Preston C: Processing of herpes simplex virus proteins and evidence that translation of thymidine kinase mRNA is initiated at three separate AUG codons. J Virol (46): 434-445, 1983.
14. Constanzo F, Campadeli-Fiume G, Foa-Tomasi L, Cassai E: Evidence that herpes simplex virus DNA is transcribed by cellular RNA polymerase B. J Virol (21): 996-1001, 1977.
15. Honess R, Roizman B: Regulation of herpesvirus macromolecular

synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. J Virol (14): 8-19, 1974.
16. Watson R, Preston C, Clements J: Separation and characterization of herpes simplex virus type 1 immediate-early mRNA's. J Virol (31): 42-52, 1979.
17. Mackem S, Roizman B: Regulation of herpesvirus macromolecular synthesis: transcription-initiation sites and domains of α genes. Proc Natl Acad Sci USA (77):7122-7126, 1980.
18. Anderson K, Costa R, Holland L, Wagner E: Characterization of herpes simplex virus type 1 RNA present in the absence of de novo protein synthesis. J Virol (34): 9-27, 1980.
19. Watson R, Clements J: Characterization of transcription-deficient temperature-sensitive mutants of herpes simplex virus type 1. Virology (91): 364-369, 1978.
20. Preston C: Control of herpes simplex virus type 1 mRNA synthesis in cells infected with wild-type virus or the temperature-sensitive mutant tsK. J Virol (29): 275-284, 1979.
21. Watson R, Clements J: A herpes simplex virus type 1 function continuously required for early and late virus RNA synthesis. Nature (285): 329-330, 1980.
22. Post L, Mackem S, Roizman B: Regulation of α genes of herpes simplex virus: expression of chimeric genes produced by fusion of thymidine kinase with α gene promoters. Cell (24): 555-565, 1981.
23. Mackem S, Roizman B: Differentiation between α promoter and regulator regions of herpes simplex virus 1: the functional domains and sequence of a moveable α regulator. Proc Natl Acad Sci USA (79): 4917-4921, 1982.
24. Batterson W, Roizman B: Characterization of the herpes simplex virion-associated factor responsible for the induction of α genes. J Virol (46): 371-377, 1983.
25. Mackem S, Roizman B: Regulation of genes of herpes simplex virus: the α 27 gene promoter-thymidine kinase chimera is positively regulated in coverted L cells. J Virol (43): 1015-1023, 1982.
26. Mackem S, Roizman B: Structural features of the herpes simplex virus α gene 4, 0, and 27 promoter-regulatory sequences which confer α regulation on chimeric thymidine kinase genes. J Virol (44): 939-949, 1982.
27. Frink R, Draper K, Wagner E: Uninfected cell polymerase efficiently transcribes early but not late herpes simplex virus type 1 mRNA. Proc Natl Acad Sci USA (78): 6139-6143, 1981.
28. Holland L, Anderson K, Shipman C, Wagner E: Viral DNA synthesis is required for the efficient expression of specific herpes simplex virus type 1 mRNA species. Virology (101): 10-24, 1980.
29. Frink R, Eisenberg R, Cohen G, Wagner E: Detailed analysis of the portion of the herpes simplex virus type 1 genome encoding glycoprotein C. J Virol (45): 634-647, 1982.
30. Garfinkle B, McAuslan B: Regulation of herpes simplex virus-induced thymidine kinase. Biochem Biophys Res Comm (58): 822-829, 1974.
31. Leung W: Evidence for a herpes simplex virus-specific factor controlling the transcription of deoxypyrimidine kinase. J Virol (27): 269-274, 1978.
32. Leung W, Dimock K, Smiley J, Bacchetti S: Herpes simplex virus thymidine kinase transcripts are absent from both nucleus and cytoplasm during infection in the presence of cycloheximide. J Virol (36): 361-365, 1980.
33. Sharp J, Wagner M, Summers W: Transcription of herpes simplex virus

genes in vivo: Overlap of a late promoter with the 3' end of the early thymidine kinase gene. J Virol (45): 10-17, 1983.

34. Bacchetti S, Graham F: Transfer of the gene for thymidine kinase to thymidine kinase-deficient human cells by purified herpes simplex viral DNA. Proc Natl Acad Sci USA (74): 1590-1594, 1977.
35. Harland R, Weintraub H, McKnight S: Transcription of DNA injected into Xenopus oocytes is influenced by template topology. Nature (302): 38-43, 1983.
36. Smiley J, Steege D, Juricek D, Summers W, Ruddle F: A herpes simplex virus 1 integration site in the mouse genome defined by somatic cell genetic analysis. Cell (15): 455-468, 1978.
37. Pellicer A, Wigler M, Axel R, Silverstein S: The transfer and stable integration of the HSV thymidine kinase gene into mouse cells. Cell (14): 133-141, 1978.
38. Luciw R, Bishop J, Varmus H, Capecchi M: Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. Cell (33): 705-716, 1983.
39. Corden J, Wasylyk B, Buchwalder A, Sassone-Corsi P, Kedinger D, Chambon P: Promoter sequences of eukaryotic protein-coding genes. Science (209): 1406-1413, 1980.
40. Grosschedl R, Birnstiel M: Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutations in vivo, Proc Natl Acad Sci USA (77): 1432-1436, 1980.
41. Benoist C, O'Hare K, Breathnach R, Chambon P: The ovalbumin gene-sequence of putative control regions. Nucl Acids Res (8): 127-142, 1980.
42. McKnight S, Gavis E, Kingsbury R, Axel R: Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. Cell (25): 385-398, 1981.
43. McKnight S, Kingsbury R: Transcriptional control signals of a eukaryotic protein-coding gene. Science (217): 316-324, 1982.
44. McKnight S: Functional relationships between transcriptional control signals of the thymidine kinase gene of herpes simplex virus. Cell (31): 355-365, 1982.
45. Read G, Summers W: In vitro transcription of the thymidine kinase gene of herpes simplex virus. Proc Natl Acad Sci USA (79): 5215-5219, 1982.
46. Roberts J, Axel R: Gene amplification and gene correction in somatic cells. Cell (29): 109-119, 1982.
47. Ostrander M, Vogel S, Silverstein S: Phenotypic switching in cells transformed with the herpes simplex virus thymidine kinase gene. Mol Cell Biol (2): 708-714, 1982.
48. Zipser D, Lipsich L, Kwoh J: Mapping functional domains in the promoter region of the herpes thymidine kinase gene. Proc Natl Acad Sci USA (78): 6276-6280, 1981.
49. Smiley J, Swan H, Pater M, Pater A, Halpern M: Positive control of the herpes simplex virus thymidine kinase gene requires upstream DNA sequences. J Virol (47): 301-310, 1983.
50. Lin S, Munyon W: Expression of the viral thymidine kinase gene in herpes simplex virus-transformed cells. J Virol (14): 1199-1208, 1974.
51. Leiden J, Buttyan R, Spear P: Herpes simplex virus gene expression in transformed cells. I. Regulation of the viral thymidine kinase

gene in transformed L cells by products of superinfecting virus. J Virol (20): 413-424, 1976.
52. Kit S, Dubbs D: Regulation of herpesvirus thymidine kinase activ in LM (TK^-) cells transformed by ultraviolet light-irradiated her simplex virus. Virology (76): 331-340, 1977.
53. Freeman M, Powell K: DNA-binding properties of a herpes simplex virus immediate-early protein. J Virol (44): 1084-1087, 1982.
54. Feldman L, Imperiale M, Nevins J: Activation of early adenovirus transcription by the herpesvirus immediate early gene: Evidence for a common cellular control factor. Proc Natl Acad Sci USA (79): 4952-4956, 1982.
55. Draper K, Frink R, Wagner E: Detailed characterization of an apparently unspliced herpes simplex virus type 1 gene mapping in the interior of another. J Virol (43): 1123-1128, 1982.
56. Nevins J: Mechanisms of activation of early viral transcription by the adenovirus E1A gene product. Cell (26): 213-230.

7

USE OF CLONED EPSTEIN-BARR VIRUS DNA TO IDENTIFY GENES THAT DETERMINE THE FATE OF VIRAL INFECTION

JEFF SAMPLE, GERALD LANCZ, AND MEIHAN NONOYAMA

SUMMARY

Transcription of the Epstein-Barr virus (EBV) genome during different cell-virus interactions was analyzed to identify the loci containing genes which may be responsible for determining the fate of an EBV infection. Transcriptionally active regions of the genome were detected by hybridization of ^{32}P-labeled cDNA (reverse-transcribed from infected-cell RNA) to the different cloned Bam Hl restriction endonuclease fragments of EBV DNA. Analysis of the transcription in restringently infected cells indicated that, in the absence of complete virus replication, transcription of the viral genome is limited to the W-Y-H region of the Bam Hl restriction map. During permissive or productive infection, transcription of approximately 90% of the viral DNA was detected. Analysis of the immediate-early transcription and transcription kinetics during permissive infection indicated that the Bam Hl M region is transcribed first following the onset of productive replication. Transcription of this region was not detected in restringently infected cells or during primary infection of adult human B lymphocytes. This suggests that expression of a gene(s) within the Bam Hl M region is required to initiate the productive cycle of virus replication. Suppression of this gene may therefore be a prerequisite to immortalization of B lymphocytes by EBV.

INTRODUCTION

Human or nonhuman-primate B lymphocytes which have been infected in vitro with Epstein-Barr virus (EBV) acquire the

Y. Becker (ed.), VIRAL MESSENGER RNA.

ability to proliferate indefinitely in cell culture, a process referred to as growth transformation or immortalization (1-3). Most lymphoblastoid-cell lines which have been established following EBV infection are restrictive to virus replication. Cell lines have been established, however, in which 1 to 5% of the cells spontaneously produce EBV particles. These cell lines have been designated as permissively infected, although the majority of the cells remain nonpermissive to virus replication. Cell lines which are completely nonpermissive to EBV replication are referred to as latently or restringently infected.

During the immortalization process, the EBV genome becomes established within the host cell so that a relatively specific number of genomes is maintained within the cells of a given cell line (4-10). The DNA genome within the virion itself is a linear double-stranded molecule approximately 170 x 10^3 base pairs (bp) in length, with a molecular weight of about 110 megadaltons (Md) (11-13). However, within the immortalized lymphocyte the genome exists as circular plasmid DNA (14,15). Several reports have suggested that in some instances the EBV genome has integrated into host cell chromosomes (16-18).

Because the complete EBV genome is maintained and propagated within established lymphoblastoid cells, utilization of these cells as a source of virus and virus DNA has greatly benefited studies of genome structure and expression. Two such cell lines which have been extensively employed in these studies are B95-8 and P3HR-1 cells. A small percentage of these immortalized lymphocytes spontaneously produce either the B95-8 or the P3HR-1 strain of EBV, respectively. In addition, the B95-8 strain of EBV has the ability to immortalize human and nonhuman-primate B lymphocytes following infection *in vitro*, and is therefore considered a prototype strain of EBV. P3HR-1 virus, however, has lost its ability to immortalize lymphocytes. Structural analysis of the genomes of these two strains of EBV has revealed that a deletion of approximately 6.5 x 10^3 bp has occurred in the P3HR-1 genome (19,20). A hypothesis has been put forth that the DNA which has been deleted from the P3HR-1 genome encodes a gene product(s) that enables prototype strains of EBV to

immortalize B lymphocytes. Another consequence of the P3HR-1 DNA deletion is that it apparently confers an ability to superinfect and induce permissive infection in cells that are restringently infected, i.e., the Raji cell line (21,22). Prototype strains of EBV, such as B95-8, are unable to induce permissive infection in Raji cells.

The regulatory events that determine whether a cell is permissive or nonpermissive to EBV replication, or the control mechanisms associated with the EBV establishment of lymphoblastoid-cell lines, are questions as yet unresolved. This is in part due to the biology of EBV which poses various technical barriers to these studies. The study of the permissive state of EBV replication employs cell lines such as B95-8 and P3HR-1. Even though 1 to 5% of these cells are undergoing a permissive cycle of infection, 95 to 99% remain nonpermissive to EBV replication. Therefore, large numbers of cells must be processed to obtain sufficient quantities of virus or virus-producing cells for experimentation. The replication efficiency of these cells can be augmented to a limited extent by the addition of 12-0-tetradecanoylphorbol-13-acetate (TPA) to the cell cultures. In the presence of TPA, 20 to 40% of the cells undergo permissive infection (23). The problem imposed by the biological properties of EBV infections has been overcome, to a great extent, by the elucidation of restriction endonuclease maps of the EBV genome and the cloning of the viral DNA in bacterial plasmids (24,25).

In this chapter, we report on the work underway in our laboratories which intends to examine the expression of the EBV genome at the transcriptional level. We have used cloned restriction endonuclease fragments of the EBV genome in hybridization experiments and have been able to identify regions of the genome that are transcribed in cells that represent different types of infection, e.g., permissive vs. restringent infection, and during the different phases of virus replication. Our data show that transcription of specific regions of the EBV genome is consistently associated with the nature or type of EBV infection, i.e., a restringent or permissive infection.

CELLS AND VIRUSES

Culture conditions for the maintenance of B95-8, P3HR-1, and Raji cells have been described elsewhere (26). Fresh human B lymphocytes were isolated from peripheral blood leukocytes on a Ficoll hypaque density gradient. Adherent cells were removed by incubation for 1 hour in a plastic tissue culture flask, and cells forming rosettes with sheep erythrocytes were removed by a second centrifugation through a Ficoll hypaque gradient. Fresh B lymphocytes (nonadherent, nonrosetting cells) were cultured in RPMI 1640 medium containing glutamine and 10% fetal bovine serum. Production of B95-8 and P3HR-1 EBV strains was accomplished by culturing each cell line with medium containing TPA* (20 ng/ml) for 3 days. Typically, 10 l of culture supernatant was filtered through a 0.45 μm filter and centrifuged at 18,000 x g for 2 hours at 4°C to pellet the virus particles. The virus pellets were resuspended to 1/300 the original culture volume of RPMI 1640 medium containing 15% newborn calf serum and were stored at -80°C. Stock preparations of B95-8 and P3HR-1 virus normally had titers of 5×10^7 transforming units (T.U.) or early antigen-inducing units (E.A.U.) per ml, respectively.

EXPERIMENTAL APPROACH

The operational regulatory mechanisms which control expression of the EBV genome can be exemplified by the manner in which the genome is transcribed during various cell-virus interactions, i.e., restringent vs. permissive infection. Therefore, by monitoring transcription of the EBV genomes in permissively and restringently infected cells, transcriptional events which reflect the regulation of genome expression can be compared and evaluated. This is the approach we have employed to determine the genomic loci which are transcribed during different types of EBV infection and identify those which may contain genes responsible for regulating different cell-virus interactions. In addition, analysis of the transcriptional events in fresh human B lymphocytes infected in vitro with the B95-8 strain of EBV was done to determine the transcriptional

*12-O-tetradecanoylphorbol-13-acetate

pattern of the genome during immortalization.

The principal method used to detect EBV-specific transcripts in EBV-established cell lines and in EBV-infected cells is a modification of the dot-blot hybridization procedure (see legend, Fig. 2), described by Kafatos, et al. (27). The Bam H1 restriction endonuclease fragments of B95-8 EBV DNA were cloned in pBR322 by Dr. James Skare (25), and were subsequently obtained from him. Cloned EBV DNA fragments were propagated in E. coli strain HB101. The Bam H1 restriction endonuclease map of the B95-8 genome is illustrated in Figure 1.

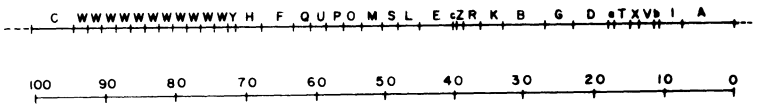

FIGURE 1. Bam H1 restriction endonuclease map of the B95-8 EBV genome as determined by Skare and Strominger (25). Scale is in megadaltons.

EXPERIMENTAL RESULTS

Transcription of the EBV genome during different states of virus replication

Transcription in restringently infected cells. Some cells, e.g., Raji cells, which were established from a biopsy of a Burkitt lymphoma patient, remain completely restrictive to complete replication of endogenous EBV. Expression of EBV-specific early antigen (EA) and viral capsid antigen (VCA), viral DNA synthesis, and production of EBV particles do not occur in these cells. These cells do express the Epstein-Barr virus-associated nuclear antigen (EBNA) and, therefore, are considered to represent a latent state of infection. However, previous studies which used DNA-RNA hybridization kinetic techniques indicated that 25% of the EBV DNA is transcribed

in Raji cells (26). This also proved true for the somatic cell hybrids, D98/Raji and D98/HR-1. This finding suggests that even in the absence of complete virus replication, limited expression of the EBV genome does occur. The term "restringent infection" is therefore a more acceptable description of this particular cell-virus interaction.

Specific regions of the EBV genome transcribed in restringently infected cells were identified in experiments using dot-blot hybridization. For these studies, cytoplasmic poly(A)+ and total nuclear RNA were isolated from Raji cells. ^{32}P-labeled cDNA hybridization probes were synthesized by reverse transcription in vitro of RNA and were hybridized to dot-blot filters (see legend, Fig. 2). Data from these experiments (not shown) indicate that the Bam Hl W, Y, and H regions (see Fig. 1) of the genome are most actively transcribed during restringent infection. These data suggest that gene products encoded within the Bam Hl W, Y, and H regions may, by one or more mechanism, be responsible for maintaining the immortalized state of the virus in the host cell. It is interesting to note that transcripts which map in the Bam HI W region (the large internal repeat region) have not been detected in the cytoplasmic poly(A)+ RNA (mRNA) fraction, suggesting that these transcripts do not function as mRNA.

Transcription in permissively infected cells. Several EBV-producing lymphoblastoid-cell lines have been employed for the in vitro production of virus and for the study of viral gene expression during permissive infection. The two most commonly studied have been the B95-8 and P3HR-1 cell lines. At a given time, 1 to 5% of these cells express EBV-specific EA and VCA indicating that virus replication is occurring in this fraction of the cell population. The fraction of cells undergoing a permissive infection was substantially increased by culturing the cells in the presence of TPA for 3 days prior to extracting the cells for RNA. The dot-blot hybridization study which reflects the transcriptional activity in TPA-induced B95-8 cells, is presented in Figure 2. These data indicate that most of the Bam Hl EBV DNA fragments are transcriptionally active during this

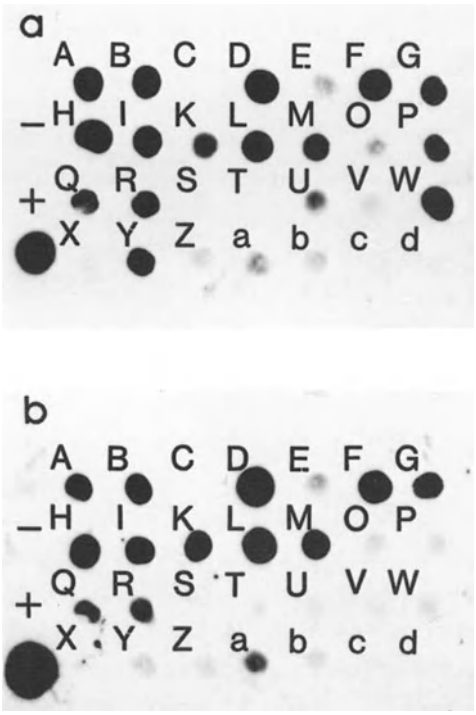

FIGURE 2. Transcription of the EBV genome in permissively infected B95-8 cells as indicated by dot-blot hybridization. ^{32}P-labeled cDNA was reverse-transcribed in vitro from 5 μg of total nuclear RNA or cytoplasmic poly(A)+ RNA which had been extracted from TPA-treated B95-8 cells. The cDNA probes (1-5 x 10^7 total CPM) were hybridized to cloned Bam Hl restriction endonuclease fragments of B95-8 DNA (A-d) fixed to nitrocellulose filters. Each filter contained the equivalent of 10 μg of EBV DNA, proportionally distributed according to the molecular weight of each fragment. pBR322 DNA (1 μg) and HeLa cell DNA (0.5 μg) served as negative (-) and positive (+) controls, respectively. Following hybridization, filters were processed by autoradiography. cDNA probes were transcribed from a) total nuclear RNA and b) cytoplasmic poly(A)+ RNA.

permissive infection. Transcripts mapping in the Bam Hl P, U, W, and Y regions of the genome were detected in the nuclear (Fig. 2a), but not in the cytoplasmic poly(A)+ (Fig. 2b), fraction. Transcripts that mapped in the Bam Hl C region of the genome were not detected in either the nuclear or the cytoplasmic poly(A)+ fraction, although previous studies did detect low levels of Bam Hl C-specific transcripts in these cells (28). Transcripts that mapped in the Bam Hl I, R, and Z regions were clearly detectable when B95-8 cells were cultured in the presence of TPA, relative to noninduced cells (28), suggesting that these transcripts were newly synthesized following TPA treatment.

Transcription of the EBV genome during the early and late phases of permissive infection

The complete (lytic) replication cycle of many viruses is mediated by a successive expression of viral genes. This succession of gene expression is stringently coordinated in such a manner that there is a specific order in which the viral genes must be expressed. For example, the genes of herpes simplex virus (HSV) have been divided into α, β, and γ genes, which are expressed in a cascade fashion, respectively, during the lytic replication of HSV (29). Genes of the α class are transcribed soon after infection and do not require virus-specific protein synthesis prior to expression. Transcription of the α genes, termed the immediate early genes, is therefore resistant to the action of chemicals which inhibit protein synthesis, e.g., cycloheximide. Transcription and expression of β, or early genes, is dependent upon prior expression of the α genes and virus-directed protein synthesis. Once expressed, products of the β genes act to turn off expression of the α genes and to turn on transcription of the γ, or late genes. In addition to virus-specific protein synthesis, γ gene expression is also dependent on viral DNA synthesis and is therefore sensitive to inhibitors of viral DNA synthesis, such as phosphonoacetic acid (PAA). Once the γ genes are expressed, transcription and expression of the β genes are turned off, assembly of the virion occurs, thus completing the replication cycle.

In the following section, we discuss studies which identify

the regions of the EBV genome which are transcribed during the early phases of permissive infection. In addition to TPA-treated B95-8 cells as a model for permissive infection, we have employed Raji cells which have been superinfected with the P3HR-1 strain of EBV. Although the Raji cells are in a restringent state of infection relative to the endogenous constituent EBV genomes they carry, these cells are permissive to superinfection by P3HR-1 EBV. This is indicated by the expression of EA and VCA, an inhibition of cellular DNA synthesis and an onset of viral DNA synthesis, and production of mature virus particles. The superinfection system has the distinct advantage in that a synchronized productive cycle of virus replication can be studied. In contrast, productive virus replication in B95-8 cells is an asynchronous process. Thus, employing the superinfection cell-virus system eliminates the inevitable mixture of early and late transcripts present in cell populations that are not synchronized with regard to the cycle of infection.

Previous studies examined persistent early and late transcripts in B95-8 cells that were cultured in the presence or absence of 100 μg of PAA per ml for 10 cell generations, and then treated with TPA for three days prior to cell harvest (30). Transcription which is sensitive to the action of PAA requires prior viral DNA synthesis and is classified as late transcription. The PAA-treated cells expressed EA but not the late viral antigen VCA. Transcription which occurs before the onset of viral DNA synthesis is resistant to PAA treatment and therefore includes the immediate early and early transcripts. PAA-resistant cytoplasmic poly(A)+ transcripts (Fig. 1, ref. 30) mapped in the Bam Hl A, B, F, G, H, L, and M regions of the EBV genome. Very weak hybridization was detected to the Bam Hl C, D, K, R, U, W, and a DNA fragments, indicating that a low level of transcription occurred from these regions as well. Hybridization of ^{32}P-labeled cloned EBV DNA to Southern blots of poly(A)+ RNA that had been extracted from TPA-induced B95-8 cells, indicated that at least 65 different EBV mRNAs are present in permissively infected cells. In

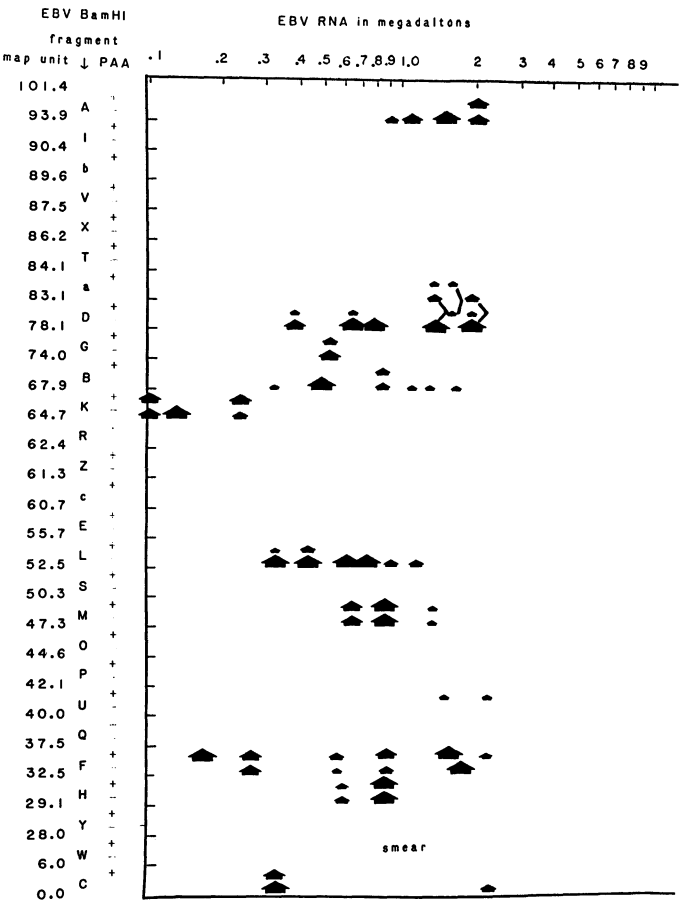

FIGURE 3. Summary of persistent early and late cytoplasmic poly(A)+ RNA in TPA-treated B95-8 cells (30). (Reprinted with permission).

the presence of PAA, the number of mRNA species was reduced to 25 (Fig. 3). The pattern of EBV-specific transcription in superinfected Raji cells, in the presence and absence of PAA, was essentially the same as in B95-8 cells (data not shown).

Immediate early transcripts were obtained from Raji cells superinfected with P3HR-1 EBV, which had been cultured in the presence of 5 μg of cycloheximide per ml from the time of

superinfection until cell harvest at 15 hr p.i. Under these conditions, protein synthesis was inhibited by at least 85%, as indicated by ^{35}S-methionine incorporation into acid insoluble protein. Cycloheximide concentrations greater than 5 μg/ml reduced cell viability significantly (less than 80%), resulting in poor recovery of RNA. Fluorescent antibody staining indicated that cycloheximide-treated cells were negative for EA and VCA. Dot-blot hybridization, used to analysize RNA isolated from superinfected Raji cells not treated with cycloheximide (Fig. 4a and 4b), revealed a pattern of transcription in the superinfected cells that is similar to that observed in TPA-induced B95-8 cells. In the presence of cycloheximide, transcription of the genome was only from the Bam Hl A, F, H, M, R, and W regions (Fig. 4c and 4d). Bam Hl W-specific transcripts were detected in the total nuclear RNA fraction but not in the cytoplasmic poly(A)+ fraction. This was also observed in permissively infected cells (Fig. 2). Longer autoradiographic exposure of dot-blot filters revealed that hybridization of the cDNA probe to other Bam Hl restriction endonuclease fragments did occur. This may be a reflection of the inability to inhibit protein synthesis greater than 85% with 5 μg of cycloheximide per ml. Thus, a degree of "leakiness" may be evident when protein synthesis is not reduced to a greater extent.

These results suggest that the immediate-early genes, expressed during productive replication of EBV, are located in the Bam Hl A, F, H, M, R, and W regions of the viral genome. To substantiate this and to determine which immediate-early region(s) of the genome is initially transcribed following initiation of productive replication, the kinetics of transcription in Raji cells superinfected with P3HR-1 EBV was determined. Transcription of the resident EBV genomes in Raji cells alone is limited to the Bam HI W, Y, and H regions (see section on restringent infection). When RNA isolated from Raji cells at 4, 8, and 12 hours following superinfection was analyzed by dot-blot hybridization, transcripts which mapped in the Bam Hl M region were the first predominant transcripts

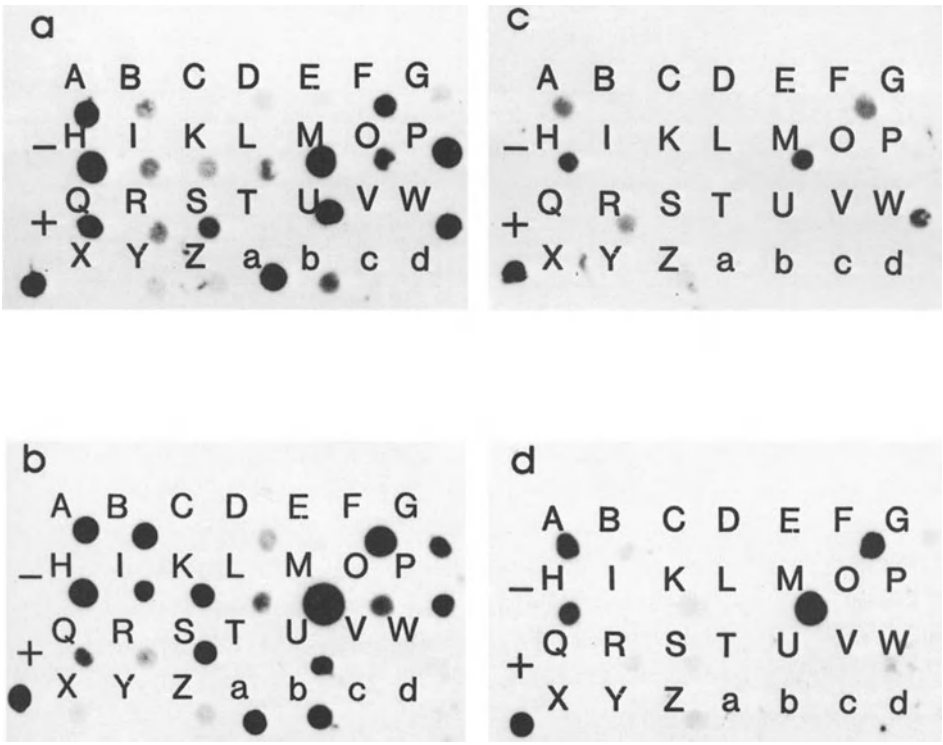

FIGURE 4. Analysis of immediate-early transcription in superinfected Raji cells. Raji cells were superinfected with P3HR-1 EBV at a M.O.I. of approximately 1.5 E.A.U. per cell, cultured in the presence or absence of 5 μg of cycloheximide per ml, and harvested 15 hr. p.i. for extraction of RNA. (a) and (b), dot-blot hybridizations in which cDNA probes were transcribed from total nuclear RNA and cytoplasmic poly(A)+ RNA, respectively, isolated from cells not treated with cycloheximide. (c) and (d), cDNA probes were transcribed from total nuclear RNA and poly(A)+ RNA, respectively, isolated from cells treated with cycloheximide.

detected (Fig. 5). Very low levels of or no transcription of the regions normally transcribed in restringently infected Raji cells was detected. When the kinetics of EA expression and genome transcription were compared, a positive correlation was found between the time course of EA expression and the transcription of the Bam Hl M region of the viral genome. Collectively, these data suggest that the Bam Hl M region of the EBV genome is the genomic locus that is initially transcribed during productive replication of EBV.

Transcription of the EBV genome during primary infection of B lymphocytes

Immortalization or growth transformation of human or nonhuman-primate B lymphocytes following infection in vitro with a prototype strain of EBV, e.g., B95-8, is the hallmark biological property of EBV. However, the P3HR-1 strain of EBV no longer has this capability. Detailed structural analysis of the P3HR-1 virus genome has revealed a deletion in the Bam Hl Y-H region of the genome (20,31). Therefore, it has been hypothesized that a gene product(s) encoded by the DNA in this region is responsible for lymphocyte immortalization. Consistent with this view is the finding that the Bam Hl W-Y-H region is the most actively transcribed region of the EBV genome during restringent infection. This hypothesis was tested by examining EBV-specific transcriptional events in fresh adult human B lymphocytes after they had been infected in vitro with the B95-8 strain of EBV.

Fresh B lymphocytes were infected with B95-8 EBV at a M.O.I. of approximately 1.5 T.U. per cell. Infected cells were harvested at 3 to 4 days p.i., RNA was extracted and analyzed by dot blot hybridization. The success of each infection was monitored by staining an aliquot of cells at 4, 7, and 14 days p.i. for EBNA.

Analysis of the dot-blot hybridization indicates that the Bam Hl F, P, and W regions of the EBV genome are transcribed following the infection in vitro of fresh B lymphocytes (Fig. 6). In some experiments, low levels of transcripts mapping in the Bam Hl A, H, and Y regions were also detected. It is apparent,

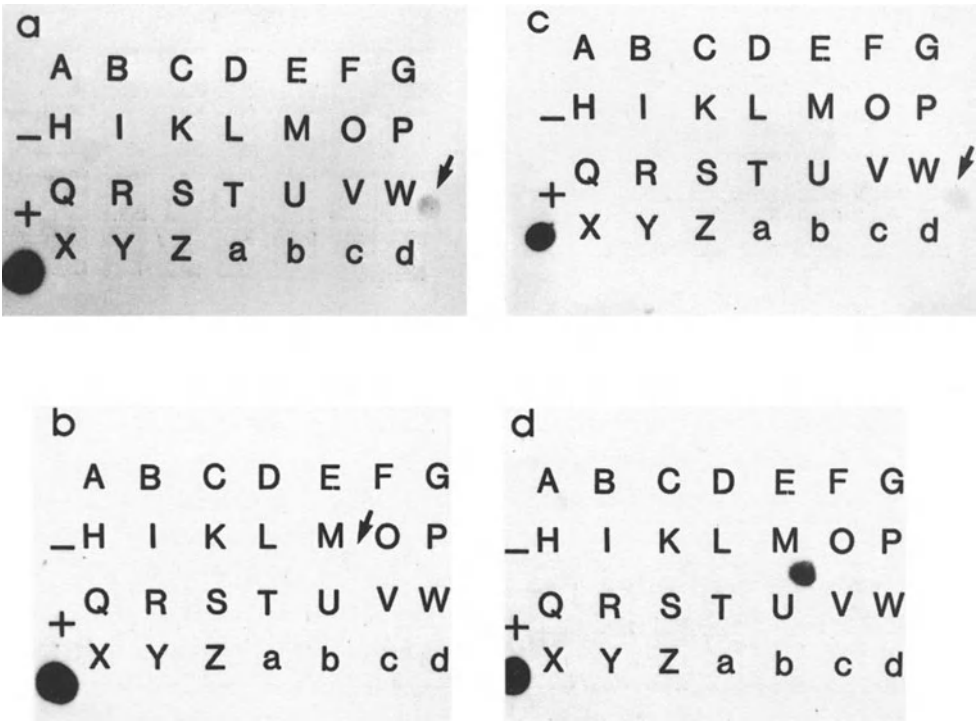

FIGURE 5. Kinetics of transcription of the EBV genome in Raji cells following initiation of productive infection by superinfection. Raji cells superinfected with P3HR-1 EBV were harvested at 8 and 12 hr. p.i. for analysis of RNA by dot-blot hybridization. (a) and (c), cDNA probes were transcribed from total nuclear RNA isolated from cells at 8 and 12 hr. p.i., respectively. (b) and (d), cDNA probes were transcribed from cytoplasmic poly(A)+ RNA isolated from cells at 8 and 12 hr. p.i., respectively. Arrows indicate regions of the filter in which a low level of hybridization was detected.

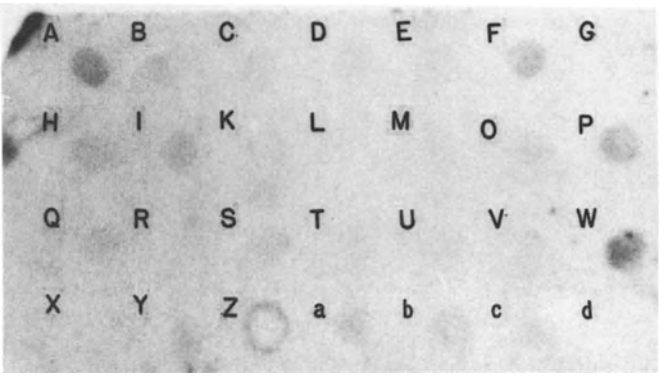

FIGURE 6. Transcription of the EBV genome during primary infection *in vitro* of adult human B lymphocytes. ^{32}P-labeled cDNA was transcribed from cytoplasmic poly(A)+ RNA and hybridized to cloned EBV DNA equivalent to 30 μg of total EBV DNA.

therefore, that regions of the genome apart from the Bam HI Y-H region are also transcribed during primary infection and are likely to be involved in cellular immortalization.

DISCUSSION

It is evident from the data we have presented that the EBV genome is transcribed differently in cells representing different types of virus infection. Analysis of dot-blot hybridization studies on RNA extracted during permissive infection of B95-8 cells, indicates approximately 90% of the EBV genome is transcribed during the productive cycle of virus replication. These data are similar to those reported by Kieff and co-workers (32-35). TPA treatment appeared to have induced transcription of the Bam Hl I, R, and Z regions of the genome, which was not detected in B95-8 cells untreated with TPA (28). Perhaps the expression of these regions is not required during productive replication of the virus.

EBV-specific transcription in restringently infected Raji cells occurred from approximately 25% of the total viral DNA

(12). Similar findings have also been reported for the restringently infected Namalwa and IB-4 cells, as well as Raji cells (32,33,35-38). The majority of the RNA transcripts from Raji cells map in the Bam Hl W, Y, and H regions of the EBV genome. This indicates that transcription of the EBV genome during restringent infection, relative to permissive infection, is highly restricted. The limited expression of the genome during restringent infection, as also reflected by the lack of EA and VCA expression, is probably required to maintain the cell in its transformed state.

Superinfection of Raji cells with the P3HR-1 strain of EBV results in a productive cycle of virus replication in these restringently infected cells. This is indicated by the induction of EA and VCA expression, initiation of viral DNA synthesis, and production of virus particles. When Raji cells are infected with the B95-8 prototype strain, permissive infection does not occur. In addition, analysis of RNA obtained from Raji cells superinfected with B95-8 EBV indicated no change in the transcription pattern following superinfection (data not shown).

Another characteristic of the P3HR-1 strain which differentiates it from prototype strains of EBV, is its inability to immortalize B lymphocytes following infection in vitro. Analysis of the structure of the P3HR-1 genome and genomes of prototype strains of EBV have shown that a 6.5×10^3 bp deletion in the Bam Hl Y-H region of the P3HR-1 genome has occurred (20). Although there is a 12×10^3 bp deletion in the Bam Hl I region of the B95-8 genome, the biological properties of this strain of EBV apparently are unaffected (39). Analysis of the genome of the virus produced by the Jijoye cell line, the parental cell line from which the P3HR-1 cell line was cloned, indicated no deletion in the Bam Hl Y-H region of the genome (40). Unlike the P3HR-1 strain of EBV, virus produced by Jijoye cells are capable of immortalizing B lymphocytes. Collectively, these findings suggest that a loss of DNA from the Bam Hl Y-H region of the genome is sufficient to compromise the immortalizing capability of EBV.

A gene product(s) encoded by the Bam Hl Y-H region of the EBV genome, therefore, may be responsible for cellular immortalization. This notion is supported by the data which show that the Bam Hl Y-H region is transcribed in B lymphocytes during primary infection in vitro and in restringently infected cells. If such a gene product is encoded by the Bam Hl Y-H region of the genome, it would explain why the P3HR-1 EBV cannot immortalize B lymphocytes. It is possible that a function of this gene product may be to repress transcription of a gene(s) responsible for initiating the productive cycle of EBV replication. Therefore, a productive cycle of virus replication ensues following superinfection of Raji cells with P3HR-1 EBV, because the gene(s) responsible for initiating productive replication is not repressed. The drawback to this conclusion, however, is that the Bam Hl Y-H encoded product expressed by the endogenous EBV genomes should repress productive virus replication following superinfection with P3HR-1 EBV.

Because superinfection of Raji cells with P3HR-1 EBV induces a synchronized productive cycle of virus replication, it has been possible to accurately analyze transcription of the EBV genome during the different phases of the replication cycle. The superinfection system has been especially useful in enabling us to identify regions of the genome which encode immediate-early transcripts during permissive infection. One or more of these regions most likely encodes a gene product(s) responsible for initiating the productive cycle of virus replication. Analysis of the RNA extracted from superinfected cells that had been cultured in the presence of cycloheximide to enhance for the presence of immediate-early transcripts, indicated that the Bam Hl A, F, H, M, R and W regions of the EBV genome encode immediate-early RNA. RNA which mapped in the Bam Hl W region of the genome was detected in the total nuclear RNA fraction but not in the cytoplasmic poly(A)+ RNA fraction. This was also true for Bam Hl W-specific transcripts detected in TPA-treated B95-8 cells. This suggests that RNA encoded by the Bam Hl W region does not function as

mRNA during permissive infection.

Data obtained by the analysis of the transcription kinetics during superinfection, suggest that the Bam Hl M region of the genome is the first region that is actively transcribed during productive replication. In addition, a positive correlation was observed when EA expression, which signals the onset of a productive cycle of replication, was compared with the transcription of the Bam Hl M region.

When the transcription patterns of the EBV genome in restringently infected Raji cells and B95-8 EBV-infected fresh adult human B lymphocytes are compared to the immediate-early transcription pattern of P3HR-1 superinfected Raji cells, all of the immediate-early regions except M are transcribed during immortalization and/or in established transformed cells. This suggests that expression of a gene(s) within the Bam Hl M region is required to initiate the productive cycle of virus replication and the suppression of M region expression may be a prerequisite to immortalization. It is also interesting to note that most of the immediate-early regions are expressed during immortalization. The mechanisms by which expression of the M region, and presumably productive replication, is restricted during restringent infection, may be either cellular or viral in nature. It is possible that a function of a gene product encoded by the Bam Hl Y-H region is to suppress expression of a gene or genes within the M region.

The actual mechanisms which control expression of the EBV genome during different types of virus infection are as yet unknown. The data obtained in our laboratories and those of others suggest that transcription of different regions of the genome is associated with the state of EBV infection in established lymphoblastoid cells. This apparently occurs in such a manner that either a latent or restringent infection is established, or alternatively, a productive cycle of virus replication predominates. Since the immediate-early transcripts are the first produced, they are believed to be important in the overall regulation of the cell-virus

interaction (29). In this regard, Pearson et al. (41) have mapped the EBV-EA complex to the Bam Hl M region of the genome. This finding supports our data which suggests that the Bam Hl M region contains immediate early genes that are expressed during the productive cycle of EBV infection.

In order to fully understand the biology of this virus, it will be necessary to characterize the functions of the EBV gene products that regulate the virus infection. The data presented here and that of others imply that the gene products encoded by the Bam Hl M and Y-H regions may have important roles in determining the outcome of an EBV infection of B lymphocytes.

REFERENCES

1. Henle, W., Diehl, V., Kohn, G., zur Hausen, H. and Henle, G. Science 157:1064-1065, 1967.
2. Pope, J., Horne, M. and Scott, W. Int. J. Cancer 3:857-866, 1968.
3. Jondal, M. and Klein, G. J. Exp. Med. 138:1365-1378, 1973.
4. zur Hausen, H. and Schulte-Holthausen, H. Nature (London) 227:245-248, 1970.
5. Nonoyama, M. and Pagano, J.S. Nature (London) New Biol. 233:103-106, 1971.
6. Nonoyama, M. and Pagano, J.S. Nature (London) 242:44-47, 1973.
7. Kawai, Y., Nonoyama, M. and Pagano, J.S. J. Virol. 12: 1006-1012, 1973.
8. Kieff, E. and Levine, J. Proc. Natl. Acad. Sci. U.S.A. 71:355-358, 1974.
9. Pritchett, R., Pedersen, M. and Kieff, E. Virology 74: 227-231, 1976.
10. Sugden, B., Phelps, M. and Domoradzki, J. J. Virol. 31: 590-595, 1979.
11. Pritchett, R.F., Hayward, S.D. and Kieff, E. J. Virol. 15:556-569, 1975.
12. Pritchett, R.F., Hayward, S.D. and Kieff, E. In: Oncogenesis and Herpesviruses (Eds. H. zur Hausen, G. de The' and M. Epstein), IARC Monograph, Lyon, 1975, pp. 171-191.
13. Hayward, S.D. and Kieff, E. J. Virol. 23:421-429, 1977.
14. zur Hausen, H., Diehl, V., Wolf, H., Schulte-Houlthausen, H. and Schneider, U. Nature (London) New Biol. 237:189-190, 1972.
15. Nonoyama, M. and Pagano, J.S. Nature (London) New Biol. 238:169-171, 1972.
16. Adams, A., Lindahl, T. and Klein, G. Proc. Natl. Acad. Sci. U.S.A. 70:2888-2892, 1973.
17. Tanaka, A. and Nonoyama, M. Proc. Natl. Acad. Sci. U.S.A. 71:4658-4661, 1974.

18. Heller, M., Henderson, A. and Kieff, E. Proc. Natl. Acad. Sci. U.S.A. 79:5916-5920, 1982.
19. Raab-Traub, N., Dambaugh, T. and Kieff, E. J. Virol. 27: 388-398, 1978.
20. Bornkamm, G.W., Hudewentz, J., Freese, U.K. and Zimber, U. J. Virol. 43:952-968, 1982.
21. Henle, W., Henle, G., Zajac, B., Pearson, G., Waubke, R. and Scriba, M. Science 169:188-190, 1970.
22. Yajima, Y. and Nonoyama, M. J. Virol. 19:187-194, 1976.
23. zur Hausen, H., O'Neill, F.J. and Freese, U. Nature (London) 272:373-375, 1978.
24. Dambaugh, T., Beisel, C., Hummel, M., King, W., Fennewald, S., Cheung, A., Heller, M., Raab-Traub, N. and Kieff, E. Proc. Natl. Acad. Sci. U.S.A. 77:2999-3003, 1980.
25. Skare, J. and Strominger, J.L. Proc. Natl. Acad. Sci. U.S.A. 77:3860-3364, 1980.
26. Tanaka, A., Nonoyama, M. and Glaser, R. Virology 82:63-68, 1977.
27. Kafatos, F.C., Jones, C.W. and Efstratiadis, A. Nucleic Acids Res. 7:1541-1552, 1979.
28. Shin, S., Tanaka, A. and Nonoyama, M. Virology 124:13-20, 1983.
29. Honess, R.W., Roizman, B. J. Virol. 14:8-19, 1974.
30. Shin, S., Donovan, J. and Nonoyama, M. Virology 124: 196-200, 1983.
31. King, W., Dambaugh, T., Heller, M., Dowling, J. and Kieff, E. J. Virol. 43:979-986, 1982.
32. Hayward, S.D. and Kieff, E. J. Virol. 18: 518-525, 1976.
33. Orellana, T. and Kieff, E. J. Virol. 22:321-330, 1977.
34. Hummel, M. and Kieff, E. J. Virol. 43:262-272, 1982.
35. Heller, M., van Santen, V. and Kieff. E. J. Virol. 44: 311-320, 1982.
36. Thomas-Powell, A.L., King, W. and Kieff, E. J. Virol. 29:261-274, 1979.
37. King, W., van Santen, V. and Kieff, E. J. Virol. 38: 649-660, 1981.
38. King, W., Thomas-Powell, A.L., Raab-Traub, N., Hawke, M. and Kieff, E. J. Virol. 36:506-518, 1980.
39. Raab-Traub, N., Dambaugh, T. and Kieff, E. Cell 22: 257-269, 1980.
40. Rabson, M., Gradoville, L., Heston, L. and Miller, G. J. Virol. 44:834-844, 1982.
41. Pearson, G.R., Vroman, B., Chase, B., Sculley, T., Hummel, M. and Kieff, E. J. Virol. 47:193-201, 1983.

8

ADENOVIRAL hnRNA IS ASSOCIATED WITH THE HOST NUCLEAR MATRIX DURING PROCESSING

W.J. VAN VENROOIJ, R. VERHEIJEN AND E.C. MARIMAN

SUMMARY

When nuclei are treated with detergents and the chromatin is degraded and removed by subsequent deoxyribonuclease and high salt treatments, a predominantly proteinaceous nuclear substructure, known as the nuclear matrix, remains. Newly synthesized RNA molecules are found to be tightly associated with this nuclear framework. When steady-state hnRNA in adenovirus-infected cells is analyzed with the S1 mapping procedure, one finds precursors, processing intermediates and products of mRNA processing to be associated with the nuclear matrix. This binding is probably mediated by proteins. In vivo crosslinking experiments and ribonuclease digestions suggest that the hnRNP group 'C' proteins might be involved in this association. Analysis of the RNA fragments complexed with the group 'C' proteins show that probably more than one of these polypeptides are associated with each molecule of RNA.

Detailed analysis of the splicing of the tripartite leader sequence indicates that several rapid cleavages in the intron regions take place before exon-exon ligation occurs. The unligated pieces of RNA are bound to the nuclear matrix and it is therefore possible that the nuclear matrix provides the structural framework that keeps the pieces of RNA in the correct spatial orientation for ligation.

INTRODUCTION

When nuclei are treated with detergents and the chromatin is degraded and removed by subsequent deoxyribonuclease (DNase) and high salt treatments, a predominantly proteinaceous nuclear substructure remains that is referred to as the nuclear matrix (for review see ref. 1). The isolated nuclear matrix consists of three morphologically distinguishable structural elements:

a. a peripheral layer, which represents the remainder of the nuclear

Y. Becker (ed.), VIRAL MESSENGER RNA.

envelope and contains pore complexes in association with a lamina;
b. residual nucleoli, and
c. internal grano-fibrillar structures.

The peripheral pore-complex-lamina has been isolated separately and its polypeptide composition has been determined (2, 3). In higher eukaryotes three distinct polypeptides, the lamins A, B and C (Mr 60,000 - 70,000) can be discerned. The chromatin-depleted nucleolar residue has recently been isolated from amphibian oocytes and shown to consist of mainly one polypeptide with a molecular weight of 145,000 (4).

About the structural polypeptides forming the intranuclear fibro-granular network little is known yet. The limited number of studies performed indicates that a complex set of proteins may be present. Experimental evidence as to which of these polypeptides form the structural backbone is, however, lacking although a recent report of Capco et al. (5) showed that actin, in a non-microfilamental form, is a main component of the matrix. The lack of firm data for the presence of specific intranuclear proteins has increased the discussion about the reality of the internal matrix structure. However, evidence is accumulating that the internal nuclear matrix represents a predominantly proteinaceous structure in the intact cell and recent studies, in which the three-dimensional structural organization of isolated matrices is viewed using electron microscopy on whole mount preparations instead of thin sections (5, 6) corroborate this idea.

Additional indications for the existence of an intranuclear structural framework can be deduced from its functional aspects (for review see ref. 1). One of these possible functions is the attachment of replicating DNA to fixed sites inside the nucleus (for reviews see 1, 7-9). The concept of the anchorage of DNA loops to a fixed matrix seems also valid for metaphase DNA (10, 11). The association of specific receptor proteins for hormones that modulate gene expression to the matrix (12) seems to support the idea that transcriptional activity is localized at fixed sites inside the nucleus as well. Indeed, most of the nuclear heterogeneous RNA (hnRNA) is found associated with the nuclear matrix and when this RNA is analyzed one finds pre-mRNA, splicing intermediates and products of RNA processing as well as snRNAs and pre-ribosomal RNAs among its sequences (1, 13-26). This has led to the hypothesis that the nuclear matrix is intimately involved in the processing and transport of nuclear RNAs.

In this review we want to focus on gene transcription and mRNA processing as nuclear matrix-associated events. We will first discuss the isolation procedure of nuclear matrices, then the association of RNAs with this nuclear ultrastructure and finally the possible function of the nuclear matrix in RNA processing.

RESULTS AND DISCUSSION

The isolation of the nuclear matrix

Particular problems in the isolation of the nuclear matrix are its extreme sensitivity to proteolytic degradation (14, 16) and the occurrence of disulphide bridge formation between nuclear proteins during cell fractionation (27). To inhibit endogenous proteolytic activities the presence of the esterase inhibitor PMSF or PMSC (phenylmethylsulfonyl-fluoride or chloride, respectively) in all buffers seems to be adequate (14). To avoid disulphide bridge formation between nuclear proteins during cell fractionation one can add iodo-acetamide (IAA) or 2-mercaptoethanol (ME) to all buffers. Recent reports, however, indicate that N-ethylmaleimide (NEM) is more effective in blocking free SH-groups (3). However, the effects of divalent cations, the extent of protein rearrangements during preparation of the matrices, the order of the various preparation steps, the use of $(NH_4)_2SO_4$ rather than NaCl for the extraction itself, the presence of endolytic enzymes other than proteases inhibited by PMSF or PMSC and many other factors which might be important to obtain a functional structure are not studied sufficiently yet.

We have recently prepared nuclear matrices with different detergents and nucleases using PMSC and NEM in all buffers and found that the protein composition of the resulting matrices was strikingly similar (Fig. 1). The morphology of nuclear matrices prepared by these methods was essentially similar as those shown in Figure 2.

The association of hnRNA with the nuclear matrix

It has been found by several groups now that rapidly labeled nuclear RNA is tightly and quantitatively associated with the isolated nuclear matrix (13-26). Similarly, particular mRNA sequences can be found in matrix-associated RNA, for example globin RNA sequences (18, 21), ovalbumin sequences (22) and virus-specific sequences (17, 20, 23-25). Various types of experiments have been performed to exclude the possibility that the

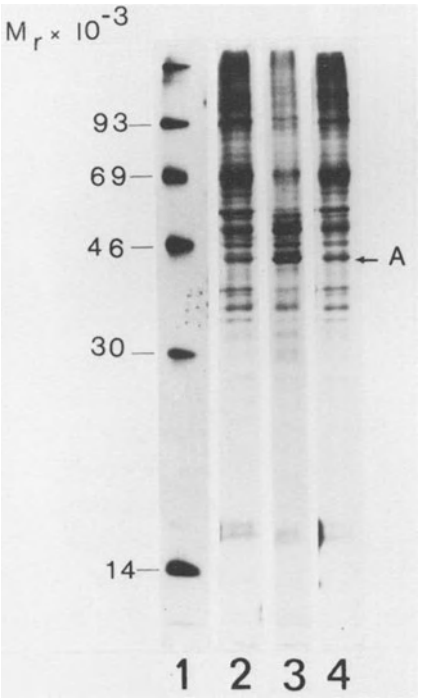

Fig. 1: Electropherogram of proteins in nuclear matrix from HeLa cells.
All buffers and solutions contained 0.5 mM PMSC; NEM (5 mM) was present in all buffers as well, except during the nuclease digestion. HeLa cells were suspended in RSB (10 mM Tris pH: 7.4, 10 mM NaCl, 1.5 mM $MgCl_2$) and homogenized by 5-10 strokes of a motor-driven Teflon pestle in a Potter-Elvehjem tissue homogenizer after addition of a DOC-Tween 40 mixture (0.5% and 1%, respectively), lane 2, or Triton X-100 (1%), lane 3, or Triton X-100 followed by an extra wash with the DOC-Tween 40 mixture (lane 4).
The nuclei ($2x10^8$ per ml) were washed with RSB and then treated with a mixture of Micrococcal nuclease (2000 U per ml) and ribonuclease A (5 mg per ml) in RSB to which 1 mM Ca^{2+} was added. Incubation for 20 min at 10°C was sufficient to degrade most of the nucleic acids. The DNA-depleted nuclei were then pelleted (5 min, 800 g) and resuspended in 0.4 M $(NH_4)_2SO_4$ in 10 mM Tris. HCl pH: 8.0 containing 1.5 mM $MgCl_2$ ($5x10^7$ nuclei per ml).
After centrifugation, the nuclear matrices were washed with RSB buffer and finally dissolved in sample buffer for gel electrophoretic analysis.
Lane 1: marker proteins
Lane 2-4: nuclear matrices prepared using various detergents (see above).
A: indicates the position of actin.

association of hnRNA with the nuclear matrix is the result of a non-specific association or caused by entrapment of the RNA or RNP complexes in the nuclear tangle remaining after the detergent and high salt treatments. Most of these control experiments have been reviewed in (28). One of the more convincing arguments is the fact that some RNA species are indeed not, or not as tightly, bound to the nuclear matrix. For example, precursors of mRNA and rRNA are more tightly bound to the nuclear matrix than the corresponding matured RNAs (22, 43).

Another indication for the specificity of the RNA-matrix interaction is the fact that the binding is very tight and resistant to various treatments unfavourable for maintaining non-specific associations (13-16). Heparin (1 mg/ml), EDTA (5-10 mM), Na-deoxycholate (5%), formamide (10 M)

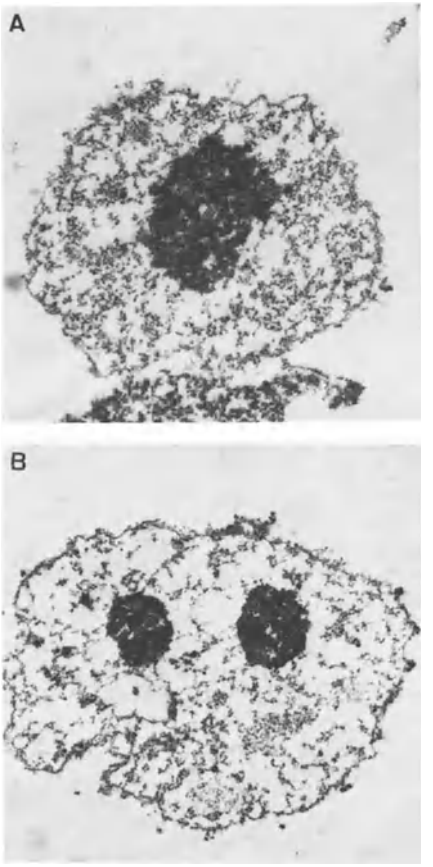

Fig. 2: Electron micrographs of HeLa nuclear matrices.
Nuclear matrices were prepared using 5 mM NEM and 0.5 mM PMSC in all solutions and buffers. In A the cells were lysed with 0.5% Triton X-100 followed by a DOC-Tween 40 (0.5% and 1%, respectively) treatment (see legend of Fig. 1). In B the cells were lysed by DOC-Tween 40 treatment. Nucleic acids were removed by Micrococcal nuclease-ribonuclease A treatment (20 min at 10°C) as described in the legend of Figure 1, followed by a wash with 0.4 M $(NH_4)_2SO_4$.

or urea (8 M) were shown to have little effect on the RNA attachment. We have tested the stability of the hnRNA attachment before the high salt treatment and in the continuous presence of disulphide bridge formation inhibitors (Table 1). After labeling HeLa cells for 15 min with ^3H precursors, nuclei were isolated in the presence of 5 mM NEM using no detergents or detergents like Triton X-100 or a Na-DOC/Tween 40 mixture (0.5% and

Table 1: EFFECT OF VARIOUS TREATMENTS ON THE RELEASE OF hnRNA FROM DNase I TREATED NUCLEI

Treatment: 5 min at 0-4°C with 10 mM Tris.HCl pH: 7.4 containing:	% of rapidly labeled nuclear RNA extracted by chemical treatment and subsequent $(NH_4)_2SO_4$ wash			left in matrices		
	A	B	C	A	B	C
10 mM NaCl, 1.5 mM $MgCl_2$	3	10	8	97	90	92
EDTA 10 mM	2	10	9	98	90	91
EGTA 2 mM	2	10	8	98	90	92
2-Mercapto-ethanol 100 mM	1	12	9	99	88	91
NaCl 2 M	2	16	15	98	84	85
KCl 2 M	N.D.	15	13	N.D.	85	87
LiCl 2 M	N.D.	16	13	N.D.	84	87
$(NH_4)_2SO_4$ 0.4M	2	15	13	98	85	87

HeLa cells (10^7 cells per ml) were incubated for 15 min with (5,6-^{3}H)-uridine (45 Ci/mmol) and (2,8,5'-^{3}H)adenosine (50 Ci/mmol), 1 µCi per ml each. All buffers and solutions in subsequent procedures contained 0.5 mM PMSC. Cells were harvested and nuclei prepared (see legend of Figure 1) by:
A: addition of a DOC-Tween 40 mixture (0.5% and 1%, respectively). In this series of experiments no NEM was added to the buffers.
B: addition of a DOC-Tween 40 mixture, but now with 5 mM NEM in all buffers and solutions.
C: addition of 0.5% Triton X-100, with NEM in all buffers and solutions.
The nuclei were incubated with DNase I (500 µg per ml, 15 min at 20°C) in RSB-PMSC, pelleted by centrifugation and resuspended in RSB, divided in an appropriate number of portions and repelleted by centrifugation (5 min, 800 g). The DNA-depleted nuclei were then resuspended in 10 mM Tris containing the designated chemicals. After 5 min at 0°C the suspensions were centrifuged at 800 g and the supernatants saved for counting. The pelleted nuclear residues were suspended in 0.4 M $(NH_4)_2SO_4$-RSB buffer and again pelleted. The supernatants were again saved for counting. The pellets (matrices) were viewed under the light microscope for the presence of internal structures (nucleoli) and for determining overall morphology. The matrices in all preparations looked normal, that is, they contained nucleoli and there were no visible signs of partial lysis or serious deformation. The matrices were then dissolved in 1% SDS (2 min, 100°C) and the soluble fraction saved for counting. Radioactivities were expressed as the percentage of combined radioactivities in the fractions. Overall yields as compared with the control ranged from 87-105%.

N.D.: Not determined.

1%, respectively). The chromatin was degraded by DNase I treatment. The nuclei were then gently suspended in various solutions (Table 1) again in the presence of NEM. After this incubation the nuclei were pelleted, resuspended in 0.4M $(NH_4)_2SO_4$ and again pelleted. The morphology of the nuclear structures (presence of nucleoli and intactness of nuclear lamina) was examined by phase contrast microscopy and the released radioactivity (rapidly labeled RNA) in each fraction was measured. The results (Table 1) show that the treatments involved did not release much hnRNA as long as the nuclear matrix structure remained intact.

Similar experiments, with minor modifications in the procedure and using 2-ME or IAA instead of NEM and using even various mixtures for the attempted hnRNA release also indicated that the hnRNA is very tightly bound to the nuclear matrix structure or, alternatively, is an essential element of the structure itself. This latter possibility has been tested experimentally.

When detergent-treated nuclei are exhaustively treated with a mixture of micrococcal nuclease and RNase A, about 5-15% of the steady-state RNA and less than 5% of the rapidly labeled RNA remain in these nuclease and thereafter high salt treated mactrices (13, Table 2). Light microscopic examination reveals an unaltered morphological appearance, that is the nuclease-treated matrices are not empty lamina structures but residual nucleoli are still present. This type of experiment suggests that most of the nuclear RNA is not essential for maintenance of the three-dimensional ultrastructure in the nuclear interior. The nature of the RNA that remains in the nuclease and high salt-treated matrices has not been analysed thoroughly but we assume that most of it has protein-protected sequences.

Core proteins involved in the RNA-matrix association

In earlier experiments we have attempted to elucidate the nature of the proteins involved in the tight binding of hnRNA with the nuclear matrix (16). The idea was that such proteins should not be released by RNase treatment of matrices and, of course, should be bound tightly to the hnRNA. Newly synthesized hnRNA is present in the nucleus in the form of a ribo-nucleoprotein (RNP) network. The protein composition of these hnRNP particles is very complex (extensively reviewed by Holoubek (29)) but most workers agree on the presence of some predominant proteins, namely the A group (32,000 - 34,000 daltons), the B group (36,000 - 37,000 daltons) and the C group (42,000 - 44,000 daltons) core proteins. When nuclear

Table 2: EFFECT OF NUCLEASE TREATMENT ON THE RNA CONTENT OF NUCLEAR MATRICES

Nuclei prepared with :	% rapidly labeled hnRNA			% steady state RNA		
	released by		left in	released by		left in
	nuclease	high salt	matrices	nuclease	high salt	matrices
Triton X-100	78	17	5	28	57	15
Triton X-100 followed by DOC-Tween 40	82	16	2	36	59	5

HeLa cells were labeled as described in the legend of Table 1 for 15 min (rapidly labeled RNA) or for 16.5 h (steady state RNA).
Nuclei were prepared as described in the legend of Figure 1 using Triton X-100 (0.5%) or Triton X-100 followed by DOC-Tween 40 (0.5% and 1%, respectively) treatment. PMSC and NEM were included in all buffers and solutions (see legend of Figure 1).
The detergent treated nuclei were incubated with Micrococcal nuclease and ribonuclease A (see legend of Figure 1) for 15 min at 10°C. After the incubation the nuclei were pelleted and resuspended in 0.4 M $(NH_4)_2SO_4$. The matrices were recovered by centrifugation and examined by phase contrast microscopy. The presence of one or more nucleoli in the matrices of all four preparations indicated that an intranuclear structure was still present after the removal of most of the RNA.
The matrices were then dissolved in 1% SDS (2 min, 100°C) and the soluble fraction saved for counting. Radioactivities in all the soluble fractions isolated were expressed as the percentage of the radioactivity present in the detergent treated nuclei.

matrices are treated with high concentrations of RNase A, the A and B core proteins are preferentially released from the matrices, but the C proteins, although present, are not (16).

To study which proteins are tightly associated with hnRNA, the UV-induced in vivo RNA-protein crosslinking was used (30). Irradiation of intact cells with ultraviolet light induces crosslinks between RNA and proteins only when the interacting molecules are close and when their reactive groups are within one bond length apart (see references in (30)). The crosslinking results showed that some hnRNA associated proteins (41,500 and 43,000 mol. wt.) are very efficiently crosslinked to hnRNA and thus are in tight contact with hnRNA in vivo. This result was also found when isolated high salt-washed matrices were irradiated. The electrophoretic mobility of these crosslinked proteins and some other indications suggested that these

proteins could be identical to the group C proteins (16, 29, 32). Although a more precise comparison of these proteins certainly is required, for example by peptide mapping, for matter of convenience we will in this review refer to these crosslinked proteins as C proteins. In fact, the tight association of the crosslinked proteins with hnRNA has also been found for the C proteins by the group of LeStourgeon who, on the basis of salt dissociation studies on hnRNP particles, described the C proteins as "interacting directly with the hnRNA" (32). The other core proteins can also be crosslinked to hnRNA although much less efficiently (17, 31, 33). These results thus strongly point to the conclusion that the A and B core proteins are associated with hnRNA but not with the matrix. In contrast, the C proteins are directly associated with both the matrix, because they are not released by the RNase treatment, and the hnRNA, because they are crosslinked so efficiently. They fulfil the requirement for proteins being involved in the association of hnRNA with the matrix.

Electron microscopic observations have shown that chromatin actively transcribed by RNA polymerase II contains fibrillar hnRNP structures resembling "beads on a string" that extend away from the DNA-protein axis. The RNP configuration for products from the same transcription unit is similar and not random with respect to the RNA sequence and clearly show that more than one RNP particle is normally found per transcript (34-37). Since the C proteins are present in isolated RNP particles it is probable that also the C proteins are bound to the RNA at multiple sites. Biochemical evidence to support this idea has been presented by Ohlsson et al. (38) and van Eekelen et al. (39) using HeLa cells infected with adenovirus. They UV irradiated intact cells to crosslink the C proteins to the RNA (16). Covalently linked hnRNA-protein complexes were treated with micrococcal nuclease and the protein-linked RNA fragments isolated. The protein moiety of these complexes was removed by proteinase K treatment and the residual pieces of RNA (about 20 nucleotides long) were 5' end labeled and then hybridized to Southern blots of adenovirus DNA restriction fragments. The results (Fig. 3) show that the crosslinked RNA annealed to all DNA fragments known to be expressed at this time after infection (39). This means that the C proteins, being the only proteins that are crosslinked under the conditions of UV irradiation used, are associated with the late transcripts at several locations from the 5' to the 3' ends of the RNA. Subsequent fine mapping of the protein-linked RNA

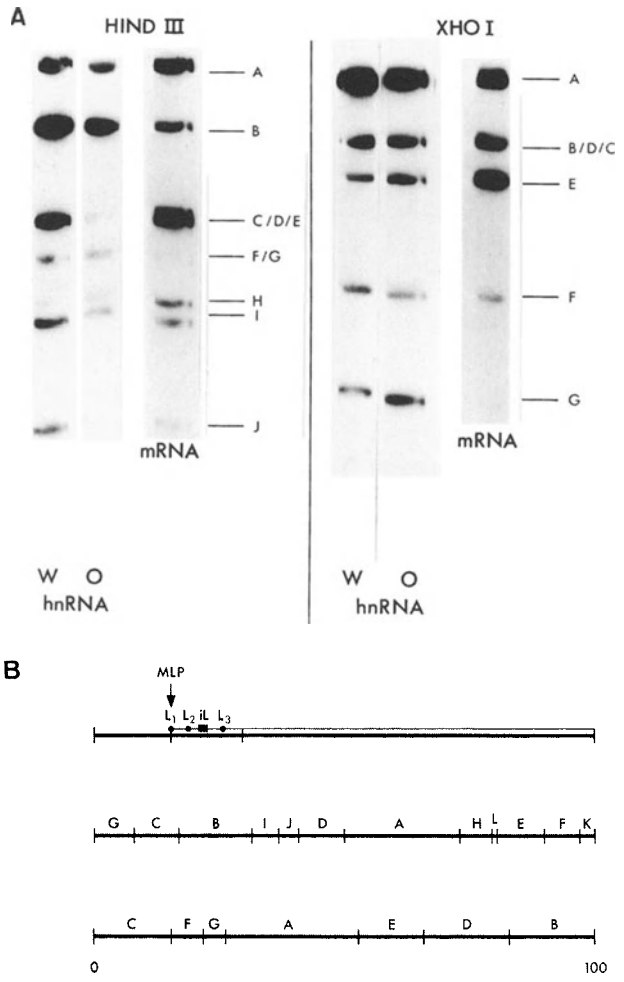

Fig. 3: Localization of C-protein-bound hnRNA fragments on the adenoviral genomic map.

A. Hybridization patterns of 5'end-labeled hnRNA probes and fragmented poly (A)-containing mRNA to Southern blots of Xho I and Hind III restricted adeno DNA. The hnRNA probes were prepared as described by van Eekelen et al. (39).
   W: RNA fragments not crosslinked to C proteins
   O: RNA fragments crosslinked in vivo to C proteins.

B. Adenovirus DNA map showing the position of the late promotor (MLP) and late transcripts (upper line) and the cleavage sites for Xho I and Hind III (lower lines).

Reprinted with permission from (39).

fragments within the Hind III B fragment of the adenoviral genome revealed regions that were more frequently crosslinked as well as regions that were less frequently crosslinked to the C proteins. The more frequently (or more efficiently) crosslinked regions are localized around the 3' end of the i-leader and around leader 3, while the less frequently crosslinked regions are found in the part of the genome encoding VA-1 and VA-2 RNA and the body of the late region 1 RNAs. Finally, intron and exon sequences were equally represented in crosslinked RNA as compared with unlinked RNA (39). The efficiency of cross-linking of the C proteins to hnRNA and the fact that they are crosslinked to

various parts of the viral transcripts indicates that probably more than one of these polypeptides are associated with each RNA molecule. These data thus support the idea of Jackson et al. (19) that RNA is bound to the matrix at multiple sites and corroborate electron microscopic data in which more than one RNP particle is usually found per transcript (34-37).

The possible meaning of such a multiple anchoring of the nascent transcript to the nuclear matrix structure is unknown yet. However, since RNA molecules "to be processed" are particularly firmly bound (21-23, 43) it may have something to do with the processing of these molecules. This will be discussed below.

The nuclear matrix and mRNA processing

After its synthesis pre-mRNA molecules are capped, methylated, polyadenylated and spliced. A function of the nuclear matrix in RNA processing seems probable since precursors of RNA are most tightly bound to this structure (22, 43). We have attempted to answer this question using adenovirus-infected cells. Adenovirus-specific hnRNA is bound to the nuclear matrix in much the same way as is host hnRNA (17). The poly (A)-containing matrix-associated RNA from infected cells was analyzed using an S1 mapping procedure described in detail elsewhere (23). When the EcoR1 B fragment of adenovirus DNA was used for the hybridization, seven major DNA:RNA hybrid bands could be distinguished (Fig. 4). Four of them, bands 4^c, 5, 6 and 7 contained processed mRNA molecules (23). Bands 1 and 2 correspond to precursors of the region L4 and L5 mRNAs and to the precursor of the DNA-binding protein (DBP)mRNA, respectively. Bands 3 and 4^n were tentatively identified as processing intermediates generated from the DBP and L4-L5 region. The implication of these results is that pre-mRNA is associated with the nuclear matrix during splicing (23).

With this finding the possibility arose that only mRNA sequences which have to be spliced are bound to the nuclear matrix. The mRNA which codes for adenovirus polypeptide IX, a structural component of the virion, is transcribed from the viral r-strand between co-ordinates 9.8 and 11.2. It has been shown that the nucleotide sequence of this mRNA is co-linear with the DNA (40) and until now this mRNA is the only known adenoviral mRNA which is unspliced. To test whether this unspliced mRNA is bound to the nuclear matrix, nuclear- and nuclear matrix-associated RNA were isolated and fractionated into polyadenylated (6% of the labeled RNA) and non-polyadenylated RNA (94% of the labeled RNA). The amount of RNA containing

nucleotide sequences specific for L3, E1B and pIX was then quantitated by filter hybridization.

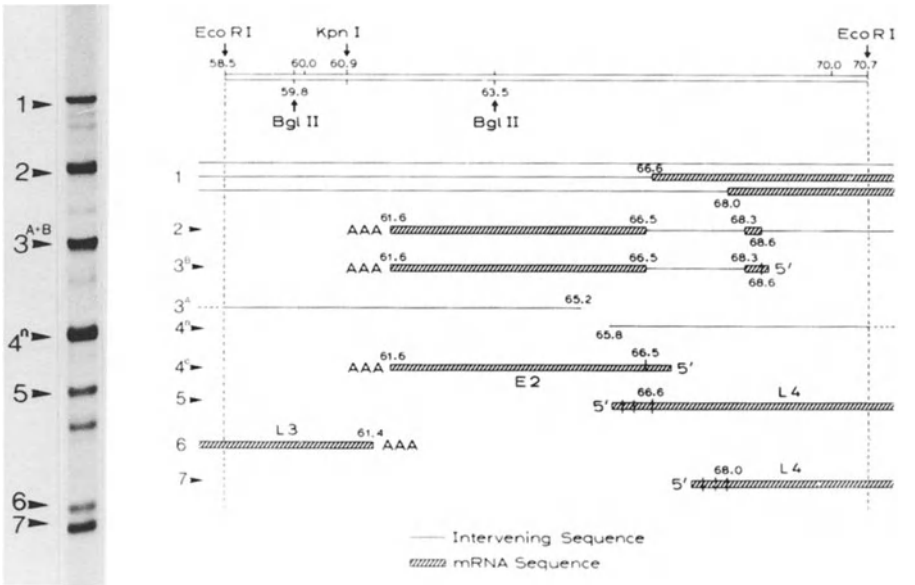

Fig. 4: Pattern of poly (A)-containing sequences present in nuclear matrices of adenovirus-infected HeLa cells.
(^{3}H)Uridine-labeled matrix RNA from virus-infected HeLa cells was prepared by phenol extraction and selected over oligo (dT)-cellulose. The poly (A)-containing RNA was hybridized to the EcoR1 B fragment of adenovirus DNA (58.5 - 70.7 map units) and analyzed via S1 mapping (23). The gel lane at the left shows an autoradiograph of the S1-resistant hybrids.
The numbers refer to the right part of the figure which summarizes the identities of these hybrid bands.
Reprinted with permission from (23).

The results (Table 3) lead to several conclusions:

a. More than 70% of the nuclear RNA containing L3, E1B and pIX-specific sequences is associated with the nuclear matrix. "To be spliced" is obviously not a requirement for association of the RNA with the matrix.

Table 3: DETERMINATION OF THE PERCENTAGES OF L3, E1B- and pIX-SPECIFIC NUCLEAR RNA SEQUENCES WHICH ARE BOUND TO THE NUCLEAR MATRIX

		Hybridization of nuclear RNA (%)	Hybridization of matrix RNA (%)	Nuclear RNA bound to the matrix (%)	Poly A(+)/ Poly A(-)
L3-specific	pA^+	7.1	6.1	86	0.8
	pA^-	9.8	7.6	78	
E1B-specific	pA^+	7.4	5.3	72	0.9
	pA^-	6.7	5.7	85	
pIX-specific	pA^+	1.0	0.7	70	0.2
	pA^-	3.8	3.2	84	

Ad2-infected HeLa cells 18 h after infection were labeled with ^{3}H-nucleosides for 30 min. After labeling nuclear and nuclear matrix RNA were prepared and fractionated into poly A(+) and poly A(-) RNA by oligo (dT)-cellulose chromatography. In each fraction of RNA the amounts of L3-, E1B- and pIX-specific sequences were determined by filter hybridization as described (24). The percentages of hybridization were corrected for non-specific adsorption by incubation of RNA with a blank filter under the same conditions. The third column shows the percentages of specific RNA sequences which are bound to the nuclear matrix. These data were calculated from the amounts of certain specific RNA sequences in nuclear and nuclear matrix RNA as shown in the first and second column. The fourth column shows the ratio between the amounts of certain specific RNA sequences in poly A(+) and poly A(-) nuclear matrix RNA. Each percentage is the mean value of the results obtained in 5 separate hybridizations. The deviation of the values in the first and second column is within 1/10 of the indicated values.
Reprinted with permission from (24).

b. Poly (A)-containing as well as non-polyadenylated hnRNA is bound to the nuclear matrix, indicating that also polyadenylation is a matrix-associated process. It is also unlikely that the poly (A)-tail is an essential element in the binding of RNA to the matrix as has been suggested by Herman et al. (13).

c. In the case of the unspliced pIX mRNA the ratio poly A(+)/poly A(-) RNA is very low. This suggests that very soon after polyadenylation, that is

after completion of processing of this particular RNA, the RNA is released from the matrix structure. In the case of L3 and E1B RNAs the ratio poly A(+)/poly A(-) RNA is much higher indicating that after polyadenylation of these RNAs some additional processing steps still have to be carried out.

Possible function of the nuclear matrix in the splicing of pre-mRNA

Using the S1 mapping procedure we have studied the cleavage reactions that occur during splicing of the tripartite leader sequences in RNAs that are derived from the major late transcription unit of adenovirus. When pulse labeled nuclear RNA was analyzed it was found that the primary transcript was cleaved at several locations. Certain cleavages are rapidly introduced in the primary transcript while other cleavage sites are used only after a significant lag time. In the latter case the cleavages occur in a processing intermediate rather than in the primary transcript. The rapid cleavages seem to occur at preferred sites in the nascent transcripts (see Fig. 5), that is, these first acts of splicing can occur before polyadenylation or termination of transcription (41).

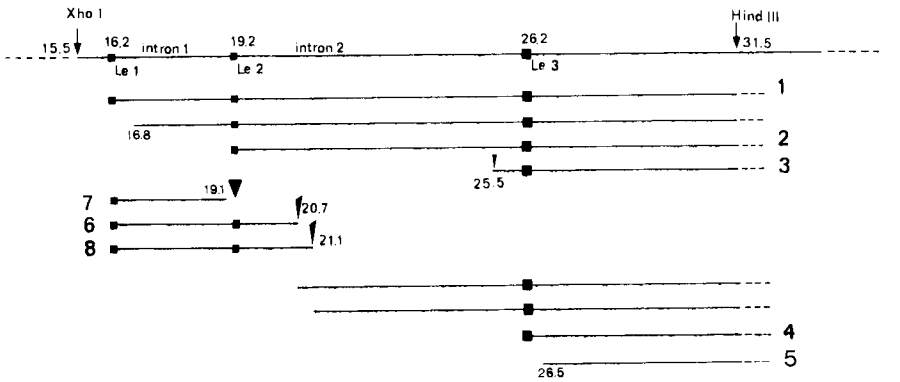

Fig. 5: Cleavage sites in newly synthesized adenovirus-specific transcripts. The figure gives a summary of the S1 mapping data that were obtained via S1 analysis of rapidly labeled nuclear RNA using the XhoI F and HindIII B fragments as DNA probes. For details see (41).
Le = leader.
Reprinted with permission from (41).

Examples of such preferred cleavage sites are co-ordinates 16.8 and 19.1 in intron 1 and 20.7 , 21.1 and 25.5 in intron 2. Cleavage of the primary transcripts at these sites occurs very soon after the passing of the RNA polymerase (within 3 min after the start of labeling). Although intron 2 in the nascent transcript can be cleaved very rapidly, Keohavong et al. (42) have shown that ligation of leader 2 to leader 3 occurs only after a time lag of 12 min. In agreement with these results we have found that cleavage at the acceptor site of intron 2 (at 26.2 map units) indeed occurs only at about 15 min after the start of transcription (41). Similar results were found when the cleavage reactions in the intron 1 region were analysed. Intron 1 is rapidly cleaved at 16.8 or at 19.1 map units. The latter preferred cleavage site is located at about 30-35 nucleotides from the acceptor site (19.2 map units) of intron 1. The last cleavage in intron 1 occurs at the acceptor site and is followed by rapid ligation of leader 1 to leader 2 (E.C.M. Mariman, unpublished observations).

The fact that cleavage and ligation are possibly independent processes implies that the unligated pieces of pre-mRNA have to be kept in place to ensure correct ligation. We know that these unligated RNA fragments are bound to the nuclear matrix and that they can be crosslinked in vivo to the C-proteins (E.C.M. Mariman, unpublished observation). So it seems a likely possibility that the nuclear matrix is involved in the spatial orientation of the 5' and 3' cleavage products to ensure correct ligation.

Concluding remarks

Based on the information discussed in the preceding paragraph we would like to draw the following conclusions:

1. RNA precursors, in particular rRNA and mRNA precursors are bound to the nuclear matrix structure soon after initiation of their synthesis. Several RNA processing steps (e.g. polyadenylation and splicing) occur when the RNA is bound to the nuclear matrix.
2. This RNA-matrix association is probably mediated by proteins. The involvement of two well-known hnRNA-associated proteins, generally referred to as C-proteins, has been suggested. Electron microscopic and biochemical evidence suggest that an RNA molecule may be bound to the matrix at multiple sites without measurable preference for intron sequences or exon sequences.
3. Recent experiments indicate that the nascent transcript is cleaved at several sites without concomittant ligation. The unligated pieces of

RNA are bound to the nuclear matrix and it is therefore possible that the nuclear matrix provides the structural framework that keeps the pieces of RNA in the correct spatial orientation for ligation.

ACKNOWLEDGEMENTS

We are very grateful to Drs. A.O. Pogo and R. Procyck for helpful discussions concerning the isolation of nuclear matrices. We thank Peter T. Sillekens and the Laboratory of Submicroscopic Morphology, University of Nijmegen (head: Prof. A. Stadhouders) for preparing the samples for electron microscopy.

This work was supported in part by the Netherlands Foundation for Chemical Research (S.O.N.) and the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

REFERENCES

1. Berezney, R. The nuclear matrix. Plenum Publ. Corp., New York, 1984.
2. Shelton, K., Egle, P. and Crochan, D. In: The nuclear envelope and the nuclear matrix (Ed. G.G. Maul) Alan R. Liss Inc., New York, 1982, pp. 157-168.
3. Kaufmann, S.H., Gibson, W. and Shaper, J.H. J. Biol. Chem. 258: 2710-2719, 1983.
4. Franke, W.W., Kleinschmidt, J.A., Spring, H., Krohne, G., Grund, C., Trendelenburg, M.F., Stoehr, M. and Scheer, U. J. Cell Biol. 90: 289-299, 1981.
5. Capco, D.G., Wan, K.M. and Penman, S. Cell 29: 847-858, 1982.
6. Capco, D.G. and Penman, S. J. Cell Biol. 96: 896-906, 1983.
7. Vogelstein, B., Nelkin, B., Pardoll, D., and Hunt, B.F. In: The nuclear envelope and the nuclear matrix (Ed. G.G. Maul) Alan R. Liss Inc., New York, 1982, pp. 169-181.
8. Berezney, R., Basler, J., Buchholtz, L., Smith, H. and Siegel, A. In: The nuclear envelope and the nuclear matrix. (Ed. G.G. Maul), Alan R. Liss Inc., New York, 1982, pp. 183-197.
9. Wanka, F., Pieck, A., Bekers, A. and Mullenders, L. In: The nuclear envelope and the nuclear matrix (Ed. G.G. Maul), Alan R. Liss Inc., New York, 1982, pp. 199-211.
10. Lebkowski, J.S. and Laemmli, U. J. Mol. Biol. 156: 300-324, 1982.
11. Lewis, C.D. and Laemmli, U.K. Cell 29: 171-181, 1982.
12. Barrack, E. In: The nuclear envelope and the nuclear matrix (Ed. G.G. Maul) Alan R. Liss Inc., New York, 1982, pp. 247-258.
13. Herman, R., Weymouth, L. and Penman, S. J. Cell Biol. 78: 663-674, 1978.
14. Miller, T.E., Huang, C.-Y. and Pogo, A.O. J. Cell Biol. 76: 675-691, 1978.
15. Long, B.H., Huang, C.Y. and Pogo, A.O. Cell 18: 1079-1090, 1979.
16. van Eekelen, C.A.G. and van Venrooij, W.J. J. Cell Biol. 88: 554-563, 1981.
17. van Eekelen, C.A.G., Mariman, E.C.M., Reinders, R.J. and van Venrooij, W.J.

Eur. J. Biochem. 119: 461-467, 1981.
18. Maundrell, K., Maxwell, E.S., Puvion, E. and Scherrer, K. Exp. Cell Res. 136: 435-445, 1981.
19. Jackson, D.A., McCready, S.J. and Cook, P.R. Nature 292: 552-555, 1981.
20. Jackson, D.A., Caton, A.J., McCready, S.J. and Cook, P.R. Nature 296: 366-368, 1982.
21. Ross, D.A., Yen, R-W. and Chae, C-B. Biochemistry 21: 764-771, 1982.
22. Ciejek, E.M., Norstrom, J.L., Tsai, M-J. and O'Malley, B.W. Biochemistry 21: 4945-4953, 1982.
23. Mariman, E.C.M., van Eekelen, C.A.G., Reinders, R.J., Berns, A.J.M. and van Venrooij, W.J. J. Mol. Biol. 154: 103-119, 1982.
24. Mariman, E., Hagebols, A.M. and van Venrooij, W.J. Nucl. Acids Res. 10: 6131-6145, 1982.
25. Ben-Ze'ev, A., Abulafia, R. and Aloni, Y. The EMBO J. 1: 1225-1231, 1982.
26. van Venrooij, W., van Eekelen, C., Mariman, E. and Reinders, R. In: The nuclear envelope and the nuclear matrix (Ed. G.G. Maul) Alan R. Liss Inc., New York, 1982, pp. 235-245.
27. Kaufmann, S.H., Coffey, D.S. and Shaper, J.H. Exp. Cell Res. 132, 105- 123, 1981.
28. van Venrooij, W. and van Eekelen, C. In: The nuclear matrix (Ed. R. Berezney) Plenum Publ. Corp., New York, 1984, in the press.
29. Holoubek, V. In: Chromosomal nonhistone proteins - Biochemistry and Biology (Ed. V. Hnilica), C.R.C. Press, Vol., 1983, in the press.
30. Wagenmakers, A.J.M., Reinders, R.J. and van Venrooij, W.J. Eur. J. Biochem. 112: 323-330, 1980.
31. Setyono, B. and Greenberg, J.R. Cell 24: 775-783, 1981.
32. Beyer, A.L., Christensen, M.E., Walker, B.W. and LeStourgeon, W.M. Cell 11: 127-138, 1977.
33. Mayrand, S. and Pederson, T. Proc. Natl. Acad. Sci. USA 78: 2208-2212, 1981.
34. Beyer, A.L., Miller, O.L. jr. and McKnight, S.L. Cell 20: 75-84, 1980.
35. Beyer, A.L., Bouton, A.H. and Miller, O.L. jr. Cell 26: 155-165, 1981.
36. Beyer, A.L., Bouton, A.H., Hodge, L.D. and Miller, O.L. jr. J. Mol. Biol. 147: 269-295, 1981.
37. Pederson, T. and Davis, N.G. J. Cell Biol. 87: 47-54, 1980.
38. Ohlsson, R.I., van Eekelen, C. and Philipson, L. Nucl. Acids Res. 10: 3053-3068, 1982.
39. van Eekelen, C.A.G., Ohlsson, R.,Philipson, L., Mariman, E., van Beek, R. and van Venrooij, W.J. Nucl. Acids Res. 22: 7115-7131, 1982.
40. Aleström, P., Akusjärvi, G., Perricaudet, M., Mathews, M.B., Klessig, D.F. and Pettersson, U. Cell 19: 671-681, 1980.
41. Mariman, E.C.M., van Beek, R.J. and van Venrooij, W.J. J. Mol. Biol. 163: 239-256, 1983.
42. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stévenin, J. Nucl. Acids Res. 10: 1215-1229, 1982.
43. Herlan, G., Eckert, W.A., Kaffenberger, W. and Wunderlich, F. Biochemistry 18: 1782-1788, 1979.

mRNA of RNA viruses

9

REOVIRUS mRNA

W. K. JOKLIK

SUMMARY

There are three mammalian reovirus serotypes, each with a set of ten genes that have evolved independently. Reovirus messenger RNAs form an attractive and readily manipulated system for examining the nature of RNA-protein and RNA-RNA interactions, including the nature of the mechanisms that cause mRNAs to become incorporated into progeny virus particles, and the factors that regulate the efficiency with which they are translated.

INTRODUCTION

Each infectious reovirus particle contains 10 discrete and distinct molecules of double-stranded (ds) RNA. Upon infection, these molecules are transcribed into plus-stranded single-stranded (ss) RNA by an RNA polymerase, the transcriptase, that is present within reovirus particles and was actually the second virus-associated transcriptase (after that in vaccinia virus) to be discovered. The transcripts fulfill two functions: They can be translated into the reovirus proteins, and they can be transcribed into minus strands with which they remain associated, thereby forming progeny dsRNA molecules. Thus they act as the reovirus messenger RNAs, and they also represent the physical as well as the genetic link between parental and progeny reovirus genomes.

Definitions

Table 1 lists the major reovirus-specified proteins. Each genome segment is transcribed into only one transcript, a transcript that is identical to its plus strand (1), and each of these transcripts is translated into only one protein (some of which are cleaved more or less extensively to provide smaller cleavage products). The dsRNA segments thus correspond to genes, and they will be referred to as such in the discussion that follows. It should be

Y. Becker (ed.), VIRAL MESSENGER RNA.

Table 1. Reovirus Proteins

Species	Molecular weight	Percent in virus particle	Approximate number of molecules per virus particle	Location
Structural proteins				
λ1	155,000	15	105	Core
λ2	140,000	11	90	Core
λ3	135,000	⩽2	⩽12	Core
μ1	80,000	2	20	Core
μ1C	72,000	35	550	Outer shell
μ2	70,000	⩽2	⩽12	Core
σ1	40,000	1	24	Outer shell
σ2	38,000	7	200	Core
σ3	35,000	28	900	Outer shell
Nonstructural proteins				
μNS	78,000			
μNSC	73,000			
σNS	42,000			

noted, however, that recent sequencing studies have revealed the existence of relatively short open reading frames in three genes in addition to the long open reading frames that encode the known reovirus specified proteins (see below). The proteins encoded by these shorter open reading frames have not yet been identified.

In several cases the relative electrophoretic migration rates of genes and cognate proteins do not coincide. For the Dearing strain of reovirus serotype 3, the prototype reovirus strain that has been investigated most extensively, the situation is as shown in Fig. 1 (2). Whether the lack of correlation has as its basis actual size differences (which would be possible because reovirus genes appear to possess variously sized untranslated regions at their 3'-termini), or whether its basis is trivial and buffer-dependent, is not yet known.

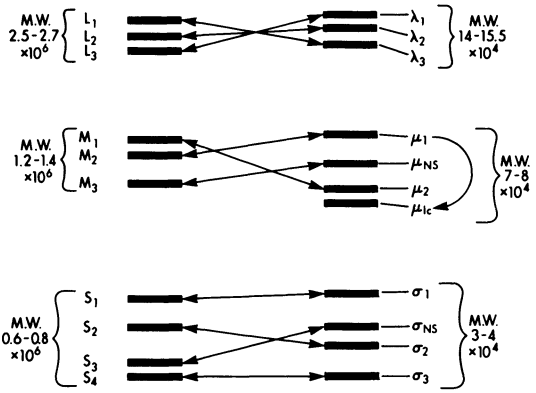

FIGURE 1. Summary of RNA coding assignments for reovirus serotype 3, strain Dearing. The positions of the RNA and protein species are those assumed in the Loening's buffer E (0.036 M Tris, 0.03 M phosphate, pH 7.8) and Tris·glycine systems, respectively (2). Reproduced with permission.

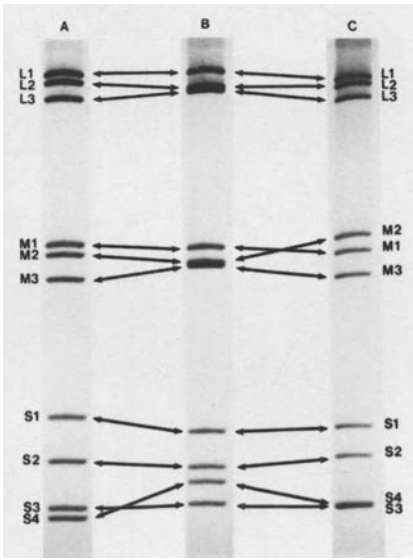

FIGURE 2. Reovirus double-stranded RNA. Lanes A, B, and C show, respectively, the genes of serotypes 3, 2, and 1 electrophoresed in 7.5% polyacrylamide-Tris·phosphate gels and stained with EtBr (4). The arrows identify cognate genes. Reproduced with permission.

It has already been pointed out that most definitive work on reovirus has so far been carried out on the Dearing strain of reovirus serotype 3. Recently strains of serotypes 1 and 2 are also being used in attempts to use the comparative virology approach for elucidating the functions of the various reovirus genes and of the proteins that they encode. The migration profiles of the genes of the three reovirus serotypes in a Tris glycine buffer system (3,4) are shown in Fig. 2.

The Reovirus Multiplication Cycle

The principal features of the reovirus multiplication cycle are summarized in Fig. 3. Briefly, reovirus adsorbs to cells via receptors specific for the outer capsid protein σ1 (5); remarkably, all three reovirus serotypes adsorb to the same receptor, although the σ1 proteins exhibit no more than 10% amino acid sequence homology (see below). The virus enters cells via phagocytic vesicles which soon fuse with lysosomes within which the outer capsid is partially disrupted/digested (6): proteins σ3 and σ1 are removed sequentially, and protein μ1C loses its N-terminal 8,000 dalton region, but remains associated with the virus particles, which are now subviral particles (SVPs) (7,8). Disruption and partial removal of the outer capsid shell activates transcription of all ten genes into complete plus strands (which does not occur in intact virus particles). The plus strands, which are capped at their 5'-ends but not polyadenylated at their 3'-ends (9), serve first as messenger RNAs that are translated into all ten primary

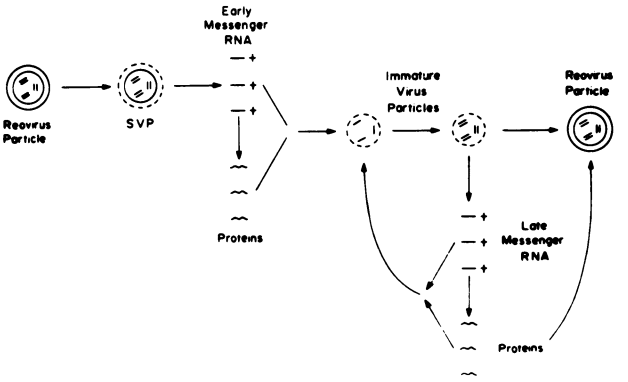

FIGURE 3. Schematic representation of the reovirus multiplication cycle.

reovirus proteins. Then, after a certain interval of time, presumably when sufficient concentrations of transcripts and of several viral proteins have been achieved, morphogenesis commences. The first stage of this process is the formation of complexes that consist of one each of all ten plus strand species and certain viral proteins within which the plus strands are transcribed once and once only into minus strands with which they remain associated, thereby forming the ten progeny reovirus genes (10,11,12). The protein composition of the particles then changes progressively which results in the formation of particles that appear to be similar to cores. These intermediate stage progeny particles transcribe the genes that they contain into plus-stranded transcripts just like the parental SVPs (13,14). Since there are many more of these immature progeny virus particles in infected cells than parental SVPs, most reovirus mRNA that is formed during the multiplication cycle is made by these incomplete progeny virus particles rather than by the parental SVPs. In fact, studies with mutants that are temperature-sensitive with respect to ability to form the earliest form of progeny virus particles have shown that about 95% of the reovirus mRNA that is made during the multiplication cycle is transcribed by immature progeny virus particles (15). Morphogenesis is then completed with the addition of the outer capsid shell, a process the final stage of which is the addition of protein σ3. Addition of this protein blocks the transcription of reovirus genes into complete mRNA molecules (7,16).

The reovirus multiplication cycle can thus be divided into two periods, one during which only parental genomes are transcribed and another during which the great excess of progeny genomes are transcribed. This second period provides the actual basis for reovirus multiplication.

More detailed information concerning molecular, genetic and biologic aspects of reovirus and its multiplication cycle may be found in reference (17).

TRANSCRIPTION IN VITRO

Discovery

The need for a reovirus-associated RNA-dependend RNA polymerase was foreshadowed by the demonstration that transcription of the reovirus genome proceeds in infected cells even if protein synthesis is inhibited (18). The actual discovery of such an enzyme then followed soon after the demonstration of the first virus particle-associated RNA polymerase, that of vaccinia virus

(19). Transcription *in vitro* by reovirus particles was demonstrated simultaneously by two groups, one in the laboratory of Graham who activated mRNA transcription (which is not catalyzed by intact virus particles) by heating virus particles at 70° for 20 s (20), and the other in the laboratory of Shatkin, who activated RNA transcription by digesting the outer capsid shell of reovirus particles with chymotrypsin, thereby generating cores (21). In practice, the latter method, yielding as it does a defined product that can be readily isolated in density gradients (cores sediment at a density of 1.43g/ml in CsCl density gradients, whereas intact virus particles sediment at 1.36g/ml) and that possesses a unique and distinctive protein complement (lacking completely proteins σ1, σ3 and μ1C) (22), has been employed exclusively for generating *in vitro* particles capable of transcribing the reovirus genome into complete mRNA molecules.

Activation

The reovirus transcriptase is not inactive in virus particles; but it does not transcribe full length mRNA molecules. Instead, transcription is initiated with the same efficiency as when full length transcripts are formed, but aborts when transcripts are on the average only 3 to 4 residues long (23). Thus "activation" of the transcription of mRNA does not involve altering the frequency or efficiency of transcription initiation, but rather the removal or abolition of constraints that prevent the relative movement of transcriptase and template (see below). How this is achieved is not known. "Activation" is a complex process that involves the successive removal by proteolytic enzymes of proteins σ1 and σ3 and either the complete removal of protein μ1C (*in vitro*) or its partial removal (*in vivo*, see above and below)(24). Borsa and his collaborators have shown that the ionic composition of the medium is of importance, Cs^+ and NH_4^+ being "facilitating" ions and Na^+ being a "nonfacilitating" ion (25). Cooperative effects also operate, since the concentration of virus particles influences the nature and extent of "activation" that can be achieved (24, 26). The products of proteolytic activation *in vitro* are cores that lack the entire outer capsid shell and that are capable of transcribing plus-stranded transcripts, the reovirus mRNAs, in fully capped form for many hours, each gene being transcribed repeatedly (27).

Kinetics and relative frequency of transcription

The relative number of each transcript species formed is inversely

proportional to its size; that is, about four times as many of each s size class transcript species are formed as of each l size class transcript species, and so on (28). This implies that each gene is transcribed at the same rate, irrespective of its size. Measurement of the initial transcriptional rates agrees with this concept: transcription proceeds rapidly, the first s, m and l size class mRNA molecules being discernible after 2, 4 and 8 min, respectively, which indicates a chain growth rate of about 600 residues/min. However, the amount of transcripts formed in 60 min is far less than what would be expected on the basis of this rate of synthesis. This result poses a dilemma; for given the evidence that the rate of transcription is independent of gene size, it implies that the intervals between the generation of intact transcripts are proportional to gene size. The reason why this should be so is not known.

Capping

In 1974 Miura et al. (29) discovered that the 5'-termini of the plus strands of reovirus genes, just like those of cytoplasmic polyhedrosis virus (30), were blocked with a novel alkali-resistant group. Shortly thereafter Furuichi et al. (31) showed that this structure was $m^7G(5')ppp(5')G^mpCp-$ and that it is also present at the 5'-termini of all reovirus mRNA species (32). Cap possession, which increases significantly the efficiency with which mRNAs are translated (33, 34), is not an exclusive feature of reovirus mRNAs, but is a property of almost all mRNAs (35-38). Its role in increasing translation efficiency will be discussed below.

Capping is effected by four successive reactions:

$$\text{pppGpC-} \xrightarrow{\text{Nucleotide phosphohydrolase}} \text{ppGpC-} + P_i$$

$$\text{pppG} + \text{ppGpC-} \xrightleftharpoons{\text{guanylyltrasferase}} \text{GpppGpC-} + PP_i$$

$$\text{GpppGpC-} + \text{AdoMet} \xrightarrow{\text{methyltransferase 1}} m^7\text{GpppGpC-} + \text{AdoHcy}$$

$$m^7\text{GpppGpC-} + \text{AdoMet} \xrightarrow{\text{methyltransferase 2}} m^7\text{GpppG}^m\text{pC-} + \text{AdoHcy}$$

(AdoMet, S-adenosylmethionine; AdoHcy, S-adenosylhomocysteine)

The structure $m^7G(5')ppp(5')G^mpCp$-is known as cap 1. During the later part of the reovirus multiplication cycle some reovirus mRNA molecules acquire cap 2 structures, namely $m^7G(5')ppp(5')G^mpC^mp$-. The second 2'-O-methylation appears to be catalyzed by a cellular methyl transferase (39).

By using inhibitors of selected steps of cap synthesis such as GMP-P(NH)P, the nonhydrolyzable analog of GTP which renders the β, -bond resistant to cleavage by the viral nucleotide phosphohydrolase (35), and AdoHcy, an analog of AdoMet (40), it can be demonstrated that capping, methylation and transcription are not obligatorily coupled.

Reovirus mRNA molecules are not polyadenylated (9). Since several of them are translated very efficiently, possession of poly(A) sequences at the 3'-terminus is not essential for mRNA function.

The fact that transcription frequency is not the same for all genes but is inversely proportional to gene size indicates that the ten reovirus genes are not linked within virus particles. This was an issue of some interest in the reovirus field, even though the evidence for it is minimal (electron micrographs showing linkage of genes at very low frequency), and the evidence against it is strong [20 gene termini within each virus particle (41), inability of reovirus genes to hybridize with each other, absence of terminal proteins, and the like]. The only manner in which linked genes could give rise to the experimentally observed transcription frequencies would be if the genes were arranged in increasing size order, transcription starting with the smallest, and the polymerase dissociated from the template as a function of distance from the initiating terminus. Such an arrangement is highly unlikely. On the contrary, the unequal transcription frequencies support the other evidence that reovirus genes are unlinked inside reovirus particles.

Abortive Transcription

Until recently, it was thought that the reovirus transcriptase is active only in cores and in the SVPs formed in cells following infection (see above), but not in virus particles. Yamakawa *et al.* (23) have recently shown that this is not so. When intact reovirus particles are incubated with Mn^{2+}(or Mg^{2+}) and the four nucleoside triphosphates, they synthesize G-terminated oligonucleotides. About two-thirds of these oliogonucleotides are (p)ppGpC, one fifth are (p)ppGpCpU, and most of the remainder are (p)ppGpCpUpA. As will be shown below, GCUA is the tetranucleotide sequence that is shared by all 10

reovirus genes. The oligonucleotides are therefore incomplete transcripts; transcription can obviously be initiated in virus particles, but aborts before more than 3 or 4 residues are polymerized. Intact virus particles synthesize no complete transcripts. The most likely explanation of these findings is that in virus particles the transcriptase is active, but that movement of enzyme relative to template is not possible. Abortive transcription is not limited to reovirus particles. In cores and SVPs also, most transcription events are abortive and result in the formation of oligonucleotides (42, 43). But in these particles removal or disruption of the outer capsid shell has apparently removed, at least partially, whatever block prevents the movement of enzyme relative to template, and some transcription events (probably 10-20%) proceed to yield intact transcripts. It seems that if transcription proceeds beyond residue 4, a complete transcript is likely to result.

The nature of the enzymes involved in transcription and capping

Little is known concerning the nature of the transcriptase or of the four capping-function enzymes, since all enzyme activities are lost when cores or reovirus particles are disrupted; nor do monoclonal antibodies directed against any of the proteins located on the outer surface of reovirus particles or cores inhibit mRNA transcription (Gaillard, Antczak and Joklik, unpublished results). It is therefore thought that the transcriptase and capping enzyme catalaytic sites are fixed on the interior surface of the core shell. Indeed, when cores transcribe dsRNA in the presence of labeled pyridoxal phosphate, which is hypothesized to react with the reactive site of the transcriptase, both protein λ1 and protein λ2 become labeled (44); and when cores transcribe mRNA in the presence of [^{32}P] GTP, protein λ2 becomes labeled, which suggests that the guanylyl transferase catalytic site is located on it (45). Further, all four capping enzymes are inhibited by pyridoxal phosphate, although this compound does not bind to any of their active sites (46). This suggests that reovirus mRNA transcription and capping are accomplished by a topographically related group of enzyme molecules. It should also be pointed out no functions have yet been assigned to the three minor proteins λ3, μ1 and μ2, which are also associated with cores, and these proteins, either free in the interior of the virus particle or attached to the inner core shell, may themselves possess transcriptase or capping activity or both. In fact, Drayna and Fields (47) have found that the pH optimum of the transcriptase is specified by gene L1, the gene that encodes protein λ3.

TRANSCRIPTION IN VIVO

Under optimal conditions of infection, the pattern and kinetics of transcription of reovirus mRNA in infected cells is very similar to that described above for the situation in vitro, that is, about 4 times as many s size class transcripts are formed as l size class transcripts, and so on (48). The amounts of mRNA species within size classes that are formed are not strictly equimolar; for example, in the l size class more l3 transcripts than l1 and l2 transcripts are usually formed, and in the m and s size classes, m3, s3 and s4 transcripts are usually formed in larger amounts than the others. However, the transcription ratios within size classes do not differ by more than 2-fold, and it is not entirely clear that this difference is not due to differences in the rates of recovery, rather than to differences in the rates of transcription.

All available evidence indicates that the relative rates of transcription are the same during the late period of the multiplication cycle (from about 4hr onwards at 37°), when transcripts are formed within immature progeny virus particles, as during the early period when they are formed within parental SVPs (48).

Graham and his collaborators reported that during the early stages of the infection cycle genes l1, m3, s3 and s4 are transcribed predominantly, and that in cells in which protein synthesis is inhibited, the same genes again are transcribed to a much greater extent than the remainder (49). Further, Shatkin and LaFiandra (50) found that SVPs isolated from cycloheximide-treated cells transcribed all 10 genes in vitro, but only genes L1, M3, S3 and S4 if re-introduced into cycloheximide-treated cells. It has therefore been suggested that some cellular protein(s) regulates reovirus transcription; but the mechanism that might be involved is not known. Further, under conditions optimal for RNA extraction from infected cells, the standard transcription pattern described above is observed even during the earliest stages of the reovirus multiplication cycle. The phenomenon described by Graham may thus have as its basis differences in the efficiences with which various mRNA species can be recovered or measured, rather than differences in transcription frequencies.

TRANSLATION IN VIVO

Reovirus does not switch off host protein synthesis rapidly. In cells infected with a multiplicity of about 10 pfu per cell at 37°, the rate of host

protein synthesis at 10hr after infection - which is quite late in the multiplication cycle, when about one third of the total virus yield has already been formed - is reduced by about 50 percent, and at that time about as much reovirus-coded protein is formed as host cell protein (48). However, if such cells are infected in the presence of 0.5μg/ml actinomycin D, more than 75 percent of the total protein synthesized at that time is virus-encoded, and even at about 6hr after infection roughly equal amounts of host and reovirus protein is synthesized. It is difficult therefore to detect and measure reovirus protein synthesis before 4hr after infection, the time when the first immature reovirus particles are formed under these conditions. It is, however, possible to measure reovirus protein synthesis well before this time by using a radioimmunoprecipitation technique in which extracts of infected cells, pre-absorbed with pre-immune serum, are mixed with reovirus antiserum pre-absorbed with extracts of uninfected cells, and the resulting antigen-antibody complexes are collected on cells of Staphyloccus aureus, from which they are then solubilized and subjected to SDS-polyacrylamide gel electrophoresis, followed by quantitative autoradiography. Under those conditions one obtains extremely "clean" protein profiles (51). Even greater resolution and precision can be obtained using not antisera, but monoclonal antibodies directed against specific reovirus proteins, such as are shown in Fig. 4 (52). Under these conditions the following reovirus-encoded proteins

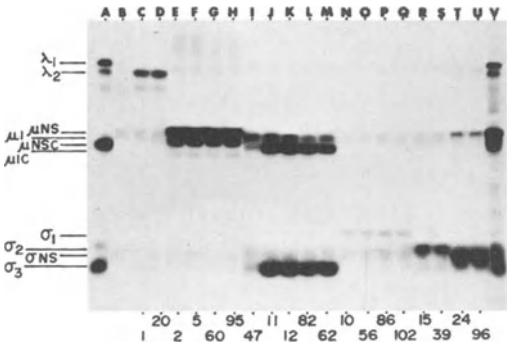

FIGURE 4. Autoradiogram of an SDS-urea-7.5% polyacrylamide gel in which labeled reovirus proteins that were precipitated by various hybridoma-secreted IgGs had been electrophoresed. The number of each clone is indicated, as is each reovirus-coded protein. Lane A, [^{35}S]methionine-labeled reovirus capsid proteins; lane B, cell lysate reacted with the culture supernatant of 45.6.TG1.7 cells (the parental myeloma cell line); lane V, [^{35}S] methionine-labeled infected cell lysate. The direction of electrophoresis was from top to bottom (52). Reproduced with permission.

have been demonstrated in infected cells:

Protein λ1 and λ1C. No monoclonal antibody against λ1 is available, but highly specific antiserum against λ1 can be made by immunizing rabbits with material excised from a SDS-polyacrylamide gel. Protein λ1C is slightly smaller than protein λ2. About 15 percent of λ1 is cleaved to λ1C (Antczak and Joklik, unpublished results).

Protein λ2 is formed in amounts that are roughly equal to those of protein λ1. Here also some protein λ2 (about 10 percent) is cleaved to protein λ2C (51). Whereas it is not known whether protein λ1C is a component of virus particles - since its size is very similar to that of protein λ2 which is present in virus particles to much greater extent - it is quite clear that protein λ2C is a component of virus particles: It is usually visible in SDS-polyacrylamide gels of highly purified reovirus particle preparations.

No monoclonal antibody is available for protein λ3; therefore it has not yet been possible to detect its synthesis in infected cells.

Protein μ1 is synthesized in large amounts from messenger RNA species m2. Protein μ1 exists in infected cells in two forms, μ1 and μ1C, a cleavage product that lacks the 8,000 dalton N-terminal region (48). About 90% of the free forms of these proteins is in the form of μ1 (52). Protein μ1C exists either in the form of a complex with protein σ3, the other major outer capsid shell component, or in the form of immature virus particles. Overall, more than 95% of intracellular protein μ1 is cleaved to protein μ1C. Since most of the free form of proteins μ1/μ1C is uncleaved μ1, and since most of the protein complexed with σ3 is protein μ1C, it is likely that the conversion of μ1 to μ1C occurs very soon after protein μ1 complexes with protein σ3 (52).

Messenger RNA species m3 is translated into large amounts of protein μNS, one of the two nonstructural proteins. About one-half of this protein exists in uncleaved form, the other half in the form of protein μNSC, which is 3,000-5,000 daltons smaller (52). The function of neither protein is known. There are indications that protein μNS may function in morphogenesis (Antczak and Joklik, unpublished results).

Messenger RNA species m1 is translated into protein μ2, a minor core component. This RNA species is translated poorly; its translation has only

been detected *in vitro*, not *in vivo* (both because only very little protein μ2 is formed, and because its size is almost identical with that of the major outer capsid shell component μ1C).

As for the four *s* size class mRNA species, they are translated into capsid proteins σ1, σ2 and σ3, and into the nonstructural protein σNS (2). No cleavage products of any of these proteins have been detected. Although all *s* size class mRNAs are formed in roughly equimolar amounts, large amounts of σ3 and σNS are formed, somewhat smaller amounts of σ2, and only very small amounts of σ1 (48,52). The ratio of the amounts of σ3 and σ1 that are formed appears to be about 50.

Zweerink and Joklik (48) measured the relative rates of transcription and translation of individual mRNA species throughout the multiplication cycle and found that they did not vary significantly. Table 2 presents estimates of the frequencies with which the individual reovirus mRNA species are translated (rate of translation/rate of transcription). They fall into three classes. Three (species *l*1, *m*1 and *s*1 are translated very infrequently; six (*l*2 and *l*3, *m*2 and *m*3, and *s*3 and *s*4) are translated frequently; and species *s*2 is translated with intermediate frequency.

Table 2. Relative rates of transcription and translation, and translation frequencies, of the ten species of reovirus serotype 3 mRNA.

mRNA Species	Relative Rate of Transcription	Relative Rate of Translation	Rate of Translation / Rate of Transcription
*l*1	0.05	< 0.01	< 0.2
*l*2	0.05	0.15	3
*l*3	0.05	0.1	2
*m*1	0.15	< 0.01	< 0.67
*m*2	0.3	1.0	3.3
*m*3	0.5	2.0	4.0
*s*1	0.5	0.025	0.05
*s*2	0.5	0.2	0.4
*s*3	1.0	1.0	1.0
*s*4	1.0	1.5	1.5

THE SIGNIFICANCE OF THE m^7G CAP IN REOVIRUS mRNA TRANSLATION

Cap possession increases the efficiency of mRNA translation (35). It appears that at least two mechanisms are involved. First, cap possession increases the efficiency with which mRNAs form initiation complexes with 40 S ribosomal subunits (33,34); second, several cap binding proteins (CBPs) exist (53,54), the best characterized of which is one with a subunit molecular weight of 24,000 (55), the function of which seems to be unwinding the 5'-terminal regions of capped mRNAs, thereby promoting the binding of ribosomal subunits (56). This is indicated by the fact that treatment of reovirus mRNA with formaldehyde (57) or replacement of G in mRNA by H (58,59), both of which cause secondary structure destabilization, lessen the requirement for both cap and CBP (56).

The structural requirements for cap function have been studied. Structures such as $m^7G(5')p$, $m^7G(5')pp$ and $m^7G(5')ppp$ inhibit translation of capped, but not of uncapped mRNAs (60), as do the corresponding ethyl and benzyl derivatives (61). The nature of the 7-substituent is therefore not important.

Cap possession increases the efficiency of translation, but is not absolutely essential for translation. Indeed, some mRNAs, like those of picornaviruses and certain plant viruses, are not capped at all. Furthermore, normally capped mRNAs can be translated in either uncapped form, or if they contain unmethylated caps. A good example of this is provided by reovirus mRNAs synthesized by cores in the absence of AdoMet which are translated in cell-free protein-synthesizing systems (62) - even in the presence of AdoHcy (to inhibit methylases that may be present in such systems) - with about 15 percent of the efficiency with which fully capped mRNAs are translated (63). It has also been found that the stringency of the cap requirement varies with the source of the cell-free extract, translation in reticulocyte (64) and ascites cell systems (65) being less dependent on cap possession than that in wheat germ extracts. The reason for this may be that CPB only reduces rather than than abolishes secondary structure near the 5' terminus of mRNAs, and that the amounts and characteristics of CBPs may differ in different systems. It should be noted that CBPs have no effect on the nature of the proteins that are synthesized, which indicates that they play no role in identifying correct initiation codons (57).

Finally, it has also been found that cap possession increases mRNA stability (66). Uncapped and capped reovirus mRNA molecules were injected into Xenopus oocytes or incubated in cell-free protein synthesizing systems derived from wheat germ or mouse L cells, and the half lives of the former were found to be one-half to one-third those of the latter.

Millward and his collaborators have recently suggested that late reovirus mRNA, that is, mRNA transcribed by immature virus particles, may not be capped. They showed that extracts of uninfected cells translated capped reovirus mRNA with high efficiency, and uncapped mRNA at a much lower efficiency; for extracts of L cells infected with reovirus the reverse was true (67); and that monoclonal antibodies directed against rabbit reticulocyte CBPs, which inhibit capped but not uncapped mRNA translation, failed to inhibit translation of reovirus mRNA in extracts of infected cells (70). They also found that the capping system in immature reovirus particles was masked (68), and that the 5'-termini of mRNA molecules synthesized by such particles were uncapped (69).

These results are consistent with the idea that reovirus modifies the cap dependence of the host translational apparatus and in that way takes over the host translational apparatus. Attractive though this notion is, it remains to be confirmed. Thach and his collaborators found no transition from cap-dependent to cap-independent translation as a result of reovirus infection (71); they found that extracts of cells infected with reovirus did not display an increased capacity to translate decapped globin mRNA, and that translation of both cellular and viral mRNAs in such extracts was sensitive to $m^7G(5')ppp$. This controversy is yet to be resolved.

TRANSLATION IN VITRO

The availability of large amounts of pure reovirus mRNA species has facilitated studies of the mechanisms involved in the initiation of translation, studies that have yielded results that are most likely also applicable to translation initiation of eukaryotic mRNAs in general.

In vitro systems for translating reovirus mRNAs were developed by McDowell et al. (62) who showed that reovirus mRNA could be translated in Krebs II mouse ascites cell, mouse L cell, Chinese hamster ovary cell, HeLa cell and rabbit reticulocyte extracts, and by Both et al. (33,34) who used a wheat germ extract to demonstrate the translation of the 10 species of reovirus mRNA into 10 protein species (72). The fact that translation

frequency is an intrinsic property of mRNAs received support from the demonstration by Levin and Samuel (73) that the relative translation frequencies of s size class reovirus mRNAs in vitro are similar to those in vivo. Accordingly in vitro protein synthesizing systems began to be used to discover the structural features of reovirus mRNAs that control frequency of translation.

Several years ago Kozak and Shatkin adapted the ribosome-protection assay devised by Steitz (74) to study coliphage MS2 RNA translation to the reovirus system in order to determine where ribosomes and ribosomal subunits bind to mRNA molecules when protein synthesis is initated. They incubated reovirus mRNA species labeled with [^{32}P] with wheat germ ribosomes or 40S ribosomal subunits in the presence of sparsomycin (in order to inhibit elongation), digested the resulting complexes with ribonuclease and sequenced the mRNA portions that were protected (75-80). Nine such protected sequences are shown in Fig. 5. In all cases a sequence is protected that includes the 5'-terminus including the cap, as well as the initiation codon and a sequence of 15-20 residues downstream from the initiation codon. The regions protected by 80S ribosomes are shorter, about 30 residues long, and are centered around the

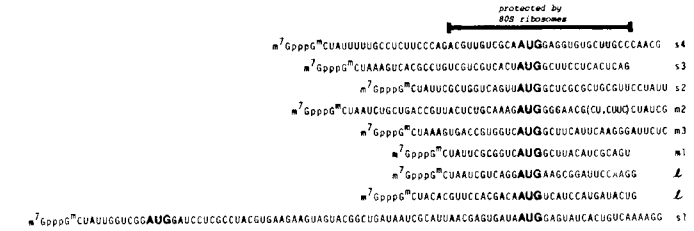

FIGURE 5. The ribosome-binding sites in 9 mRNAs of reovirus serotype 3. The presumptive AUG initiator codon in each message is shown in boldface (81). For mRNA species s1 the second AUG codon is aligned with the 5'-proximal AUG codons of the other species; but it is, in fact, the 5'-proximal AUG codon that is the initiating codon for the translation of protein σ1 (Cashdollar, Chmelo and Joklik, unpublished results). The sequences presented here agree almost completely with those of the 5'-terminal regions of the corresponding mRNA species presented by Antczak et al. (86). The upper 1 mRNA sequence shown here is that of species 13, the lower that of species 11 (86).

initiation codons. The initiation codon is in each case that closest to the 5'-terminus, except in the case of s1 mRNA in which the ribosome-binding assay detects a second functional initiation codon, in a different reading frame, starting at residue 71 (80). Although the latter appears to be a strong initiation signal as judged by the strength of ribosome binding, the former is in fact used to initiate translation of the primary s1 gene product, protein σ1, since it is succeeded by an open reading frame of 345 codons, whereas the latter is succeeded by an open reading frame of only 120 codons, the protein corresponding to which is not known (Cashdollar, Chmelo and Joklik, unpublished results; see below).

Kozak and Shatkin postulate that 40S ribosomal subunits bind to the 5'-end of mRNA, either to the cap, to the CBP or to the region that is unwound by the CBP (see above), and then proceeds downstream until they encounter the first AUG triplet, where they pause, combine with 60S subunits, and then initiate translation (82). The pausing occurs only if the secondary structure around the initiation codon is intact, since pausing does not occur if fragments of mRNA are used; if some of the G residues in mRNA are replaced by H residues which causes a weakening of the secondary structure (83); or in the presence of formaldehyde, bisulfite or other reagents that destroy or weaken secondary structure (84). Ability to recognize AUG initiation codons is also abolished in the presence of inhibitors of protein synthesis such as edeine (82).

All evidence points to proximity to the 5'-terminus being the overridingly important characteristic of AUG codons that are used for initiation. This thesis was recently confirmed by Kozak (85) in an elegant experiment in which she constructed plasmids containing multiple reiterated copies of sequences containing the natural preproinsulin initiation codon, and showed that translation is initiated in each case at the 5'-proximal initiation codon. The only exceptions to this rule appeared to be reovirus s1 mRNA which contains two functional ribosome binding sites (80) - but the first is used to code for protein σ1; and mRNA species 12 in which the first AUG starts at position 6 but is followed 9 codons later by a termination codon (86) (see also above). It should of course be pointed out that it is impossible to know whether a particular AUG is an initiator codon unless one knows either the N-terminal sequence of the corresponding protein or the complete mRNA sequence so as to be able to ascertain the availability of open reading frames. Genes S2 (87), S1 (Cashdollar, Chmelo and Joklik, unpublished

results) and S3 (88) of reovirus serotype 3 strain Dearing, as well as the S1 genes of reovirus serotypes 1 and 2 (Cashdollar, Chmelo and Joklik, unpublished results) have now been sequenced, and in each case it is the 5'-proximal AUG that is followed by a long open reading frame of exactly the size required to encode the cognate protein.

Kozak examined many other eukaryotic mRNAs for features close to the initiation codon that correlated with translation frequency and found that strong initiation codons, that is, initiation codons from which translation was initiated frequently, almost always had a purine in position -3 (the A of the AUG being position +1) and a G in position +4 (89). This rule holds well for the mRNAs of reovirus serotype 3. The l2, l3, m2, m3, s3 and s4 mRNA species are translated frequently, and all have a purine both in position -3 and in position +4, as does the moderately frequently translated s2 mRNA. By contrast, the very infrequently translated l1 mRNA has U in position +4, and the very infrequently translated s1 mRNA has C in position -3; however the very infrequently translated m1 mRNA does possess purines in both positions -3 and +4 (86). The rule also holds for the l3, m3 and s2 mRNA species of serotypes 1 and 2, all of which possess strong initiation codons, and for the s1 mRNA species of serotypes 1 and 2, both of which possess weak initiation codons (that is, a pyrimidine in either or both positions -3 and +4).

Another approach to the problem of what controls translation frequency was adopted by Thach and his collaborators, who studied the relative translation frequencies of reovirus and host mRNAs in reovirus infected cells, and their ability to compete with each other for limiting factors (90-93). They interpreted their results as indicating that all mRNAs, both viral and cellular, must compete for a limiting message-discriminatory component before binding to 40S ribosomal subunits, and that translation frequency is determined by their relative affinity for this component. In general, host cell messenger RNAs have a higher affinity for this factor than reovirus mRNAs, but during the later stages of the multiplication cycle so much reovirus mRNA is synthesized that it is translated predominantly simply because it is present in larger amounts. They developed a quantitative *in vitro* assay system for measuring the ability of mRNA to compete for this factor and, in partial agreement with the results presented in Table 2, found that mRNA species m2 and m3 were translated very efficiently. Strangely enough, however, they did not find that any of the s size class mRNA species are translated efficiently, though they obviously are *in vivo* as well as *in*

vitro. The nature of the ribosome-independent discriminatory factor for which mRNAs are postulated to compete and which selects against mRNAs for which its affinity is lowest - though it permits all mRNAs to be translated with maximum efficiency if it is present in excess - remains to be discovered.

THE SEQUENCES OF THE mRNAs OF THE THREE REOVIRUS SEROTYPES

Characteristics of the three reovirus serotypes

The numerous strains of reovirus that circulate in nature in mammalian hosts can be divided into three groups or types on the basis of their immunological properties. Type-specificity is the property of one of the ten reovirus proteins, namely protein σ1, which is encoded by gene S1 (94, 52). In other words, three quite distinct S1 genes exist in nature that have evolved from a common ancestor and that encode three σ1 protein whose antigenic determinants are by now quite different, although they retain other common functions such as fitting into the reovirus outer capsid shell, ability to hemagglutinate (95), ability to act as the cell attachment protein (5), and ability to react with cells of the immune mechanism (96).

The other nine genes of the three reovirus serotypes are related much more closely than the S1 genes. As judged by the ability of their genomes to hybridize, reovirus serotypes 1 and 3 are related to the extent of about 70 percent to each other, and about 10 percent to serotype 2 (4). In order to estimate the relatedness of the individual genes of the three reovirus serotypes, Gaillard and Joklik (4) constructed hybrid genes whose plus and minus strands derived from cognate genes of different serotypes and measured their resistance to ribonuclease digestion under standard conditions. The results of this analysis are presented in Table 3. The gene that has diverged most markedly during evolution is the S1 gene; the genes that have diverged least are the three L genes. In all cases the serotype 2 and serotype 1 or 3 genes exhibit no more than 20 percent, and often less than 10 percent homology. This indicates that the gene sets of reovirus serotypes 1, 2 and 3 evolved independently of each other.

Sequences at the Termini of the Ten Species of mRNA of Reovirus Serotype 3

The sequences at the 5'-termini of the ten species of reovirus serotype 3 mRNA are shown in Fig. 6 (86). All mRNAs share a common 5'-terminal tetranucleotide GCUA-, and all possess an initiation codon within 13-33 residues of the 5'-terminus. Several mRNAs possess additional initiation

Table 3. Percent homology among the genes of reovirus serotypes 1, 2 and 3 (see ref. 4).

Gene	+(1)[a] -(2)	+(1) -(3)	+(3) -(1)	+(3) -(2)	+(2) -(1)	+(2) -(3)
L1	15	86	84	18	21	11
L2	23	85	94	13	12	7
L3	12	88	79	13	15	14
M1	9	66	69	5	8	3
M2	5	31	42	9	11	8
M3	10	63	67	5	6	8
S1	4	12	6	1	1	4
S2	14	56	44	9	6	13
S3	12	69	81	14	8	12
S4	11	57	51	11	11	15

[a] Percent ribonuclease-resistant material in the hybrid gene consisting of a serotype 1 plus strand and a serotype 2 minus strand.

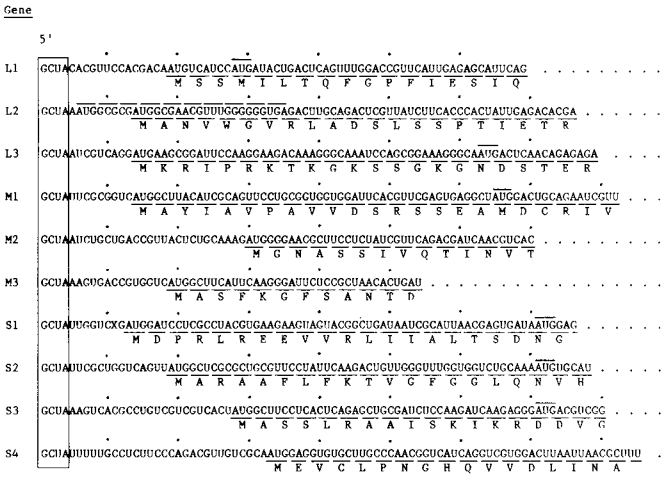

FIGURE 6. The sequences of the 5'-terminal regions of the ten mRNAs of reovirus serotype 3 (strain Dearing) (86). Reproduced with permission.

codons, either in phase or out of phase, further downstream. It is impossible to known which of these initiation codons are used in the absence of further sequencing information, but since the recognition signal for ribosome attachment appears to be the 5'-terminal capped region, it is very likely that it is indeed the first initiation codon that is used (see above). mRNA species s1, s2 and s3 have been sequenced (see below), and in them it is indeed the 5'-proximal initiation codon that is used, because it is the only one that is followed by an appropriate length open reading frame.

The amino acid sequences of proteins starting at the first initiation codon (or the second, in gene L2) are shown beneath each sequence (86). For most genes, the number of hydrophobic and charged amino acid residues is about equal, and in most cases the net charge of the first 15-20 amino acid residues is either 0 or 1. The only exceptions are the proteins encoded by genes M2 and S2, which encode proteins μ1/μ1C and σ2, respectively, in which the net charge is +3 and +4, respectively, and the protein encoded by gene L3, protein λ1, a major core constituent. The nature of the first 22 amino acids of this protein is striking; there is only one hydrophobic amino acid residue and no fewer than nine basic amino acid residues and two acidic ones, which gives the N-terminal 22 amino acid residues a net charge of + 7. The amino terminus of protein λ1 is thus very hydrophilic and basic.

At the 3'-termini the only sequence shared by all ten genes is the 3'-terminal pentanucleotide UCAUC-3'. Multiple termination codons are present in all ten sequences (86).

The availability of sequence data concerning all ten reovirus mRNAs presents the opportunity for asking questions such as

(a) Can any evolutionary relationships be discerned among the ten genes (that is, do some genes appear to be derived from others?)

(b) Do the genes possess features that may function in gene assortment during morphogenesis?

(c) Do they possess features that may regulate translation frequency?

(d) Can binding recognition sites for RNA polymerases and capsid proteins be discerned?

The sequences were therefore examined for the following by means of computer programs: (1) Homologies (that is, the presence of identical or very similar same-sense sequences); (2) Symmetries (that is, the presence of identical or very similar opposite-sense sequences); (3) Dyad symmetries (that is, the presence of complementary sequences); (4) Dinucleotide frequences; and

(5) Complementary sequences between the 5'-terminal sequences of the mRNAs and the 3'-terminal consensus sequence of eukaryotic 18S ribosomal RNA. To assess whether any features found were significant, identical searches were carried out on 20 random sequences constructed using a random number table (86).

Possible Evolutionary Relationships. The 5'-and 3'-terminal regions of reovirus genes possess 22 sequences 20-30 residues long that are very similar to other sequences, either in the same sense or in the opposite sense. No such regions were present in the 20 random sequences. These sequences are widely distributed among reovirus mRNAs. Only s1 mRNA does not possess at least one; mRNA l1 possesses no fewer than 6, and 11, 6 and 5 of these sequences are present in l, m, and s size class mRNAs, respectively. A remarkable feature of these sequences is that no fewer than 7 exhibit either homology or symmetry not only to one but also to two or even three other sequences; that four sequences exhibit homology with one sequence and symmetry with another; and that two sequences exhibit homology with two sequences and symmetry with a third. This results in an extraordinarily complex network of closely related sequences the origin of which is not readily apparent, but which may represent vestiges of gene-duplication events or of aberrant transcription of both plus and minus strands (86).

Features That May Function in mRNA Assortment. The 20 mRNA sequences were examined for signals that might explain how sets of 10 reovirus plus strands are assembled during the initial phase of reovirus morphogenesis. Numerous regions exhibiting complementarity and potentially capable of forming reasonably strong associations by means of hydrogen-bonding can be found; in fact, some interactions have free energies as low as -25 Kcal and even lower that are highly significant energetically and thermodynamically stable. The problem is that random sequences exhibit an equal frequency of such interactions. Thus, the terminal regions of reovirus mRNAs are undoubtedly capable of interacting by base-pairing, but such interactions may not possess the specificity necessary for such interactions to form the basis of a highly selective association process. In fact, no mechanism could be detected capable of directing the unique incorporation of one copy of each of the 10 plus strands into an immature virus complex. In particular, the sequences were examined for ability to interact in unique arrangements such as one plus strand species acting as a "collector" strand for the other nine, or for

head-to-head, tail-to-tail, or head-to-tail and tail-to-head associations that would result in the formation of circles upon insertion of the tenth element. No evidence supporting such models could be detected (86). One must conclude either that recognition signals of this nature are not present in the terminal regions, or that the associations occur via mechanisms that do not depend on base-pairing, or that the assembly mechanism involves not only nucleic aicd-nucleic acid but also nucleic acid-protein interactions.

Features That May Control Frequency of Translation. Attempts were made to discern signals that might function to control the frequency of translation of the ten reovirus mRNAs. Several factors could be eliminated readily (86). For example, almost none of the reovirus plus strands exhibit significant ability to form hairpin loops in the 5'-terminal regions; and although most of them can form rather stable, and even very stable, interactions between sequences in their 5'-terminal and 3'-terminal regions, thereby forming rings, they do so neither more, nor significantly less frequently than random sequences, and the relative stability of such interactions does not correlate, positively or negatively, with relative translation frequency. Secondary structure considerations based on hairpin loop and ring formation therefore do not seem to enter into controlling frequency of translation. Nor does ability to base-pair with the consensus 3'-terminal region of mammalian ribosomal 18 S RNA appear to be a factor; such interactions are rather weak, do not occur more frequently with the 5'-terminal regions of reovirus mRNAs than with their 3'-terminal regions, and do not occur more or less frequently than with random sequences. Nor could frequency of translation be correlated with distance of the first initiation codon from the 5'-terminus or with the presence of additional initiation codons. Two factors, however, do correlate with relative frequency of translation. First, for 9 of the 10 reovirus mRNAs translation frequency correlates with possession of purines in positions -3 and +4 with respect to the initiation codon (89) (see above); second, reovirus mRNA species that are translated frequently possess AG-rich regions upstream and surrounding the first initiation codon, whereas mRNA species that are translated infrequently do not possess such AG-rich regions (86).

Protein-Binding Recognition Sites. Apart from the 5'- and 3'- terminal tetra- and pentanucleotides, the ends of the ten mRNA species possess no sequences sufficiently similar for them to be interpreted as representing common

protein-binding sites. If such sites exist, they must share features other than similarity of base sequence (86).

Sequences at the Termini of the l3, m3, sl and s2 mRNAs of reovirus serotypes 1, 2 and 3.

Li et al. (97) and Gaillard et al. (98) sequenced the terminal regions of the l3, m3, sl and s2 mRNAs of reovirus serotypes 1, 2, and 3. This selection of mRNAs provides information concerning representatives of all three size classes, comparison between two mRNAs of the s size class, and information concerning mRNAs that encode both structural and nonstructural proteins. The 5'-terminal regions of these mRNAs are presented in Fig. 7.

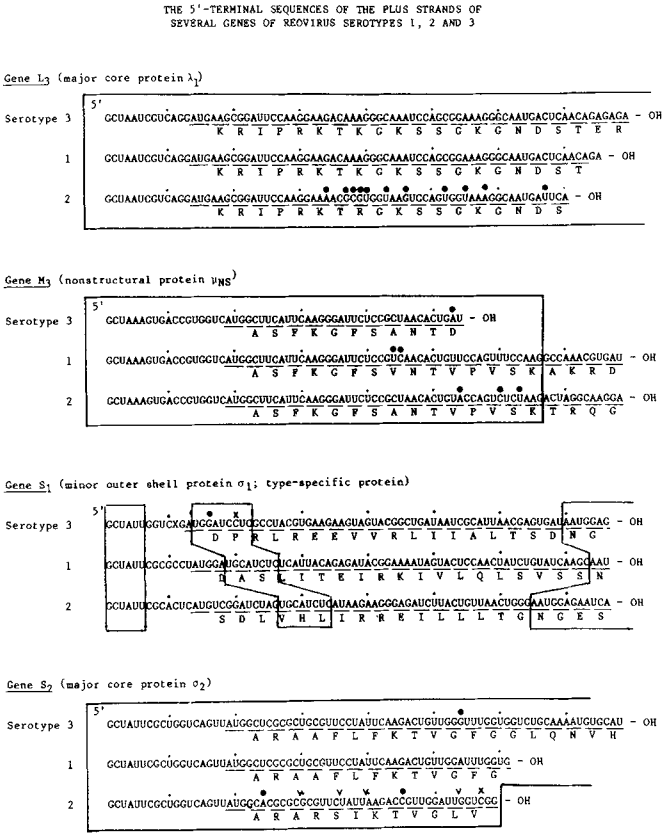

Figure 7. The sequences of the 5'-terminal regions of the l3, m3, sl and s2 species of reovirus serotype 3, 1 and 2 mRNA (97,98).

The 5'- and 3'-terminal regions of the l3, m3 and s2 mRNAs of the three serotypes are extraordinarily similar. Where there are substitutions in the 5'-terminal regions, they are mostly in third codon positions, so that the amino acid sequences remain the same. There is no sequence divergence in the first 18 codons of the three l3 mRNAs, but the sequences of the serotype 1 and 2 m3 mRNAs do diverge after the 15th codon. In the case of the s2 mRNAs, the serotype 1 and 3 mRNAs are almost identical, but in the serotype 2 mRNA there are 7 changes in the first 59 residues, including deletions that change the reading frame, so that the amino acid sequence of the serotype 2 σ2 protein differs from that of the serotype 1 and 3 σ2 proteins in five of the first 12 amino acids. As for the three s1 mRNAs, they are almost totally dissimilar, both in the regions upstream from the initiation codon and in their coding regions, so that the amino acid sequences of the three σ1 proteins are quite different (note thtat the shared 5'-proximal nonapeptides are read in different reading frames).

At their 3' ends, the four sets of cognate mRNAs display similar patterns: the sets of l3, m3 and s2 genes are very similar, the three s1 mRNAs much less so (97,98).

The most unexpected feature of this analysis is the extent of similarity between the serotype 2 mRNAs on the one hand and the serotype 1 and 3 mRNAs on the other. Since the relatedness of serotype 2 genes to serotype 1 and 3 genes is no more than about 15 percent (50; see also above), sequence divergence among these genes must be concentrated in their internal regions. This is curious because the reason for terminal homology is clearly not a requirement for polymerase and encapsidation recognition signals that possess the same sequence; note that the three s1 mRNAs show only very limited homology in their terminal regions (95).

No doubt it will be possible to answer these and other questions better when the complete sequences of all ten reovirus mRNAs are known.

CLONING THE REOVIRUS GENES

Cashdollar, Chmelo and Joklik (unpublished results) have cloned all 10 genes of all three reovirus serotypes into pBR322 by tailing both strands of each gene with poly(A), transcribing them with reverse transcriptase, tailing the cDNA strands with oligo dC, isolating molecules of the correct size by electrophoresis in alkaline agarose gels, reannealing the plus and minus strands, and cloning the tailed ds cDNA molecules by standard procedures (85).

All cloned genes were shown to be complete gene copies by sequencing their termini and showing that these termini were identical with those of the reovirus genes themselves as determined by Antczak et al.(84).
At least five reovirus genes have been sequenced. Cashdollar et al. (85) sequenced the S2 gene of reovirus serotype 3 which is 1329 nucleotides long and possesses a single long open reading frame that extends for 331 codons, exactly the length required to encode a protein the same size as the known S2 gene product, the major core component σ2 (M_r, 38,000). Interestingly, the S2 gene possesses a second 85 codon long open reading frame in a different phase that starts a short distance beyond the end of the long open reading frame. However, there is no evidence that this reading frame is operative, as no 10,000-dalton protein such as it would encode, has yet been detected in cells infected with reovirus.

Richardson and Furuichi (88) sequenced the S3 gene of reovirus serotype 3. It is 1198 nucleotides long and possesses a long open reading frame that extends for 366 codons. The untranslated region at its 3'-terminus is only about 70 residues long.

Cashdollar, Chmelo and Joklik (unpublished results) have sequenced the S1 genes of all three reovirus serotypes. The serotype 3 gene is 1416 nucleotides long and possesses a long open reading frame that starts at the 5'-proximal AUG codon and extends for 345 codons, which is the size required for encoding a protein with the molecular weight of about 40,000, the molecular weight of protein σ1. Interestingly, this gene possesses another for encoding a protein with the molecular weight of about 40,000, the molecular weight of protein σ1. Interestingly, this gene possesses another open reading frame that commences at the second AUG codon, at residue 71, and that is 120 codons long. It could encode, in a different reading frame from that for σ1, a basic protein with a molecular weight of about 13,000. Such a protein has not yet been found in infected cells. Interestingly, this reading frame is also open in the other two only very distantly related S1 genes (Cashdollar, Chmelo and Joklik, unpublished results).

The availability of intact cloned reovirus genes should permit work in several areas that have so far not been accessible to experimentation. For example, it will be possible to clone reovirus genes into expression vectors which will provide large amounts of individual species of native reovirus proteins which can then be examined for enzymic activity and other functions, as well as make possible in vitro studies of reovirus morphogenesis and of the

protein-protein and protein-nucleic acid interactions that this involves. Further, it will be possible to alter the cloned genes by introducing point mutations, deletions, and substitutions, and in that way provide mRNAs encoding altered proteins, which should be invaluable in defining reovirus protein functions. The ability to alter mRNAs should also permit identification of their structural features that control translation frequency.

ACKNOWLEDGMENTS

I would like to thank numerous colleagues for their collaboration and stimulating discussions, particularly Bill Cashdollar, Richard Chmelo, Dick Gaillard and Jim Antczak.

REFERENCES

1. Li, J. K.-K., Scheible, P. P., Keene, J. D. and Joklik, W. K. Virology 105:282-286, 1980.
2. McCrae, M. A. and Joklik, W. K. Virology 89:578-593, 1978.
3. Sharpe, A. H., Ramig, R. F., Mustoe, T. A. and Fields, B. N. Virology 84:63-74, 1978.
4. Gaillard, R. K. and Joklik, W. K. Virology 123:152-164, 1982.
5. Lee, P.W.K., Hayes, E. C. and Joklik, W. K. Virology 108:156-163, 1981.
6. Silverstein, S. C. and Dales, S. J. Cell Biol. 36:197-230, 1968.
7. Chang, C. and Zweerink, H. J. Virology 46:544-555, 1971.
8. Silverstein, S. C., Astell, C., Levin, D. H., Schonberg, M. and Acs, G. Virology 47:797-806, 1972.
9. Stoltzfus, C. M., Shatkin, A. J. and Banerjee, A. K. J. Biol. Chem. 248:7993-7997, 1973.
10. Acs, G., Klett, H., Schonberg, M., Christman, J., Levin, D. H. and Silverstein, S. C. J. Virol. 8:684-689, 1971.
11. Sakuma, S. and Watanabe, Y. J. Virol. 10:628-638, 1972.
12. Zweerink, H. J., Ito, Y. and Matsuhisa, T. Virology 50:358-359, 1972.
13. Morgan, E. M. and Zweerink, H. J. Virology 68:455-466, 1975.
14. Sakuma, S. and Watanabe, Y. J. Virol. 8:190-196, 1971.
15. Ito, Y. and Joklik, W. K. Virology 50:189-201, 1972.
16. Astell, C., Silverstein, S. C., Levin, D. H. and Acs, G. Virology 48:648-654, 1972.
17. Joklik, W. K. (Ed.):The Reoviridae. Plenum Press, New York and London, 1983.
18. Watanabe, Y., Millward, S. and Graham, A. F. J. Mol. Biol. 36:107-116, 1968.
19. Kates, J. A. and McAuslan, B. R. Proc. Natl. Acad. Sci. USA 58:134-138, 1967.
20. Borsa, J. and Graham, A. F. Biochem. Biophys. Res. Commun. 33:895-900, 1968.
21. Shatkin, A. J. and Sipe, J. D. Proc. Natl. Acad. Sci. USA 61:1462-1466, 1968.
22. Smith, R. E., Zweerink, H. J. and Joklik, W. K. Virology 39:791-810, 1969.
23. Yamakawa, M., Furuichi, Y., Nakashima, K., LaFiandra, A. J. and Shatkin, A. J. J. Biol. Chem. 256:6507-6512, 1981.

24. Joklik, W. K. Virology 49:700-715, 1972.
25. Borsa, J., Sargent, M. D., Copps, T. P., Long, D. G. and Chapman, J. D. J. Virol. 11:1017-1024, 1973.
26. Borsa, J., Sargent, M. D., Lievaart, P. A. and Copps, T. P. Virology 111:191-200, 1981.
27. Levin, D. H., Mendelsohn, N., Schonberg, M., Klett, H., Silverstein, S., Kapuler, A. M. and Acs, G. Proc. Natl. Acad. Sci. USA 66:890-895, 1970.
28. Skehel, J. J. and Joklik, W. K. Virology 39:822-831, 1969.
29. Miura, K.-I., Watanabe, K., Sugiura, M. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 71:3979-3984, 1974.
30. Miura, K.-I., Watanabe, K. and Sugiura, M. J. Mol. Biol. 86:31-40, 1974.
31. Furuichi, Y., Muthukrishnan, S. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 72:742-746, 1975.
32. Furuichi, Y., Morgan, M., Muthukrishnan, S. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 72:362-366, 1975.
33. Both, G. W., Banerjee, A. K. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 72:1189-1193, 1975.
34. Both, G. W., Furuichi, Y., Muthukrishnan, S. and Shatkin, A. J. Cell 6:185-193, 1975.
35. Shatkin, A. J. Cell 9:645-648, 1976.
36. Furuichi, Y. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 73:3448-3452, 1976.
37. Furuichi, Y., Muthukrishnan, S., Tomasz, J. and Shatkin, A. J. J. Biol. Chem. 251:5043-5053, 1976.
38. Desrosiers, R. C., Sen, G. C. and Lengyel, P. Biochem. Biophys. Res. Commun. 73:32-40, 1976.
39. Langberg, S. R. and Moss, B. J. Biol. Chem. 256:10054-10061, 1981.
40. Furuichi, Y., Muthukrishnan, S., Tomasz, J. and Shatkin, A. J. J. Biol. Chem. 251:5043-5051, 1976.
41. Millward, S. and Graham, A. F. Proc. Natl. Acad. Sci. USA 65:422-427, 1970.
42. Zarbl, H., Hastings, K.E.M. and Millward, S. Arch. Biochem. Biophys. 202: 348-355, 1980.
43. Yamakawa, M., Furuichi, Y., Nakashima, K., LaFiandra, A. J. and Shatkin, A. J. J. Biol. Chem. 256:6507-6513, 1981.
44. Morgan, E. M. and Kingsbury, D. W. Biochem. 19:484-491, 1981.
45. Shatkin, A. J., Furuichi, Y., LaFiandra, A. J. and Yamakawa, M. In: Double-stranded RNA viruses (Eds. R. W. Compans and D.H.L. Bishop), Elsevier Biomedical, New York, 1983, pp. 43-54.
46. Morgan, E. M. and Kingsbury, D. W. Virology 113:565-575, 1981.
47. Drayna, D. and Fields, B. N. J. Virol. 41:110-118, 1982.
48. Zweerink, H. J. and Joklik, W. K. Virology 41:501-518, 1970.
49. Lau, R. Y., Van Alstyne, D., Berckmans, R. and Graham, A. F. J. Virol. 16:470-478, 1975.
50. Shatkin, A. J. and Lafiandra, A. J. J. Virol. 10:698-708, 1972.
51. Gaillard, R. K. and Joklik, W. K. Virology 107:533-536, 1980.
52. Lee, P.W.K., Hayes, E. C. and Joklik, W. K. Virology 108:134-146, 1981.
53. Filipowicz, W., Furuichi, Y., Sierra, J. M., Muthukrishnan, S., Shatkin, A. J. and Ochoa, S. Proc. Natl. Acad. Sci. USA 73:1559-1563, 1976.
54. Sonenberg, N. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 74:4288-4292, 1977.

55. Sonenberg, N., Morgan, M. A., Merrick, W. C. and Shatkin, A. J. Proc. Natl. Acad. Sci. USA 75:4843-4847, 1978.
56. Sonenberg, N., Gurtin, D., Cleveland, D. and Trachsel, H. Cell 27:563-571, 1981.
57. Kozak, M. and Shatkin, A. J. Cell 13:201-212, 1978.
58. Kozak, M. Cell 19:79-89, 1980.
59. Morgan, M. A. and Shatkin, A. J. Biochem. 19:5960-5968, 1980.
60. Hickey, E. D., Weber, L. A. and Baglioni, C. Proc. Natl. Acad. Sci. USA 73:19-23, 1976.
61. Adams, B. L., Morgan, M., Muthukrishnan, S., Hecht, S. M. and Shatkin, A. J. J. Biol. Chem. 253:2589-2598, 1978.
62. McDowell, M. J., Joklik, W. K., Villa-Komaroff, L. and Lodish, H. F. Proc. Natl. Acad. Sci. USA 69:2649-2653, 1972.
63. Levin, K. H. and Samuel, C. E. Virology 77:245-259, 1977.
64. Bergmann, J. E. and Lodish, H. F. J. Biol. Chem. 254:459-469, 1979.
65. Samuel, C. E., Farris, D. A. and Levin, K. H. Virology 81:476-488, 1977.
66. Furuichi, Y., LaFiandra, A. J. and Shatkin, A. J. Nature 266:235-239, 1977.
67. Skup, E. and Millward, S. Proc. Natl. Acad. Sci. USA 77:152-156, 1980.
68. Skup, D. and Millward, S. J. Virol. 34:490-496, 1980.
69. Zarbl, H., Skup, D. and Millward, S. J. Virol. 34:497-505, 1980.
70. Sonenberg, N., Skup, D., Trachsel, H. and Millward, S. J. Biol. Chem. 256:4138-4141, 1981.
71. Detjen, B. M., Walden, W. E. and Thach, R. E. J. Biol. Chem. 257:9855-9860, 1982.
72. Both, G. W., Lavi, S. and Shatkin, A. J. Cell 4:173-183, 1975.
73. Levin, K. H. and Samuel, C. E. Virology 106:1-13, 1980.
74. Steitz, J. A. Nature 224:957-961, 1969.
75. Kozak, M. and Shatkin, A. J. J. Biol. Chem. 251:4259-4266, 1976.
76. Kozak, M. and Shatkin, A. J. J. Mol. Biol. 112:75-96, 1977.
77. Kozak, M. and Shatkin, A. J. J. Biol. Chem. 252:6895-6908, 1977.
78. Kozak, M. Nature 269:390-393, 1977.
79. Kozak, M. J. Virol. 42:467-477, 1982.
80. Kozak, M. J. Mol. Biol. 156:807-820, 1982.
81. Shatkin, A. J. and Kozak, M. In:The Reoviridae (Ed. W. K. Joklik), Plenum Press, New York and London, 1983, pp. 79-106.
82. Kozak, M. and Shatkin, A. J. J. Biol. Chem. 253:6568-6577, 1978.
83. Kozak, M. Cell 19:79-90, 1980.
84. Kozak, M. J. Mol. Biol. 144:291-304, 1980.
85. Kozak, M. Cell 34:971-978, 1983.
86. Antczak, J. B., Chmelo, R., Pickup, D. J. and Joklik, W. K. Virology 121:307-319, 1982.
87. Cashdollar, L. W., Esparza, J., Hudson, G. R., Chmelo, R., Lee, P.W.K. and Joklik, W. K. Proc. Natl. Acad. Sci. USA 79:7644-7648, 1982.
88. Richardson, M. A. and Furuichi, Y. Nucleic Acids Res. 11:6399-6408, 1983.
89. Kozak, M. Nucleic Acids Res. 9:5233-5252, 1981.
90. Walden, W. E., Godefroy-Colburn, T. and Thach, R. E. J. Biol. Chem. 256:11739-11746, 1981.
91. Brendler, T., Godefroy-Colburn, T., Carlill, R. D. and Thach, R. E. J. Biol. Chem. 256:11747-11754, 1981.
92. Brendler, T., Godefroy-Colburn, T., Yu, S. and Thach, R. E. J. Biol. Chem. 256:11755-11761, 1981.

93. Godefroy-Colburn, T. and Thach, R. E. J. Biol. Chem. 256:11762-11773, 1981.
94. Weiner, H. L. and Fields, B. N. J. Exp. Med. 146:1305-1310, 1977.
95. Weiner, H. L., Ramig, R. F., Mustoe, T. A. and Fields, B. N. Virology 86:581-584, 1978.
96. Finberg, R., Weiner, H. L., Fields, B. N., Benacerraf, B. and Burakoff, S. J. Proc. Natl. Acad. Sci. USA 76:442-446, 1979.
97. Li, J. K.-K., Keene, J. D., Scheible, P. P. and Joklik, W. K. Virology 105:41-51, 1980.
98. Gaillard, R. K., Li, J. K.-K., Keene, J. D. and Joklik, W. K. Virology 121:320-326, 1982.

10

TRANSCRIPTION OF VESICULAR STOMATITIS VIRUS GENOME RNA

AMIYA K. BANERJEE, BISHNU P. DE, and ANGELES SANCHEZ

Vesicular stomatitis virus (VSV) - a prototype of negative-strand rhabdoviruses, serves as one of the model systems to study transcription and replication of viral genome RNA in vitro. The linear single-strand genome RNA of VSV is tightly associated with approximately 2000 molecules of a nucleocapsid protein designated N (molecular weight, 47,000) (1,2). The N protein renders the genome RNA resistant to ribonuclease and imparts a high degree of stability to the ribonucleoprotein (RNP) complex (3,4). Also associated with the RNP within the matured virions are two minor proteins which are essential for the negative-strand genome RNA to be transcribed into five distinct mRNA species in vitro (5); these include approximately 40 molecules of a large protein designated L (molecular weight, 240,000) and 300 molecules of a phosphoprotein designated NS (molecular weight, 25,000). The transcribing RNP is packaged within the characteristic bullet-shaped lipid-containing shell that contains two other virus-specific proteins: a surface antigen glycoprotein designated G (molecular weight, 69,000) and a membrane-associated protein designated M (molecular weight, 26,000) located underneath the lipid bilayer.

By far, the most interesting feature of VSV is that the purified virion contains a RNA-dependent RNA polymerase which, when activated, synthesizes mRNA in vitro (6,7). Activation is carried out by treating the virus with a nonionic detergent, rendering the RNP accessible to ribonucleoside triphosphates for mRNA synthesis. Moreover, the RNP-associated minor proteins, L and NS, can be removed from the N-RNA complex rendering both fractions transcriptionally inactive (5). Efficient reconstitution of RNA synthesis occurs when L and NS proteins are mixed with the template N-RNA complex (5,8,9). Thus, the VSV system provides a unique opportunity to study the mechanism of viral gene expression and its control in vitro without host involvement. Because only three proteins (L, NS, and N) are involved in

Y. Becker (ed.), VIRAL MESSENGER RNA.

transcription, the precise roles of each protein in this process can be studied. Moreover, since NS is a phosphoprotein, the system is convenient for studying the possible role of protein phosphorylation in gene control.

In vitro transcription process

Over the last ten years, a great deal of work has been carried out to characterize the RNA products transcribed by the RNP in vitro and also to understand the mechanism underlying the transcription process. It has now been unequivocally established that the transcribing RNP synthesizes sequentially from the 3'-end of the genome RNA a small leader RNA (47 bases) (10) followed by sequential synthesis of five 5' capped and 3' polyadenylated mRNA species coding for, in order, the N, NS, M, G, and L proteins (11,12).

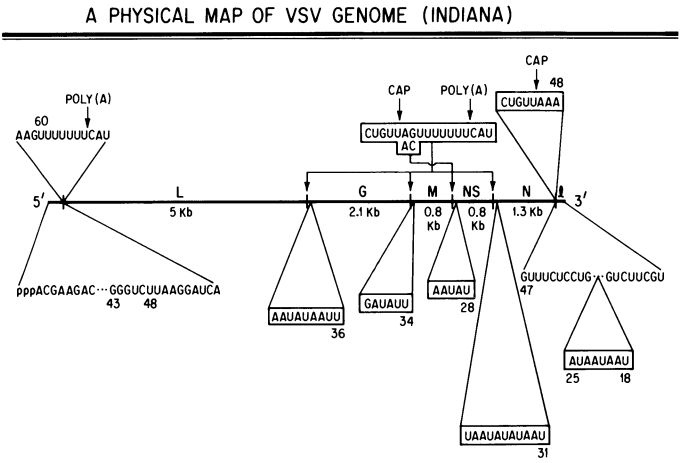

FIGURE 1. L, G, M, NS, and N correspond to viral structural proteins. ℓ represents the region of the genome coding for the leader RNA. The boxed nucleotide sequences (below the genome) are the AU-rich sequences present in the genome RNA, as indicated (16). The boxed nucleotide sequences (above the genome) are the intragenic sequences, as indicated. The G is replaced by C in the corresponding sequence between the M and NS genes.

The leader RNA is uncapped and contains polyphosphate at the 5'-terminus (Fig. 1) (13). Its synthesis terminates at the 47th base from the 3'end of the genome RNA and is present in highest molar amount in the reaction products. There are only three A residues that separate the leader template and the next gene coding for the N-mRNA (14,15). These three A residues are not transcribed in vitro. The start site for the N-mRNA is located next to the A triplet and the termination of the N-mRNA occurs at a sequence, 5' UUUUUUUGCA 3' (16). The polyadenylation at the 3' end of the mRNA probably occurs by repeated slippage

of the polymerase on the U_7 residues. The initiation, termination, and polyadenylation of the subsequent mRNAs occur when the polymerase encounters similar sequences located at identical sites resulting in the synthesis of the matured mRNAs. It should be noted that (a) the terminal 60 nucleotides at the 5' end of the genome RNA are not transcribed *in vitro* (17), and (b) five dinucleotide sequences located between the cap site and the U_7 residues (16) are also not transcribed. Whether 5' capping of the mRNAs occurs concomitantly with the initiation of RNA synthesis or post-transcriptionally (after elongation of a few nucleotides) has not been firmly established.

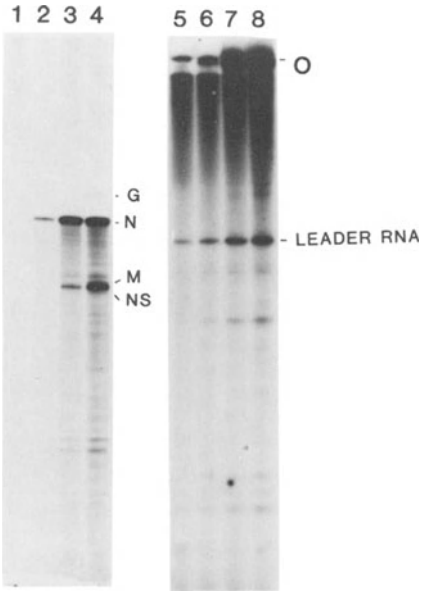

FIGURE 2. Sequential synthesis of mRNA *in vitro* in a reconstituted transcription reaction. Template containing N protein RNA complex (680 µg/ml) and transcriptase containing the L and NS proteins (360 µg/ml) were isolated from purified RNP as detailed elsewhere (28). Equal aliquots of purified template (3.6 µg) were mixed with transcriptase (1.8 µg) in several tubes and RNA syntheses were carried out at 30°C. At different intervals equal aliquots of the reaction products were analyzed for VSV mRNAs in 5% PAGE after removing the poly(A) tails (62) (lanes 1 to 4) and for leader RNA in 20% PAGE (lanes 5 to 8). Reaction times: 5 min, lanes 1 and 4; 10 min, lanes 2 and 5; 20 min, lanes 3 and 6; 40 min, lanes 4 and 8. G, N, M, and NS represent migration positions of corresponding poly(A)-mRNAs.

The unique polar effect on transcription was demonstrated first by using ultraviolet transcription mapping (11,12). However, sequential appearance of individual mRNA species can also be demonstrated directly using transcribing RNP (18) or using a completely reconstituted system. In the latter case, the

RNP-associated L and NS proteins were removed with 0.8 M NaCl (28) Addition of optimum amount of a mixture of L and NS proteins to the N-RNA complex resulted in a sequential appearance with time of the leader RNA and the five mRNAs species. As shown in Fig. 2, the only RNA that was synthesized following 5 min of reconstituted transcription was the leader RNA and no full-length poly(A)-containing mRNAs was detected. However, after 10 min and 20 min of transcription, poly(A)-containing N and NS mRNAs, respectively, appeared sequentially. The M and G mRNAs appeared after 20 min of transcription. Thus,it seems from the above results that the movement of the virion-associated RNA polymerase is somehow attenuated at the junction points between the genes (20). It is, as yet, unclear whether this restriction in the polymerase movement is manifested by the secondary structure of the transcribing RNP (21,22) or by the polyadenylation step at each junction point causing the polymerase to attenuate (20). In the latter case it would be difficult to explain the attenuation observed at the leader RNA-N gene junction where no polyadenylation of the leader RNA is involved. Thus, the precise mechanism by which the virion-associated RNA polymerase maintains the observed sequential appearance and the relative amounts of the mRNAs during transcription in vitro remain poorly understood.

Models for RNA synthesis in vitro

The unique polar effect on transcription in vitro described above suggested that the virion-associated polymerase would most likely initiate at a single site at the 3'-end of the viral genome RNA. Initially, it was proposed that the mRNAs were formed in vitro by cleavage of the growing nascent chains initiated at the 3'-end of the genome RNA (23; Fig. 3, model 1). The capping at the 5'-ends and polyadenylation at the 3'-ends of the cleaved RNAs are carried out by specific virion-associated enzymes. This model was essentially based on two important findings: (a) the ultraviolet transcription mapping experiments indicated a polar effect on transcription (11,12), and (b) the unique 5'-terminal cap structure of VSV mRNAs contained the α and β phosphates of the blocking GTP and the α-phosphate of the 5'-penultimate adenosine, $G_{(5')}^{\alpha\beta}ppp-p^{\alpha}ApA.....$ (24). This structure suggested that the cap was formed by the interaction of GTP with the 5'-end of the mRNAs carrying a single 5'-phosphate residue, i.e. possibly on a cleaved RNA molecule. This model lacked support mainly due to insufficient data to demonstrate cleavage reactions in vitro or isolate large uncleaved precursor RNA in vitro. Moreover,

the decreased order of synthesis of mRNAs in vitro, leader RNA>N>NS>M>G>L (11), clearly indicated that some additional steps were involved in the biosynthesis of individual mRNAs besides processing. The second model termed as start-stop

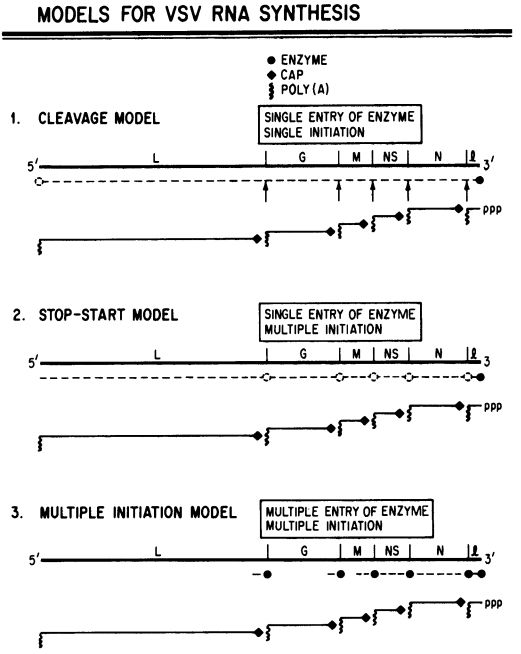

Fig. 3

model (Fig. 3, model 2) proposed that the virion-associated RNA polymerase initiates RNA synthesis at the 3'-end of the genome RNA and synthesizes first the leader RNA. The same enzyme then reinitiates RNA chains on the N gene and continues its synthesis till it terminates chains at the 3'-end and adds poly(A) by slippage on the U_7 residues (see Fig. 1). The same enzyme continues synthesis of NS-mRNA by initiation. The process continues until the enzyme reaches the end of the L gene. The enzyme may fall off the template at any point during transcription but may bind only at the 3'-end of the genome RNA to initiate RNA synthesis. This process will generate a concentration gradient of the RNA species in the order leader RNA>N>NS>M>G>L. The crux of the model is that there is an obligatory requirement for the virion-associated RNA polymerase to enter at a single site at the 3'-end of the genome RNA throughout the transcription process. The evidence for this model has been provided by Emerson (25) who demonstrated that in an incomplete reaction condition containing ATP and CTP, the reconstituted complex synthesized only a dinucleotide pppAC representing the 5'-dinucleotide of the leader RNA. The

oligonucleotide pppAACA representing the 5'-terminal sequence of the mRNAs (26) was synthesized in a similar partial reaction only when RNA synthesis was allowed to proceed (thus allowing the enzyme to move on the template) by the reconstituted complex. Although the model seems highly plausible, it is inconsistent with some findings. For example, the model predicts that no initiation at the N gene would occur unless the complete leader RNA is synthesized. However, it has been shown that under certain experimental conditions both capped (27) and uncapped (18,28,29) oligonucleotides representing the 5'-terminal N-mRNAs are synthesized in the absence of leader RNA synthesis.

An alternate model has been proposed (18) in which the polymerase is located at multiple internal sites in addition to the leader template (Fig. 3, bottom). RNA synthesis occurs by multiple initiations at the putative promoter sites beginning with synthesis of leader RNA. The extension of the initiated mRNA chains occurs sequentially with the synthesis of N-mRNA and so on. The secondary structure of the transcribing RNP is proposed to play a role in this cascade mode of transcription. The bulk of the evidence for this model came from the fact that, in addition to the leader RNA, distinct 5'-polyphosphorylated transcripts representing the 5'-terminal N-mRNA and NS-mRNA were found in the transcription products (18). These transcripts were synthesized very early during transcription and continued to be synthesized with time, whereas the mature mRNAs appeared later and sequentially. Multiple internal bindings of the polymerase were also demonstrated in reconstitution experiments (28). The N-RNA complex was first irradiated for different periods of time and subsequently reconstituted with active soluble transcriptase. The extent of RNA synthesis was decreased by 78% and 88% after 2 min and 3 min of ultraviolet irradiation, respectively. In contrast, the synthesis of the leader RNA and several initiated transcripts representing N and NS mRNA remained virtually constant. These results suggested that each of the RNA species was initiated through independent binding of transcriptase at its putative promoter sites. Also, during reconstitution, at low ionic strength when the matrix protein (M) was present, synthesis of the leader RNA was abolished while the synthesis of small, initiated N-mRNA transcripts remained unaffected (28). The involvement of the secondary structure of the transcribing RNP was demonstrated by using the photoreactive compound, 4-substituted psoralen (21,22). Photoreaction modifies the genome RNA in the N-RNA complex such that at a concentration of 10^{-4} M psoralen the transcription

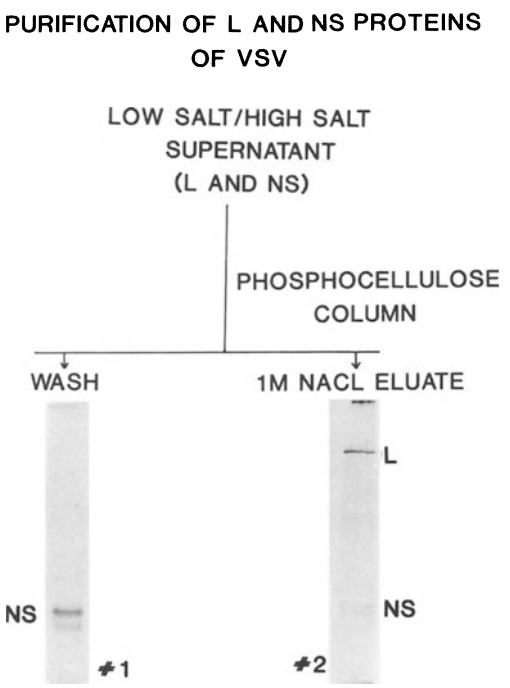

FIGURE 4. Separation of L and NS proteins of VSV. RNP was purified from Triton X-100 disrupted virions. The L and NS proteins were removed from RNP with high concentration of NaCl. The low salt/high salt supernatant was further fractionated by phosphocellulose column chromatography. Concentrations of purified NS protein (#1) and L protein (#2) were 120 µg/ml and 75 µg/ml, respectively (19).

of genome RNA *in vitro* was inhibited by more than 90%. By using [^{3}H]-labeled psoralen, it was shown that photoreaction occurs at a cytosine residue within the N gene near the 3'-end of the viral genome RNA (22). Binding of psoralen to a specific site on the N gene eliminated not only the formation of the N-mRNA but also the transcription of the other viral genes. The synthesis of the small initiated transcripts, however, continued. These results suggested that the secondary structure of the N-gene may have a role in the overall transcription of the genome. Again, as with the other models, several findings are also inconsistent with this model. Failure to chase the initiated transcripts into matured mRNAs (30,31) suggests that these molecules may be abortive products of transcription, although their synthesis may reflect interaction of transcriptase with the template. Thus, understanding of the precise mode of VSV mRNA synthesis *in vitro* still remains unclear. A detailed study on the roles played by the purified L, NS, and N-RNA complex in the

transcription process would certainly shed light on the mechanism of the unique mode of transcription by VSV. In the studies described below we have purified individual components of the transcription complex and studied in detail their requirements, functions, and specificity in the transcription process in vitro.

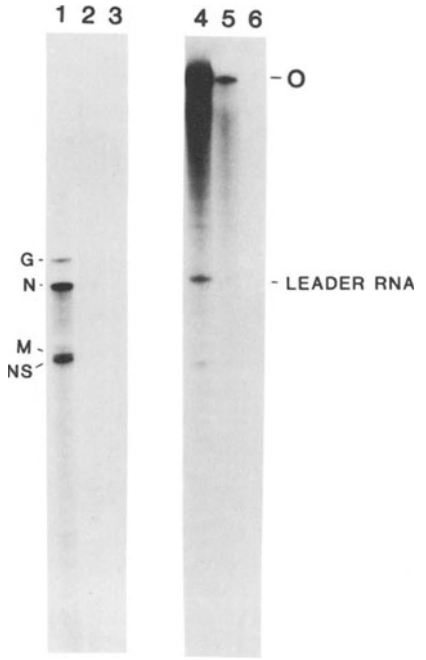

FIGURE 5. Analyses of RNA products synthesized in vitro in reconstitution reactions. RNA syntheses were carried out using saturating concentrations of (0.2 μg) L protein and NS protein (1.4 μg) with N-RNA template (2.2 μg) in the presence of [α-^{32}P]CTP as the labeled precursor. The incubation was at 30°C for 2 hr. Equal aliquots of the reaction products were analyzed for both poly(A)-mRNAs (5% polyacrylamide gel, lanes 1 to 3) and leader RNAs (20% polyacrylamide gel, lanes 4 to 6). Lanes 1 and 4, L+NS+template; lanes 2 and 5, L+template; lanes 3 and 6, NS+template. Migration positions of mRNAs for G, N, M, and NS and leader RNA are indicated. O represents the origin of the gel.

Requirements and possible roles of L and NS proteins in the transcription process in vitro

By using reconstitution experiments, it has been well established that both L and NS proteins are required for mRNA synthesis in vitro (5,9). The L and NS protein appeared to function as a dimeric enzyme since a molar ratio of 1:1 was found to be optimal for transcriptase activity. Although these findings were made some time ago, the precise roles of the individual proteins in the transcriptive process have not been clearly understood. The studies may have

been complicated by the extreme lability of purified L protein (5). Recently, in our laboratory an attempt has been made to purify L, NS, and N-RNA complex to reinvestigate the requirements and possible functions of the individual components in the transcriptive process in vitro.

The L and NS proteins were purified by high salt (0.8 M NaCl) treatment of RNP which initially had been purified by disruption of purified virions with Triton X-100 in the presence of 0.4 M NaCl followed by centrifugation. By this procedure a mixture of L and NS proteins was obtained which was virtually free from G and M proteins (19). The L protein was bound to phosphocellulose column and eluted at 1 M NaCl (Fig. 4). Upon rechromatography on phosphocellulose at least two times traces of NS protein were removed. The purified NS protein, on the other hand, contained a small amount (approximately 5 to 10%) of N protein (Fig. 4). The N-RNA template was purified by treatment of high salt-washed RNP with renografin followed by CsCl banding (19). The RNA products synthesized in a typical reconstitution reaction with L and NS proteins is shown in Fig. 5. Synthesis of mRNAs (lane 1) was monitored by electrophoresing the RNA products through 5% polyacrylamide after removal of poly(A) tails (using oligo(dT) and RNase H), a procedure that enabled better separation of the individual mRNA species. The synthesis of the leader RNA was monitored by electrophoresis of

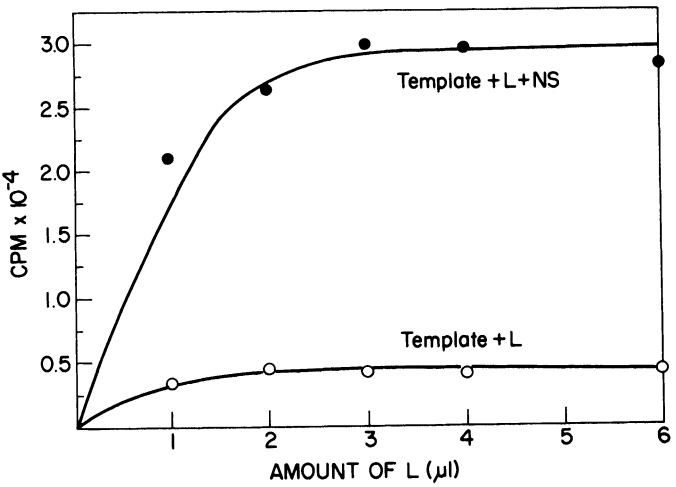

FIGURE 6. Rate of RNA synthesis with saturating NS protein and increasing concentrations of L protein. RNA synthesis was carried out using template (2.2 μg), saturating concentration of NS protein (1.4 μg), and increasing amounts of L protein. The RNA synthesized at each point was determined by measuring cold trichloroacetic acid insoluble radioactivity retained on nitrocellulose filters.

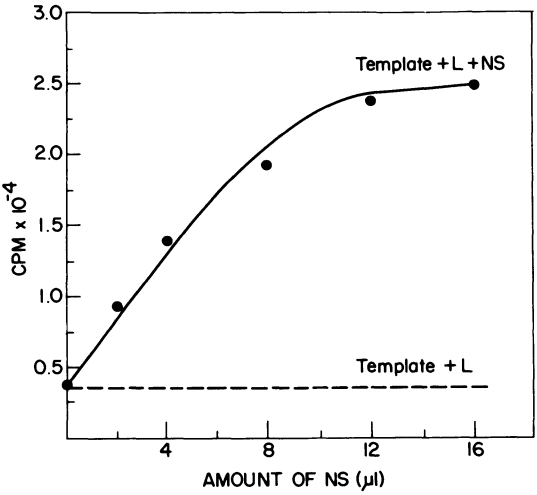

FIGURE 7. Rate of RNA synthesis with saturating L protein and increasing concentrations of NS protein. RNA synthesis was carried out using template (2.2 µg), saturating concentration of L protein (0.2 µg), and increasing amounts of NS protein, as indicated. The RNA synthesized at each point was determined by measuring cold trichloroacetic acid insoluble radioactivity retained on nitrocellulose filters.

the RNA products on a 20% polyacrylamide gel (lane 4). It can be seen that addition of L protein alone to the template produced virtually no complete mRNA species (lane 2), whereas trace amounts of leader RNA and some incomplete RNAs were synthesized (lane 5). The stimulation of RNA synthesis (upon addition of NS protein) ranged between 20- to 30-fold. The residual RNA synthesis obtained using L alone with the N-RNA template (lanes 2 and 5) presumably was due to contaminating NS protein in the L preparation. This conclusion was supported by the fact that repeated chromatography of L protein on phosphocellulose increased the extent of stimulation of transcription (upon addition of NS protein plus N-RNA template) concomitant with the removal of the NS protein from the L protein. The NS protein alone, on the other hand, failed to synthesize any completed mRNA (lane 3), leader RNA or smaller RNAs (lane 6) when added to the N-RNA template. These results confirm the original observations of Emerson and Yu (5) that both L and NS proteins are required for RNA synthesis *in vitro*. The L protein isolated from this virus strain (Indiana serotype, Mudd-Summers strain) was quite stable. Full enzymatic activity was retained after storage in liquid N_2 for more than two weeks. Purified NS protein was highly stable and remained active for many months of storage in

liquid N_2. The N-RNA complex was the most stable of the transcription complex components, retaining its template function for months even when stored at 4°C, although it was routinely stored in liquid N_2.

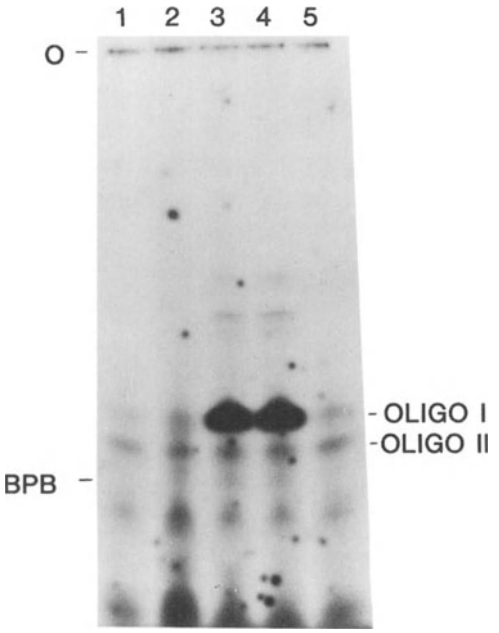

FIGURE 8. Synthesis of oligonucleotides in reconstituted reactions. Oligonucleotide AC (oligoI) and AAC and AACA (oligoII) were synthesized in reconstitution reactions (200 µl) using ATP (1 mM) and 100 µCi of [α-^{32}P]CTP (30 µM) in the standard reaction conditions (30). After incubation for 2 hr at 30°C, the oligonucleotides were isolated and treated with calf intestinal alkaline phosphatase (30), analyzed by electrophoresis in a 20% polyacrylamide gel and autoradiographed. The migration position of bromophenol blue (BPB) is shown. The amounts of the template, L, and NS proteins used in reactions were 2.2 µg, 0.2 µg, and 1.4 µg, respectively. Lane 1, template alone; lane 2, NS+template; lane 3, L+template; lane 4, L+NS+template; lane 5, template+L protein heated at 60°C for 2 min.

At a fixed concentration of template and a saturating concentration of NS protein, RNA synthesis was significantly stimulated by the addition of increasing concentrations of L protein (Fig. 6). However, optimal RNA synthesis occurred at a concentration of 0.2 µg/200 µl reaction (viz. 2.5 µl). In contrast, the requirement of NS protein for optimal RNA synthesis was significantly higher than that for the L protein. As shown in Fig. 7, at a saturating concentration of L, RNA synthesis was optimal at an NS concentration of 1.4 µg/200 µl reaction (viz. 12 µl). These results indicated that the requirement for NS protein is at least seven-fold higher (on a weight basis)

than the L protein requirement for optimal RNA synthesis. Using the molecular weights for L, NS, and N proteins of 240,000 (32), 25,000 (33), and 47,000 (33), respectively, and assuming 2,500 molecules of N protein/RNA (2), the molar ratio of RNA:L:NS determined here is 1:40:2800 (thus N:NS = 1:1). This value is similar to the molar ratio of the protein components found in purified virions (1:40:80, ref. 2), except that the NS protein is present in suboptimal levels.

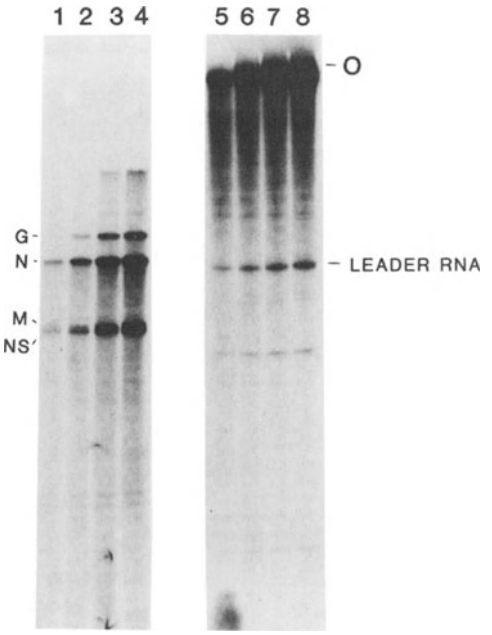

FIGURE 9. Analysis of the RNA products synthesized at saturating L protein concentration and increasing amounts of NS protein. The mRNAs (lanes 1 to 4) and leader RNA (lanes 5 to 8) synthesized in vitro in reconstituted reactions containing template (2.2 µg), L protein (0.2 µg) and various amounts of NS protein were analyzed as in Fig. 5. Amounts of NS protein in lanes 1 and 5 were 0.24 µg; lanes 2 and 6 were 0.48 µg; lanes 3 and 7 were 0.96 µg; and lanes 4 and 8 were 1.4 µg. The migration positions of the mRNAs for G, N, M, NS, and leader RNA are shown. O represents the origin of the gel.

In previous studies (18,25,30) initiation of RNA synthesis by purified or reconstituted RNP was monitored by the synthesis in vitro of pppAC (representing leader RNA initiation) and of pppAAC and pppAACA (representing mRNA initiation) using an incomplete reaction containing ATP and [α-^{32}P]CTP. Similar studies were carried out using purified L, NS and N-RNA template preparations. As shown in Fig. 8, purified template synthesized trace

quantities of AC (oligo I) and AAC+AACA (oligo II) (lane 1). Similarly, very little oligonucleotide was synthesized when purified NS protein was added to the template (lane 2). In contrast, addition of purified L protein to the template significantly increased (20-fold) oligo I and (4-fold) oligo II synthesis (lane 3). Addition of a saturating amount of NS protein to the L fraction did not stimulate the oligonucleotide synthesis (lane 4). Heating the L fraction at 60°C for 2 min abolished the oligonucleotide synthesis (lane 5). These results suggest that the L protein is involved in the initiation of RNA synthesis, and that the function of NS is at some post-initiation step, such as chain elongation.

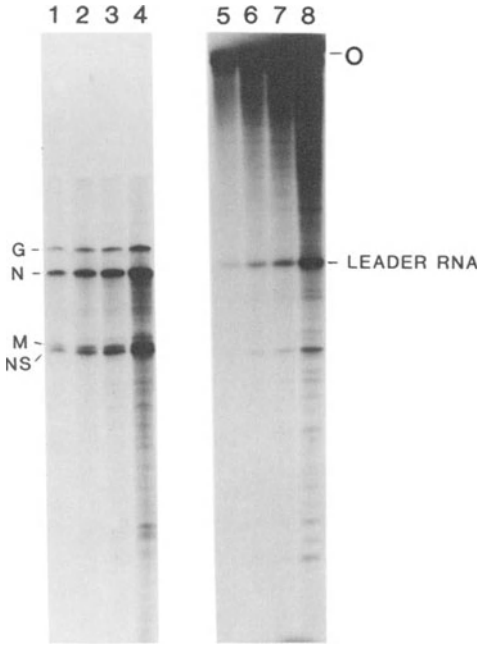

FIGURE 10. Analysis of RNA products synthesized at saturating concentration of NS protein and increasing amounts of L protein. The mRNA products (lanes 1 to 4) and leader RNA (lanes 5 to 8) synthesized at saturating concentration of NS protein (1.4 µg), and increasing L protein amounts were analyzed by electrophoresis in 5% and 20% polyacrylamide gels electrophoresed, as in Fig. 9. Amounts of L protein in lanes 1 and 5 were 0.02 µg; in lanes 2 and 6 were 0.04 µg; in lanes 3 and 7 were 0.08 µg; lanes 4 and 8 were 0.2 µg, respectively. The migration positions of the mRNAs for G, N, M, NS, and leader RNA are shown. O represents the origin of the gel.

In order to understand better the role of NS protein in the transcription process, a series of reconstitution experiments was performed using limiting to saturating amounts of NS or L proteins. Each reaction product was analyzed for

synthesis of individual mRNA species by electrophoresis in 5% polyacrylamide gels and for leader RNA synthesis by analysis on 20% polyacrylamide gels. Figures 9 and 10, and Table 1 show the results of these experiments using various concentrations of NS protein (Fig. 9) and L protein (Fig. 10). When L protein was kept at a constant saturating concentration and the NS protein concentration was increased from limiting (0.24 µg/200 µl) to saturating (1.4 µg/200 µl) (Fig. 9), the ratio of N mRNA to leader RNA increased steadily with the increase of NS protein concentration (Table 1). In contrast, at a constant, saturating concentration of NS protein, the increase of L protein concentration from limiting (0.02 µg/200 µl) to saturating (0.2 µg/200 µl) (Fig. 10) resulted in a constant ratio in the N mRNA to leader RNA over this concentration range (Table 1). These results indicated that at saturating concentrations of L protein, all RNA chains are rapidly initiated, whereas chain elongation was rate limiting due to limiting concentration of NS protein. On the other hand, when NS protein was in excess, the rate limiting step was the initiation of RNA synthesis by L protein. Thus, any RNA chain initiated was quickly extended and completed due to the presence of an excess of NS protein.

Expt. No.	Additions			Radioactivity incorporated (cpm)		Ratio N-mRNA
	N-RNA µg	L µg	NS µg	Leader	N-mRNA	Leader
A	1.5	0.12	0.16	791	652	0.8
	1.5	0.12	0.32	1111	11937	1.7
	1.5	0.12	0.64	1398	6139	4.4
B	1.5	0.015	0.80	176	642	3.6
	1.5	0.030	0.80	269	1044	3.9
	1.5	0.06	0.80	373	1253	3.4

Table 1. Reconstitution reactions were carried out using indicated amounts of N-mRNA complex, L, and NS proteins. The mRNAs and leader RNA synthesized in experiments A and B were analyzed by polyacrylamide gel electrophoresis as described in Figs. 9 and 10, respectively. The labeled bands representing the N mRNA and the leader RNA under each reaction condition were excised from the gel and radioactivity quantitated by Cerenkov counting.

In order to investigate further the role of NS protein in RNA chain elongation, a heat inactivation experiment was performed. Purified L or NS proteins were heated at 40°, 60°, and 80°C for 2 min and used at saturating

concentrations for RNA synthesis in reconstitution experiments. Heating purified L protein at various temperatures virtually abolished synthesis of both completed mRNA (Fig. 11, lanes 4, 5, and 6), and leader RNA synthesis (lanes 11, 12, and 13). In contrast, heated NS preparations were able to synthesize leader RNA as well as a series of small, preterminated RNA species migrating between the origin and the leader RNA (lanes 8, 9, and 10), but synthesis of completed mRNA decreased sharply with the increase of temperature of heating. These results are consistent with the interpretation that heat treatment of NS protein effectively eliminated its capability to elongate and complete the large mRNA chains but had little effect on its capacity to complete small RNA chains. In contrast L, being thermolabile, failed to initiate RNA chains after being heated; consequently, no RNA was synthesized even though excess NS protein was present.

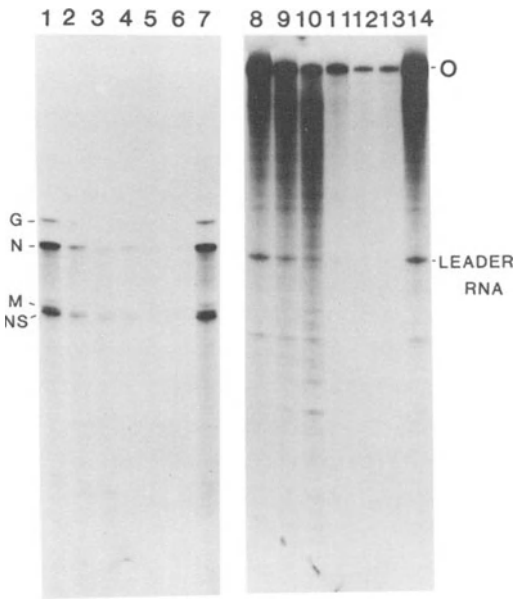

FIGURE 11. Effect of temperature on the L and NS proteins on mRNA and leader RNA synthesis in in vitro reconstitution reactions. The mRNAs (lanes 1 to 7) and leader RNA (lanes 8 to 14) were synthesized using heated L (0.2 µg) or NS proteins (1.4 µg) and template (2.2 µg) and analyzed as described in Fig. 9. NS protein was heated at 40°C for 2 min (lanes 1 and 8); 60°C for 2 min (lanes 2 and 9); 80°C for 2 min (lanes 3 and 10). L protein was heated at 40°C for 2 min (lanes 4 and 11); 60°C for 2 min (lanes 5 and 12); 80°C for 2 min (lanes 6 and 13). Lanes 7 and 14 represent control reactions at 30°C for 2 hr. The migration positions of mRNAs for G, N, M, NS, and leader RNA are shown. O represents the origin of the gel.

It appears from the above results that L protein by itself lacks the ability to synthesize full-length mRNA but possesses the ability to initiate RNA chains. The NS protein, on the other hand, is not able to initiate RNA chains but is involved in promoting chain elongation by L by virtue of its interaction with the L and the N protein. Moreover, it appears that the requirement of NS protein for an optimal rate of transcription is approximately 7-fold higher on a weight basis (or approximately 70-fold higher on a molar basis) than the L protein. This conclusion is different from that reported earlier (5,29) where a 1:1 molar ratio of L and NS proteins was found to be required for optimal RNA synthesis in vitro. One of the reasons for this discrepancy may be due to the extreme lability of the L protein used in the previous reports.

The above results, taken together, strongly suggest that the L protein is the polymerase which interacts with the template and initiates RNA chains while the NS protein is required for chain elongation to produce full-length mRNAs. The NS protein possibly interacts with the L protein as well as with the N protein and alters the template so that the L protein can move along it. Perhaps NS protein is an RNA unwinding protein which displaces the N protein from the genome RNA for L protein to obtain access to the genome. This possibly explains the high molar requirement of NS relative to L. For each L protein molecule, on the order of 70 NS molecules might be required to interact with the region of N-RNA, preceding the moving site of new RNA synthesis. It will be of interest to study whether any other agent, for example, cellular protein or synthetic proteins, can replace the unwinding function of the NS protein in vitro. Stimulation of RNA synthesis in vitro by host cell components has been reported (34). It remains to be seen whether the putative host factors facilitate NS function or can replace it in vitro.

Specificity of interaction of the L and NS proteins with heterologous N-RNA complex

The role of the N protein in the transcription process appears to be to impart stability to the genome RNA by maintaining the structural integrity of the RNP. In addition, direct interaction of L and NS proteins with the N protein is needed for transcription of the genome RNA. In order to probe the specificity of these interactions in vitro, two serologically distinct VSV were used: VSV(Indiana) and VSV(New Jersey). The complete nucleotide sequences of the N-mRNAs of both serotypes revealed an identity of more than 80% of the amino acid sequences between the N proteins when conservative replacements of

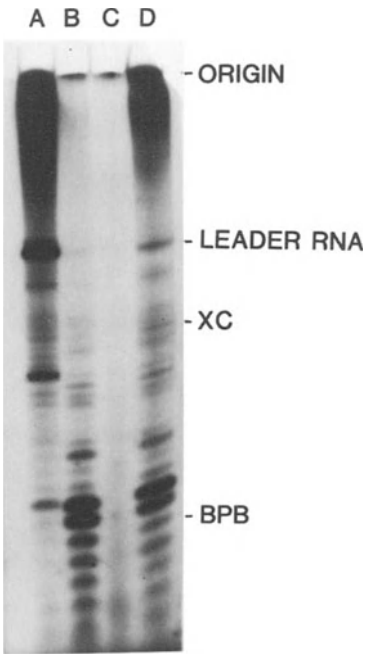

FIGURE 12. Heterologous reconstitution reactions using N-RNA(IND) template. Reconstitution reactions using N-RNA(IND) template in vitro were carried out as detailed elsewhere (19). The RNA products were analyzed by electrophoresis on 20% polyacrylamide gel followed by autoradiography. Reconstitution with N-RNA(IND) template and L(NJ) + NS(NJ) (B); NS(IND) (C); L(NJ) + NS(NJ) + NS(IND) (D); reconstitution with N-RNA(NJ) + L(NJ) + NS(NJ) (A).
(Reprinted with permission from ref. 19).

amino acids were considered (35). Moreover, previous biochemical studies had revealed that the sequences of leader RNA templated by the 3'-end of the genome RNA of these serotypes are very similar (36); there is one base change in the first 24 bases and 9 base changes in the last 23 bases. In addition, the 5'-terminal capped AACAG sequence is conserved in all five mRNAs (37), and extensive homology is observed in the polyadenylic acid adjacent sequences of the N-mRNA of the two serotypes (38). Thus, the N-RNA complex of both serotypes offers an excellent template to probe the specificity of interaction of L and NS proteins with heterologous templates.

When the L and NS fractions of VSV(NJ) were added to purified N-RNA(IND), as shown in Fig. 12 (lane B), no mRNA or leader RNA were synthesized (19). A small amount of oligonucleotides was synthesized, the majority of which migrated with the bromophenol blue dye. However, when purified NS(IND) was included in the transcription mixture, RNA synthesis ensued (lane D). As

expected, purified NS(IND) alone was inactive when added to the N-RNA(IND) (lane C). In a separate series of experiments using purified L and NS proteins of New Jersey, it was shown that a mixture of L(NJ) and NS(IND) was unable to support transcription unless NS(NJ) was present. Thus, it seems that for transcription of N-RNA(IND) complex, the L(IND) protein function can be replaced by L(NJ) whereas homologous NS(IND) was essential to form a stable RNA polymerase complex. The obligatory requirement of NS(NJ) was probably for stabilizing the L(NJ)--N-RNA(IND) complex. If, indeed, the L protein is the RNA polymerase that initiates RNA chains and the NS protein is involved in the RNA chain elongation process (see above), the heterologous resonstitution experiments indicate that N-RNA(IND) template recognized the L(NJ) for initiation of RNA synthesis whereas the chain elongation function was serotype-specific. Thus, there is specific interaction of NS(IND) with its homologous template where L(NJ) and L(IND) share common binding sites that interact with

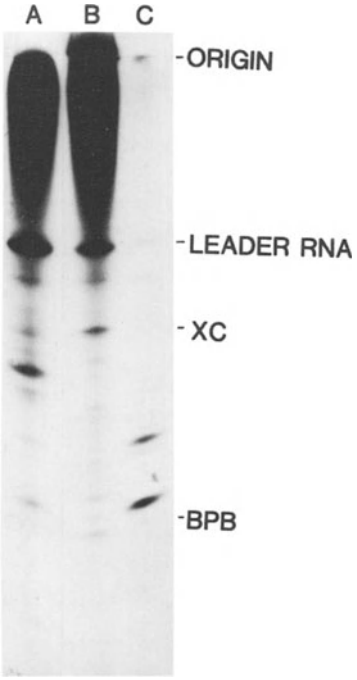

FIGURE 13. Heterologous reconstitution using N-RNA(NJ) template. Reconstitution reactions in vitro were carried out using N-RNA(NJ) + L(NJ) + NS(NJ) (A); N-RNA(IND) + L(IND) + NS(IND) (B); and N-RNA(NJ) + L(IND) + NS(IND) + NS(NJ) (C), as described elsewhere (19). The RNA products were analyzed by electrophoresis on 20% polyacrylamide gel followed by autoradiography. (Reprinted with permission from ref. 19).

the N-RNA(IND) template. In this respect the NS(IND) resembles the sigma factor of bacterial DNA-dependent RNA polymerase (39) which imparts to the core enzyme specificity for binding and initiating RNA chains on the DNA template without having RNA synthesizing capacity by itself.

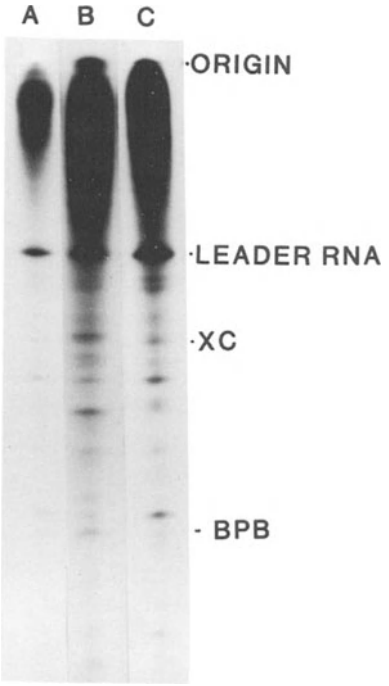

FIGURE 14. Requirement for transcription of N-RNA(NJ) template in heterologous reconstitution reactions. In vitro reconstitution reactions were carried out as in Fig. 13 and RNA products analyzed by polyacrylamide gel electrophoresis. RNA products synthesized by N-RNA(NJ) template with L(NJ) alone (A); L(NJ) + NS(NJ) (B); and L(NJ) + NS(IND) (C), are shown. (Ref. 19, with permission).

The requirements for transcription of N-RNA(NJ) were quite different from those of heterologous N-RNA(IND) (19). As shown in Fig. 13, a mixture of L(IND), NS(IND), and NS(NJ) failed to synthesize any RNA from N-RNA(NJ) template (lane C). These unexpected results suggested that L(IND) and NS(IND) may not form a stable complex with the heterologous N-RNA(NJ) template. Alternatively, NS(NJ) may not interact with L(IND) to perform the chain elongation function. It was further demonstrated that the N-RNA(NJ) template required for RNA synthesis is homologous L(NJ), whereas NS protein could be interchanged. As shown in Fig. 14, when NS(NJ) was added to L(NJ) there was a 6-fold stimulation of RNA synthesis (lane B). Identical stimulation was

obtained with NS(IND) when added to L(NJ) (lane C) indicating that L(NJ) coul interact with the NS protein of both serotypes. The small amount of RN synthesis by L(NJ) alone (lane A) was due to trace contamination with the N protein. The simplest explanation for this specificity is that L(NJ) possesse domains at which both NS(NJ) and NS(IND) interact, whereas L(IND) lacked domain recognized by NS(NJ). Such putative sites or domains in the L protei may be inherently located in the amino acid sequence or be generated b specific interactions with the N proteins and RNA template. Although th precise reasons for these specific requirements for transcription o heterologous templates are not presently understood, the above results clearl indicate that specific interactions between the L and NS proteins and also wit the N-protein RNA complex determine the capability of the complex to transcrib the genome template. It is interesting to note that despite considerabl sequence homology between the N genes of both serotypes, a monoclonal antibod raised against N(IND) failed to cross-react with N(NJ) (40), indicating tha there is an exposed epitope in the N-RNA(IND) complex which is highly specifi to the Indiana serotype.

Phosphorylation of NS protein and transcription of VSV genome RNA in vitro

It is well established that the NS protein is the major phosphoprotein o VSV in infected cells; phosphorylation occurs predominantly in the serine an threonine residues of the polypeptide (41-43). In addition, M protein is als found to be phosphorylated at similar sites in vivo (44). However, in contras to NS protein, M protein contains a small portion of tyrosine residue phosphorylated in vivo (45). Since NS protein, in association with the protein, constitutes the RNA polymerase complex, regulation of phosphorylatio of this protein may play a role in the replicative process of the virus. number of in vitro and in vivo studies have strongly suggested that the degre of phosphorylation of NS protein plays a regulatory role in the in vitr transcription processes (41,46,47,48,49,50).

Phosphorylation of NS protein in vivo is generally considered to b mediated by a cellular protein kinase since purified virions exhibit a protei kinase activity which phosphorylates virus structural proteins in vitro whe detergent-disrupted virions are incubated with [γ-^{32}P]ATP (42-44,48,51). I vitro phosphorylation occurs predominantly at the serine residues of th proteins and interestingly enough, significant phosphorylation at tyrosin residues is also seen in each of the structural proteins (45). These result indicated that possibly a cellular tyrosine-specific protein kinase may hav

been packaged within the virion during the final stages of maturation. Although the virion-associated protein kinases in VSV appear from several studies to be of cellular origin, some results lend credence to the idea that part of the protein kinase activity may be of viral origin (52). Since phosphorylation of NS protein appears from several studies to be important for the transcription process in vitro, it was of interest to study the presence, if any, of protein kinase activity specific for NS protein in purified virions.

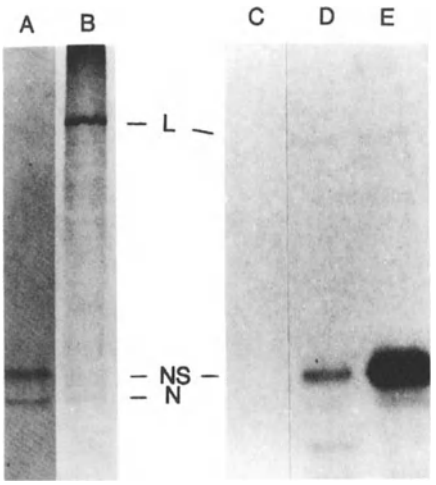

FIGURE 15. Protein kinase activity of purified L and NS proteins. L (lane B) and NS (lane A) proteins were purified by phosphocellulose chromatography and analyzed by polyacrylamide gel electrophoresis followed by silver staining. Aliquots of NS (20 µl, lane C), L (20 µl, lane D), and L+NS (10 µl each, lane E) were assayed for protein kinase activity using [γ-^{32}P]ATP as substrate and labeled protein bands were visualized by autoradiography after polyacrylamide gel electrophoresis (53). Migration positions of L, NS, and N are shown. Concentration of L and NS proteins were 150 µg/ml and 120 µg/ml, respectively.

Purified virions were disrupted with detergent in the presence of 0.4 M NaCl and purified RNP was prepared free from the G and M proteins by centrifugation (53). The L and NS proteins associated with the RNP were subsequently removed by treatment with 0.8 M NaCl. The L and NS proteins were further purified by chromatography on a phosphocellulose column; NS protein was recovered in the flow-through and the bound L protein was recovered by elution with 1 M NaCl. The protein kinase activity in each of these fractions was determined separately and in various combinations. As shown in Fig. 15, lane C, the purified NS fraction contained no demonstrable autophosphorylating

activity. The L protein fraction incorporated a small amount of radioactivity into the "bound" NS protein (lane D). However, when equal aliquots of NS and L fractions were mixed and protein kinase activity assayed, a 30-fold increase of [^{32}P]-incorporation into the NS protein was observed (lane E). These results suggested that the L protein may possess a protein kinase activity, that phosphorylated NS. The L protein (1 µg) incorporated approximately 1 pmol of [^{32}P] per µg of NS protein in 30 min. Subsequent analyses indicated that phosphorylation occurred predominantly at serine residues (95%) with a small amount (5%) in threonine residues in the NS protein. No tyrosine phosphorylation was observed.

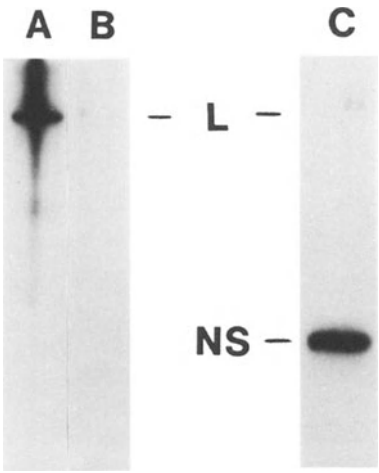

FIGURE 16. Labeling of L protein with [γ-^{32}P]-8-azido ATP. L and NS proteins were obtained as in Fig. 15, except that the L protein was further purified by double chromatography on a phosphocellulose column. Binding of L or NS protein with [γ-^{32}P]-8-azido ATP (10 µM) was carried out in 50 µl of water and exposed to ultraviolet light with a peak wavelength of 364 nm for 30 min. The proteins were precipitated with ice-cold CCl_3COOH and analyzed by electrophoresis on 10% polyacrylamide gels and autoradiographed. Lane A, L protein (15 µl) UV-cross-linked with [γ-^{32}P]-8-azido ATP; lane B, L protein (15 µl) UV-cross-linked with [γ-^{32}P]-8-azido ATP in the presence of unlabeled ATP (70 µM). Lane C, protein kinase activity of L + NS fractions (15 µl each) using [γ-^{32}P]-8-azido ATP as phosphate donor. Autoradiogram exposure time for lanes A and B was 24 hr, whereas in lane C it was 1 hr.

To confirm that the L protein and not some cellular contaminant is involved in mediating phosphate transfer from ATP to the NS protein, we used 8-azido-adenosine-5'-triphosphate as a photoaffinity probe (54,55) to locate the binding site of ATP. Purified L fraction was reacted with 8-azido[γ-^{32}P]ATP in the presence of ultraviolet light and the proteins were

subsequently analyzed by electrophoresis in a polyacrylamide gel. The concentration of labeled azido ATP was kept at 10 μM (Km for protein kinase reaction) which was considerably lower than the Km for ATP (500 μM) in the transcription reaction (56). As shown in Fig. 16, lane A, a distinct labeled band was associated only with the L protein. The label in the band was virtually removed (>90%) when the reaction was carried out in the presence of an excess of unlabeled ATP. In contrast, purified NS protein failed to react with 8-azido-[γ-^{32}P]ATP. The 8-azido-[γ-^{32}P]ATP was effectively used as a phosphate donor, as shown by the labeling of NS protein in the presence of L protein (Fig. 16, lane C). Since no other labeled protein band appeared in the gel, the above results strongly suggest that the L protein possesses a protein kinase activity and, following photoreaction, 8-azido-ATP was bound to its active site.

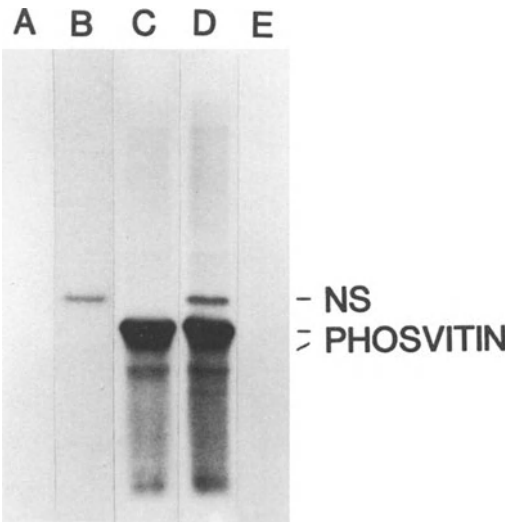

FIGURE 17. Phosphorylation of exogenous phosphate acceptor proteins. L and NS proteins were obtained as in Fig. 15, and the kinase reaction was carried out using [γ-^{32}P]ATP. Lane A, L protein (15 μl); lane B, L+NS proteins (15 μl each); lane C, L protein (15 μl) + phosvitin (20 μg); lane D, L protein (15 μl) + NS protein (15 μl) + phosvitin (20 μg); lane E, NS protein (15 μl) + phosvitin (20 μg). The labeled proteins were analyzed by polyacrylamide gel electrophoresis followed by autoradiography.

To test whether the L-associated protein kinase activity was specific for the NS protein, we used phosvitin as a phosphate acceptor protein. As shown in Fig. 17, purified L protein lacked autophosphorylating activity (lane A), but effectively phosphorylated the exogenously added phosvitin (lane C). It should

be noted that the purified L protein used in these experiments lacked contaminating NS protein. This was due to removal of traces of NS protein by repeated phosphocellulose chromatography of the L protein. The phosphorylation of NS protein (lane B), however, remained virtually unchanged even in the presence of an excess of phosvitin (lane D). Purified NS protein failed to phosphorylate exogenous phosvitin (lane E). Moreover, addition of increasing amounts of phosvitin failed to inhibit phosphorylation of NS by L protein. In contrast, when casein (1 μg) was used as a phosphate acceptor, phosvitin could effectively compete with its phosphorylation. These results suggest that the L protein possesses a protein kinase activity which is specific for phosphorylation of NS protein.

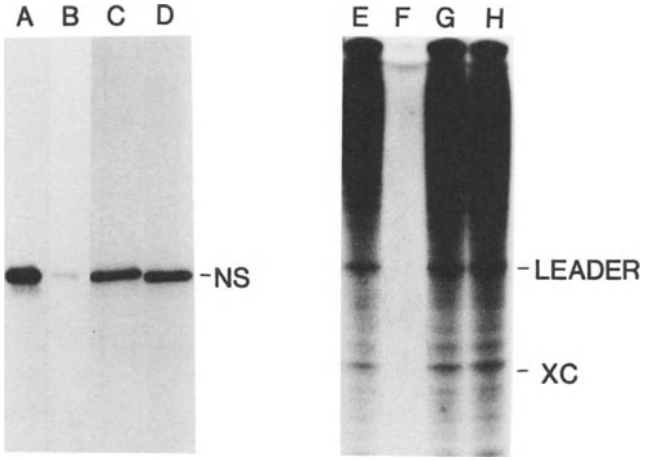

FIGURE 18. Effect of photoreaction of L protein with 8-azido ATP on protein kinase activity and transcription in vitro. Aliquots of purified L (15 μl) and NS proteins (15 μl) (Fig. 15) were photoreacted with unlabeled 8-azido ATP, as described in Fig. 16. Protein kinase activity of the mixture of L and NS proteins was assayed and labeled proteins analyzed by polyacrylamide gel electrophoresis. Lane A, UV-irradiated L fractions + NS protein; lane B, L protein UV-cross-linked with 8-azido ATP + NS protein; lane C, L fraction + UV-irradiated NS protein; lane D, L fraction + NS protein UV-cross-linked with 8-azido ATP.

In vitro RNA synthesis using N-RNA complex reconstituted with L and NS proteins was carried out as described (19). RNA products were labeled with [α-^{32}P]CTP (30 μCi) purified and analyzed by electrophoresis on a 20% polyacrylamide slab gel containing 8 M urea. Lane E, UV-irradiated L fraction + NS protein; lane F, L protein UV-cross-linked with 8-azido ATP + NS protein; lane G, L protein fraction + UV-irradiated NS protein; lane H, L fraction + NS protein UV-cross-linked with 8-azido ATP. Leader represents leader RNA, and migration positions of NS protein and xylene cyanol are shown.

In order to study whether the differential phosphorylated state of NS protein has any contribution to the transcription process in vitro, the L protein was photoreacted with unlabeled 8-azido ATP and its effect on phosphorylation of NS protein was studied. As shown in Fig. 18, lane B, phosphorylation of NS protein was inhibited by more than 90% when L protein was photoreacted, compared to the control reaction (lane A) where L protein was irradiated in the absence of the ATP analog. In contrast, photoreaction of NS protein had no effect on the L-associated protein kinase activity (Fig. 18, lanes C and D). In the same manner, RNA synthesis virtually ceased when photoreacted L was reconstituted with NS and N-RNA complex (lane F), compared to the control experiment (lane E). In contrast, photoreaction of NS protein had no effect on RNA synthesis (Fig. 18, lanes G and H) when reconstituted with L and N-RNA complex. Since phosphorylation of NS and overall transcription were strongly inhibited by photoreaction of L with 8-azido ATP, these results suggest that the degree of phosphorylation of NS protein may have a role in the transcription process in vitro. It is important to note that, in the absence of uv-irradiation, 8-azido ATP (1 mM) cannot substitute for ATP in RNA transcription in vitro and also will not inhibit RNA transcription in the presence of 1 mM ATP.

From the above studies it appears that the L-associated protein kinase activity is probably virally coded. The major part of the activity found associated with the mature virion may be cellular in origin. Several recent reports (41,46,47,48,50,51) are consistent with this contention, in particular the incorporation of src protein into mature VSV grown in baby hamster kidney cells transformed with avian sarcoma virus (44). Although the majority of cellular protein kinases could be removed by extensive purification of RNP, the L-associated protein kinase activity remains strongly bound to the RNP. The precise roles of L protein-associated protein kinase and phosphorylation of NS protein in VSV transcription remain to be explored.

CONCLUSION

The apparently simple structure of VSV N-RNA complex, the ease of dissociability of the L and NS proteins from the RNP, and the restoration of the transcription process by reconstitution make VSV one of the important model systems to study transcription in vitro. It became progressively clear that the interactions of three proteins with the genome template leading to mRNA synthesis is more complex than previously thought. The precise mechanism by

which the polar effect on transcription is maintained is still unclear. Specifically, the mechanism of termination of the leader RNA at the 47th base from the 3'-end remains an enigma. The catalytic and stoichiometric requirements for RNA synthesis of L and NS proteins, respectively, suggest that the secondary structure of RNA undergoes a dynamic change during the transcription process. The NS protein may be involved in maintaining a correct structure of the N-RNA complex which is compatible for transcription by the L protein. It is intriguing to note that during the transcription process, the RNA within the N-RNA complex remains inaccessible to the action of exogenous ribonuclease. Thus, if L or NS proteins are involved in "lifting" the N protein from the template to allow access to a region being transcribed, the site must be immediately protected following transcription. Whether L protein-associated protein kinase activity and differentially phosphorylated NS protein are involved in this process of interaction with the N protein remains to be explored. It is also important to study the domains of L and NS polypeptides that are directly involved in the interaction with N-RNA template. The results presented above clearly indicate that there are specific interactions between the L, NS, and N proteins which lead to mRNA synthesis in a heterologous VSV system.

Although the capping, methylation, and polyadenylation of mRNAs have been demonstrated convincingly in the VSV system, the proteins mediating these processes have not been unequivocally established. A protein-GMP complex, similar to that shown for vaccinia and eucaryotic cell capping enzymes (57-60), has not been obtained in the VSV system. However, NS protein has been shown to tightly bind with GDP (61). Similarly, the mechanism by which the unique cap structure in VSV, $G_{(5')}\overset{\alpha\beta}{pp}\text{-}\overset{\alpha}{p}_{(5')}A$, is formed has not been established. Thus, the VSV transcription system still remains an interesting system where challenging questions remain unresolved and where important new findings are bound to emerge in the future.

REFERENCES

1. Wagner, R.R. In: Comprehensive Virology (Eds. H. Fraenkel-Conrat and R.R. Wagner), Plenum Publishing Corp., New York, 1975, Vol. 4, pp. 1-80.

2. Bishop, D.H.L. and Roy, P. J. Virol. 10:234-243, 1972.

3. Bishop, D.H.L. In: Comprehensive Virology (Eds. H. Fraenkel-Conrat and R.R. Wagner), Plenum Publishing Corp., New York, 1977, Vol. 10, pp. 117-276.

4. Thornton, G.B., Kopchick, J.J., Stacey, D.W. and Banerjee, A.K. Biochem. Biophys. Res. Commun. 116:1160-1167, 1983.
5. Emerson, S.U. and Yu, Y.H. J. Virol. 15:1438-1356, 1975.
6. Baltimore, D., Huang, A.S. and Stampfer, M. Proc. Natl. Acad. Sci. USA 66:572-576, 1970.
7. Moyer, S.A. and Banerjee, A.K. Cell 4:37-43, 1975.
8 Abraham, G. and Banerjee, A.K. Virology 230:230-241, 1976.
9 Naito, S. and Ishihama, A. J. Biol. Chem. 251:4307-4314, 1976.
10. Colonno, R.J. and Banerjee, A.K. Cell 8:197-204, 1976.
11. Abraham, G. and Banerjee, A.K. Proc. Natl. Acad. Sci. USA 73:1504-1508, 1976.
12. Ball, L.A. and White, C.N. Proc. Natl. Acad. Sci. USA 73:442-446, 1976.
13. Colonno, R.J. and Banerjee, A.K. Cell 15, 93-101, 1978.
14. Rowland, D.J. Proc. Natl. Acad. Sci. USA 76:4793-4797, 1979.
15. Keene, J.D., Schubert, M. and Lazzarrini, R.A. J. Virol. 33:789-794, 1980.
16. Rose, J.K. Cell 19:415-421, 1980.
17. Schubert, M., Keene, J.D., Herman, R.C. and Lazzarrini, R.A. J. Virol. 34:550-559, 1980.
18. Testa, D., Chanda, P.K. ande Banerjee, A.K. Cell 21:267-275, 1980.
19. De, B.P. and Banerjee, A.K. J. Virol. 51:628-634, 1984.
20. Iverson, L.E. and Rose, J.K. Cell 23:477-484, 1981.
21. Nakashima, K., Chanda, P.K., Deutsch, V., Banerjee, A.K. and Shatkin, A.J. J. Virol. 32:838-844, 1979.
22. Talib, S. and Banerjee, A.K. Virology 118:430-438, 1982.
23. Banerjee, A.K., Abraham, G. and Colonno, R.J. J. Gen. Virol. 34:1-8, 1977.
24. Abraham, G., Rhodes, D.P. and Banerjee, A.K. Cell 5:51-58, 1975.
25. Emerson, S.U. Cell 31:635-642, 1982.
26 . Rhodes, D.P. and Banerjee, A.K. J. Virol. 17:33-42, 1976.
27. Talib, S. and Hearst, J.E. Nuc. Acids Res. 11:7031-7042, 1983.
28. Thornton, G.B., De, B.P. and Banerjee, A.K. J. Gen. Virol. 63:663-668, 1984.
29. Pinney, D.F. and Emerson, S.U. J. Virol. 42:897-904, 1982.
30. Chanda, P.K. and Banerjee, A.K. J. Virol. 39:93-103, 1981.
31. Lazzarrini, R.A., Chien, I., Yang, F. and Keene, J.D. J. Gen. Virol. 58, 429-441, 1982.
32. Schubert, M., Harmison, G.G. and Meier, E. J. Virol. 51:505-514, 1984.
33. Gallione, C.J., Greene, J.R., Iverson, L.E. and Rose, J.K. J. Virol. 39:529-535, 1981.
34. Rose, J.K., Lodish, H.F. and Brock, M.L. J. Virol. 21, 683-693, 1977.
35. Banerjee, A.K., Rhodes, D.P. and Gill, D.S. Virology 137, 432-438, 1984.

36. Colonno, R.J. and Banerjee, A.K. Nuc. Acids Res. 5:4165-4176, 1978.
37. Franze-Fernandez, M.T. and Banerjee, A.K. J. Virol. 26:179-187, 1978.
38. Rhodes, D.P. and Banerjee, A.K. Virology 105:297-300, 1980.
39. Chamberlin, M.J. In: RNA Polymerase (Eds. R. Losick and M.J. Chamberlin), Cold Spring Harbor Laboratory Publications, Cold Spring Harbor, New York, 1976, p. 1768.
40. De, B.P., Tahara, S.M. and Banerjee, A.K. Virology 122:510-514, 1982.
41. Clinton, G.M., Burge, B.W. and Huang, A.S. J. Virol. 27:340-346, 1978.
42. Imblum, R.L. and Wagner, R.R. J. Virol. 13:113-124, 1974.
43. Moyer, S.A. and Summers, D.F. J. Virol. 13:455-465, 1974.
44. Clinton, G.M., Guerina, N.G., Guo, H. and Huang, A.S. J. Biol. Chem. 257:3313-3319, 1982.
45. Clinton, G.M. and Huang, A.S. Virology 108:510-514, 1981.
46. Hsu, C-H, Morgan, E.M. and Kingsbury, D.W. J. Virol. 43:104-112. 1982.
47. Kingsford, L. and Emerson, S.U. J. Virol. 33:1097-1105, 1980.
48. Sinacore, M.S. and Lucas-Lenard, J. Virology 121:404-413, 1982.
49. Testa, D., Chanda, P.K. and Banerjee, A.K. Proc. Nat. Acad. Sci. USA 77:294-298, 1980.
50. Witt, D.J. and Summers, D.F. Virology 197:34-49, 1980.
51. Harmon, S.A., Marnell, L.L. and Summers, D.F. J. Biol. Chem. 258:15283-15290, 1983.
52. Tan, K.B. Virology 64:566-570, 1975.
53. Sanchez, A., De, B.P. and Banerjee, A.K. (Submitted), 1984.
54. Banerjee, R.K. and Racker, E. J. Biol. Chem. 252:6700-6706, 1977.
55. Haley, B.E. Fed. Proc. 42:2831-2836, 1983.
56. Testa, D. and Banerjee, A.K. J. Biol. Chem. 254:2053-2058, 1979.
57. Schuman, S. and Hurwitz, J. Proc. Nat. Acad. Sci. USA 78:187-191, 1981.
58. Venkatesan, S. and Moss, B. Proc. Nat. Acad. Sci. USA 79:340-344, 1982.
59. Wang, D., Furuichi, Y. and Shatkin, A.J. Mol. Cell. Biol. 2:993-1001, 1982.
60. Mizumoto, K., Kaziro, Y. and Lipmann, F. Proc. Nat. Acad. Sci. USA 79:1693-1697, 1982.
61. De, B.P. and Banerjee, A.K. Biochem. Biophys. Res. Commun. 114:138-147, 1983.
62. De, B.P., Thornton, G.B., Luk, D. and Banerjee, A.K. Proc. Nat. Acad. Sci. USA 79:7137-7141, 1982.

11

INFLUENZA VIRAL RNA TRANSCRIPTION

ROBERT M. KRUG, ANN BEATON, JANET BRAAM AND MICHAEL KATZE

SUMMARY

Influenza viral messenger RNA (mRNA) synthesis is initiated by cap 1 ($m^7GpppNm$)-containing primers generated from heterologous RNAs. We have elucidated the functions and movements of the three viral P (PB1, PB2 and PA) proteins during viral mRNA synthesis catalyzed by purified virion nucleocapsids in vitro. Viral mRNA synthesis occurs in the nucleus of infected cells, leading to: (i) disruption of the metabolism of host cell RNA polymerase II transcripts; and (ii) the modification (methylating of internal A residues and splicing) of viral mRNAs by host nuclear enzymes. In the infected cell, another type of viral transcript, full-length transcripts, are synthesized, which, unlike the viral mRNAs, are initiated without a primer and contain a copy of the last 17-22 nucleotides at the 5' ends of the virion RNAs. We describe the development of an in vitro system that catalyzes the synthesis of these transcripts, the presumed templates for virion RNA synthesis. Finally, we will show that influenza virus establishes a translational system that selectively translates viral and not host mRNAs.

INTRODUCTION

Influenza virus, which has a segmented RNA genome of negative polarity, employs a unique mechanism for the synthesis of its viral messenger RNAs (mRNAs). This synthesis requires primers generated from heterologous cap 1 ($m^7GpppNm$)-containing RNAs (1-6). These RNAs are cleaved by a viral endonuclease at a purine residue 10 to 13 nucleotides from the cap, and the resulting capped fragments serve as primers. Transcription is initiated by the addition of a G residue onto the 3' end of the capped primer fragments, directed by the

Y. Becker (ed.), VIRAL MESSENGER RNA.

penultimate C residue of the virion RNA (vRNA) templates (4). The viral mRNA chains are then elongated until a site of termination and polyadenylate (poly A) addition is reached 17-22 nucleotides from the 5' ends of the vRNA templates (7-9).

This process is catalyzed *in vitro* by viral nucleocapsids (or cores) (4), which contain four virus-specific proteins (10,11). The nucleocapsid (NP) protein, which is the predominant species (representing over 90% of the total protein), presumably serves largely a structural role. The other three viral proteins, the P proteins, distribute during two-dimensional gel electrophoresis into two basic species, PB1 and PB2, and one acidic species, PA (11,12). Several experimental approaches have established that one of the P proteins, the PB2 protein, is the cap 1-recognizing protein. Ultraviolet (UV) light-induced crosslinking studies employing a capped RNA labeled with ^{32}P only in its cap structure showed that the PB2 protein is closely associated with the cap during the endonuclease reaction (11). In addition, it was shown that PB2 specifically binds to a photoreactive derivative of m^7GTP (13). Subsequent experiments using temperature-sensitive virus mutants established that the physical association of the PB2 protein with the cap 1 structure seen in crosslinking studies reflects a functional association. Temperature-sensitive mutants with a defect in the PB2 protein were shown to exhibit temperature-sensitivity in cap recognition during viral mRNA synthesis *in vitro* (14).

The ability of UV-crosslinking experiments to discern the function of the PB2 protein in viral mRNA synthesis prompted us to use this methodology to delineate the functions and movements of the three viral P proteins during viral mRNA synthesis (15). These studies, which will be summarized here, allowed us to formulate a coherent model for the functions and movements of the viral P proteins as they transcribe the influenza vRNA segments into viral mRNAs (15).

Viral mRNA synthesis occurs in the nucleus of infected cells (16,17), utilizing capped RNA primers cleaved from newly synthesized polymerase II transcripts (16). As a result of the cleavage of their 5' ends, the subsequent processing and transport of polymerase II transcripts might be affected. If a significant proportion of

various polymerase II transcripts were cleaved, then the production of new host cell mRNAs might be inhibited. Initial experiments indicate that this is in fact the case for at least two polymerase II transcripts.

The interaction with host cell nuclear functions continues after the viral mRNAs are synthesized. Strong evidence has been obtained that two of the viral mRNAs, those coding for the M1 and NS1 proteins, are themselves spliced to form smaller mRNAs (18). Several lines of evidence indicate that cellular enzymes catalyze this splicing: (i) the splice junctions in the viral mRNAs are closely similar to those found in RNA polymerase II-catalyzed transcripts (19-21); and (ii) the RNA transcripts copied off cloned viral DNAs (coding for M1 and NS1) that had been inserted into simian virus 40 vectors are also spliced, utilizing the same splice junctions (22). This work has been carried out by Lamb and coworkers and has been summarized elsewhere (18,23,24). Cellular enzymes in the nucleus have also been implicated in the other major modification of influenza viral mRNAs, the methylation of their internal A residues to form N6 methyl adenosine (m^6A). In the total viral mRNA polulation, there are approximately two m^6A residues per chain in the virus-coded region (i.e., after the host-donated primer sequences have been removed) (25,26). Our results indicate that the two m^6A residues in the virus-coded region are found in the same sequences (GAC and AAC) as in DNA-directed RNA polymerase II transcripts. The role of internal m^6A residues is not known, but it has been postulated that these residues, which have been shown to occur at a specific subset of the available GAC and AAC sequences in several DNA-directed transcripts, are involved in RNA processing and/or transport from the nucleus (27-30). If m^6A residues serve a role in the splicing of influenza viral mRNAs, the simplest situation would be that only those viral mRNAs that undergo splicing (the M1 and NS1 mRNAs) would contain m^6As in the virus-coded regions. We will show here that all eight of the major viral mRNAs contain internal m^6As.

Because they contain host-derived 5' sequences and are not complete copies of the vRNA segments, the viral mRNAs are not suitable templates for vRNA replication. The presumed template RNAs, or full-length transcripts, have been identified in infected cells

(7,16,31). These RNAs, unlike the viral mRNAs, are initiated without a primer and contain a copy of the last 17-22 nucleotides at the 5' ends of the vRNAs (7,31). The template RNAs are not polyadenylated and require ongoing protein synthesis, presumably virus-coded, for their synthesis (31,32). In order to understand the mechanism of the switch from viral mRNA to template RNA synthesis - the first step in vRNA replication, it is essential to develop an *in vitro* system in which the synthesis of both template RNA and viral mRNA occurs. We will describe here the development of such an *in vitro* system (33).

In addition to its actions in the nucleus, influenza virus might be expected to affect the expression of host cell mRNAs in the cytoplasm, as host cell protein synthesis is shut off during infection (34,35). This would presumably involve some mechanism for blocking the translation of host mRNAs synthesized both before and after infection. We will show that this is accomplished by the establishment of an influenza virus-specific translational system that selectively translates viral and not host mRNAs (36).

RESULTS AND DISCUSSION

Functions and Movements of the Viral P Proteins During Capped RNA-Primed Viral mRNA Synthesis. The initiation of viral mRNA synthesis occurs via the incorporation of a G residue onto the 3' ends of the heterologous capped fragments generated by the viral endonuclease (4). The viral protein catalyzing this initiation should be closely associated with this residue. Similarly, the viral protein responsible for the subsequent addition of ribonucleotides during chain elongation would be expected to be at the very 3' end of the growing transcript, coincident with the addition of the 3' terminal residue. To identify the protein catalyzing initiation, we incubated purified viral nucleocapsids with a capped primer in the presence of (α-^{32}P)GTP as the only ribonucleoside triphosphate. Under these conditions, one or a few G residues are added to the capped primer fragment (Figure 1A, lane 1) (4,11,15). To identify the elongation protein, we performed transcription reactions that result in the synthesis of incomplete mRNA chains containing ^{32}P-label only in the 3' terminal nucleotide (15). This was possible to accomplish because the initial part of the viral-coded region of

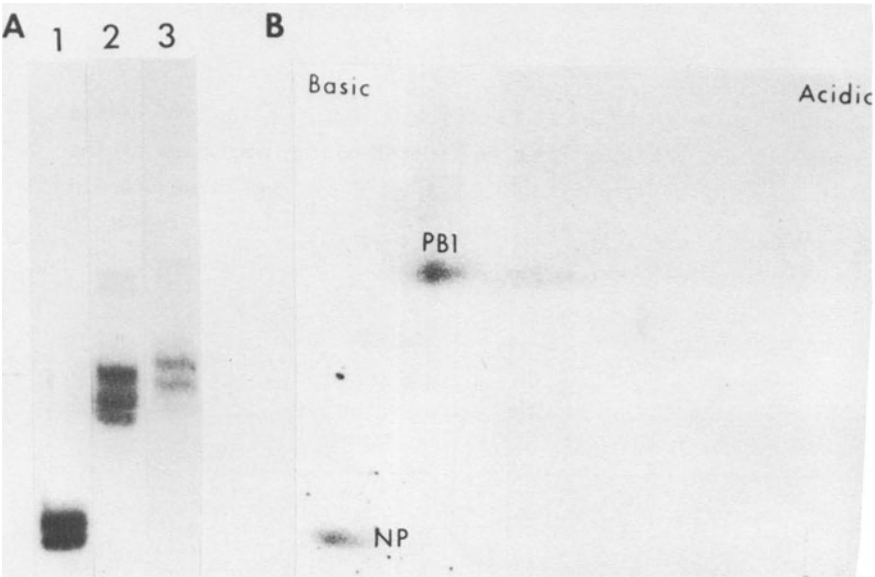

FIGURE 1. Identification of the P Protein that Catalyzes the Addition of the 3' Terminal Nucleotide onto Growing Viral mRNA Chains. Viral cores purified as described previously (4,11) were incubated for 20 min at 31C in a 100µl reaction mixture containing AlMV RNA 4 and unlabeled GTP, CTP and ATP, each at 100µM. These cores with the nascent chains were freed of triphosphates by spin chromatography through Sephacryl S-200, and were then incubated for 20 min at 31C in the presence of (α-^{32}P) UTP (10µM, 300 Ci/mmole) as the only ribonucleoside triphosphate. (A) The labeled RNA products of this reaction were electrophoresed on a 20% acrylamide 7M urea gel (lane 3). The RNAs synthesized in two other 100µl reaction mixtures were analyzed on the same gel: viral cores incubated for 20 min at 31C with AlMV RNA 4 and either (α-^{32}P) GTP alone (10µM, 15 Ci/mmole) (lane 1) or three ribonucleoside triphosphates, (α-^{32}P) GTP, CTP and ATP (lane 2). (B). After removal of labeled (α-^{32}P) UTP from the first reaction described above using spin chromatography, the sample was irradiated at 254nm at 0C (dose: 6×10^4 ergs/mm^2); digested for 2 hr at 37C with pancreatic, T1 and T2 RNases; and then analyzed by two-dimensional gel electrophoresis (11,15). From (15) with permission.

---

each mRNA chain lacks uridine (U) residues (Figure 2). Our strategy then was to incubate viral cores with a capped RNA primer (alfalfa mosaic virus RNA 4 [AlMV RNA 4]) and unlabeled GTP, CTP and ATP and then to isolate the viral cores containing the nascent partial mRNAs

mRNA SEGMENT	PRIMER REGION	VIRAL-CODED REGION
1, 3	m^7GpppGm. . . A	G C G A A A G C A G G (U)
7	m^7GpppGm. . . A	G C A A A A G C A G G (U)
5, 8	m^7GpppGm. . . A	G C A A A A G C A G G G (U)
4	m^7GpppGm. . . A	G C A A A A G C A G G G G (U U)
6	m^7GpppGm. . . A	G C A A A A G C A G G A G (U U)
2	m^7GpppGm. . . A	G C G A A A G C A G G C A A A C C A (U U U)

FIGURE 2. The 5' Terminal Sequences of the Eight Viral mRNA Chains. The viral mRNA chains or segments are numbered 1 to 8 in order of decreasing molecular weight. The sequence of the 5' viral-coded region of the mRNAs is from Robertson (37). From (15) with permission.

free from unincorporated triphosphates. Transcription was then resumed by the addition of (α-^{32}P)UTP as the only ribonucleoside triphosphate. As predicted, the size of the RNA (Figure 1A, lane 3) is 1-3 nucleotides longer than the partial transcripts terminated prior to the first U residue (lane 2) (15).

To identify the protein(s) closely associated with either the initial labeled G residue or with the labeled U residue at the end of the partially elongated chains, the two reaction mixtures described above were exposed to UV light irradiation to covalently crosslink the labeled nucleotide to the protein(s) closely associated with this nucleotide (11,15). After nuclease digestion, the labeled protein(s) was identified by two-dimensional polyacrylamide gel electrophoresis. In both reactions, PB1, the larger of the two basic P proteins, is essentially the only P protein labeled (Figure 2B shows the results for the partial elongation reaction). The crosslinking of NP occurs to varying degrees in different experiments and is almost certainly nonspecific. We have shown that NP, because it constitutes greater than 90% of the protein in viral nucleocapsids, crosslinks nonspecifically to labeled primer fragments and product RNAs (11). Thus, PB1 is the protein which is closely associated with the initial G residue added to the primer fragment and also with the 3' terminal residue added to growing mRNA chains.

Pulse-chase experiments directly demonstrated that the PB1 protein initially associated with the initiating G residue moves away from this residue as soon as elongation begins (15). Our results indicate that PB1 is moving down the growing viral mRNA chain and is found associated with the last residue added to the chain. Consequently, PB1 most likely catalyzes each nucleotide addition to the growing mRNA chain, specifically including the first G residue. This conclusion has subsequently been confirmed by others, using pyridoxal phosphate to label the nucleotide binding site of PB1 (38).

To determine whether either of the other two P proteins are also associated with the growing mRNA chains, we synthesized uniformly labeled partial transcripts terminated before the first viral-encoded U residue and exposed this reaction mixture to UV light (15). All three P proteins were associated with the uniformly labeled partial transcripts (Figure 3), strongly suggesting that the P proteins are present as a complex. If this were the case, the three P proteins should move together from this initial region of the viral mRNA chains to downstream regions as the mRNA chains are further elongated. These initial partial chains can be extended to various lengths by a second incubation with unlabeled UTP and two other unlabeled triphosphates. After this second incubation, the close association of the three P proteins with the initial partial chains was greatly reduced (80-90%) (15), indicating that these proteins move away from the initial labeled portion of the partial chains after a subsequent elongation. By labeling the extended region, and not the initial region, of the partial chains, we showed directly that the three P proteins move to the extended region (15). Consequently, the three P proteins move down the mRNA chains together during elongation.

Since the PB2 protein is part of this complex, we determined whether the PB2 protein remains with, or dissociates from, the cap during chain elongation. We used capped AlMV RNA 4 labeled with ^{32}P only in its cap and carried out transcription reactions which yielded mRNA chains extended to various lengths. As before, UV-crosslinking was employed to monitor the association of the PB2 protein with the labeled cap. It was found that PB2 dissociates from the cap soon after 11 to 13 nucleotides have been added to the capped primer

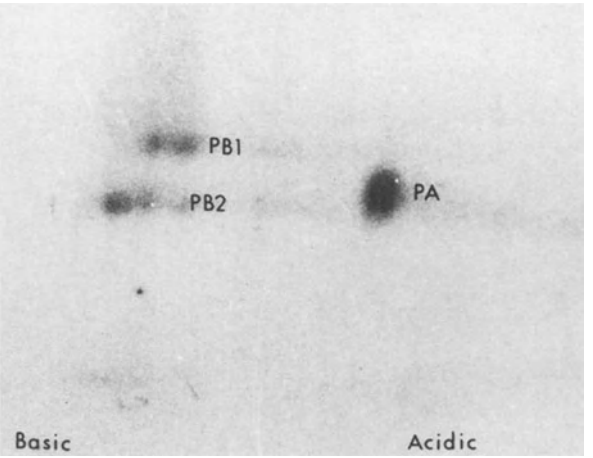

FIGURE 3. Identification of the P Proteins Associated with the Growing Viral mRNA Chains. Viral cores were incubated for 20 min at 31C with AlMV RNA 4 and three (α-^{32}P)-labeled ribonucleoside triphosphates, GTP, CTP and ATP, each at 20μM and 100 Ci/mmole, in a final volume of 100μl. The labeled triphosphates were removed by spin chromatography, and after UV irradiation and nuclease digestion, the proteins were analyzed by two-dimensional electrophoresis. From (15) with permission.

---

fragments (15).

Because of our results, we can formulate a coherent model for the functions and movements of the three influenza virus P proteins during capped RNA-primed transcription (15). This model is shown in Figure 4. The top diagram shows the most likely arrangement of the P proteins during the generation of capped primer fragments. The P proteins are in the form of a complex which must be located at the 3' end of the vRNA templates where transcription initiates. Presumably, individually or cooperatively, the P proteins recognize and bind to the common 3' terminal sequence of the vRNA templates. The PB2 protein in this complex recognizes and binds to the cap 1 structure of the primer RNA, allowing cleavage of the primer at a purine residue 10-13 nucleotides from the cap. We do not yet know which protein catalyzes this cleavage. The simplest idea is that PB2, the cap-recognizing protein, spans the 5' terminal 10-13 nucleotides of the primer RNA and is capable by itself of catalyzing the cleavage reaction. Alternatively, the PB2 protein might catalyze this

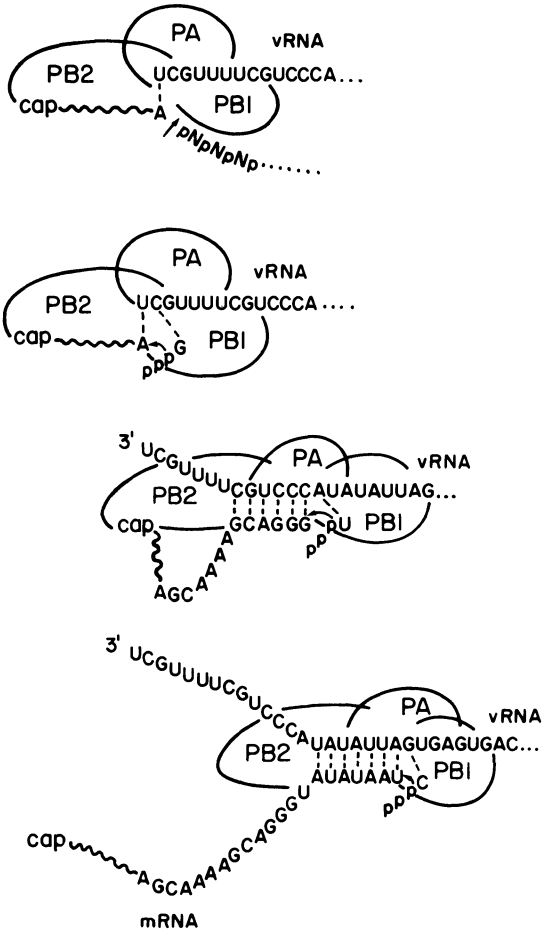

FIGURE 4. Model of the Functions, Interactions, and Movements of the Three P Proteins during Capped RNA-Primed Viral mRNA Synthesis. The sequence shown is that of vRNA and mRNA chain number 5 coding for the nucleocapsid protein. From (15) with permission.

---

cleavage only when in association with one or both of the other two proteins; or one of the other P proteins might catalyze cleavage. The PB1 protein is a possible candidate, as it is located at the appropriate position at the 3' end of the primer fragment poised to add the first G residue (second diagram). The third diagram shows

that after PB1 adds the initial G residue, it moves down the elongating viral mRNA chains in a complex with PA and PB2. The PB1 protein leads the way, positioned at the growing end of the mRNA chain, where it most likely catalyzes the addition of each nucleotide to the chain. Although the PA protein is part of the complex, we have not identified a specific catalytic function for this protein in viral mRNA synthesis. For the first 11 to 15 nucleotides of chain growth, PB2 is bound to both the growing chain and the cap, possibly causing the formation of a loop at the 5' end of the growing chain. Subsequently, PB2 dissociates from the cap, and the polymerase complex continues to move down the chain (last diagram). This complex has an aggregate molecular weight of about 255,000 and must thread its way down the vRNA templates through the multiple sites at which the NP protein (56,000 molecular weight) is associated with the vRNAs. The complex continues down the chain until it reaches a tract of 5 to 7 U residues about 17-22 nucleotides from the 5' ends of the vRNAs (9). At this site, polyadenylation occurs, presumably resulting from stuttering or repetitive copying of the U tract. Transcription then terminates and the viral mRNA chains are released from the viral nucleocapsids (15). It is not known what happens to the P protein complex at this point. Does it fall off the nucleocapsids and/or recycle to initiate viral mRNA chains on the same or a different vRNA nucleocapsid template? This remains an important unanswered question about viral mRNA synthesis.

Interactions of Viral mRNA Synthesis with Host Cell Nuclear Functions. Because influenza viral mRNA synthesis occurs in the nucleus utilizing primers derived from newly synthesized polymerase II transcripts (16), the metabolism of various, or all, polymerase II transcripts might be affected as a result of virus infection. To examine this, we have been focusing on several specific transcripts synthesized in chicken embryo fibroblasts (CEFs) and HeLa cells. In CEFs, we have been examining the effects of virus infection on the metabolism of the transcripts coding for β-actin and for avian leukosis virus (ALV) proteins. Proviral DNA was integrated into host cell DNA as a result of prior infection by ALV. In initial experiments the kinetics of labeling (with ^{3}H uridine) of these two transcripts synthesized in infected cells were measured. The results

indicate that within one to two hours after infection these newly synthesized transcripts are degraded in the nucleus and that little or no new β-actin and ALV mRNAs appear in the cytoplasm. Experiments are in progress to determine whether other polymerase II transcripts in CEFs and HeLa cells are similarly affected. If most of these other transcripts are also degraded in the nucleus, this would indicate that influenza virus effectively shuts down the production of new host cell mRNAs.

Influenza viral mRNAs synthesized in vivo contain on the average three m^6A residues per viral mRNA chain, with two being in the virus-coded region (25,26). Because of the postulated role of m^6A in splicing, one possibility was that only those viral mRNAs that are subsequently spliced, namely the NS1 and M1 viral mRNAs, contain m^6A residues in the virus-coded region. The previous estimate of two m^6A residues per viral mRNA chain was based on an analysis of total (unfractionated) viral mRNA (25,26).

To examine the virus-coded region only, we annealed poly(A)$^+$ ^{3}H-methyl labeled in vivo mRNA to an excess of vRNA (39). These hybrids were then digested with T2 RNase to remove the host-coded primer sequences found at the 5' end of the in vivo mRNA and to digest any host cell RNAs that might be present. To ensure that the integrity of the 5' end of the viral-coded region was maintained, we carried out a titration of T2 RNase using double-stranded viral RNA preparations in which the 3' end of the vRNA was labeled by ligation to [5'-^{32}P]pCp. We chose a level of T2 RNase which preserved the 3' end of the vRNA and hence the 5' end of the mRNA in the viral hybrids resolved by gel electrophoresis. When hybrids of vRNA and ^{3}H-methyl-labeled in vivo viral mRNA were hydrolyzed with T2 RNase under these conditions, ^{3}H-radioactivity was found in each of the vRNA-mRNA hybrids (Figure 5A) (39), strongly suggesting that all of the viral mRNAs, both those that are and those that are not spliced to form smaller mRNAs, contain internal m^6A residues. To verify that the ^{3}H-radioactivity was indeed in m^6A, hybrids containing the NP mRNA (which is not subsequently spliced) and NS1 mRNA (which is subsequently spliced) were eluted from the gel, heat-denatured and digested with P1 nuclease and calf intestine phosphatase. The digests were analyzed by thin-layer chromatography. As shown in

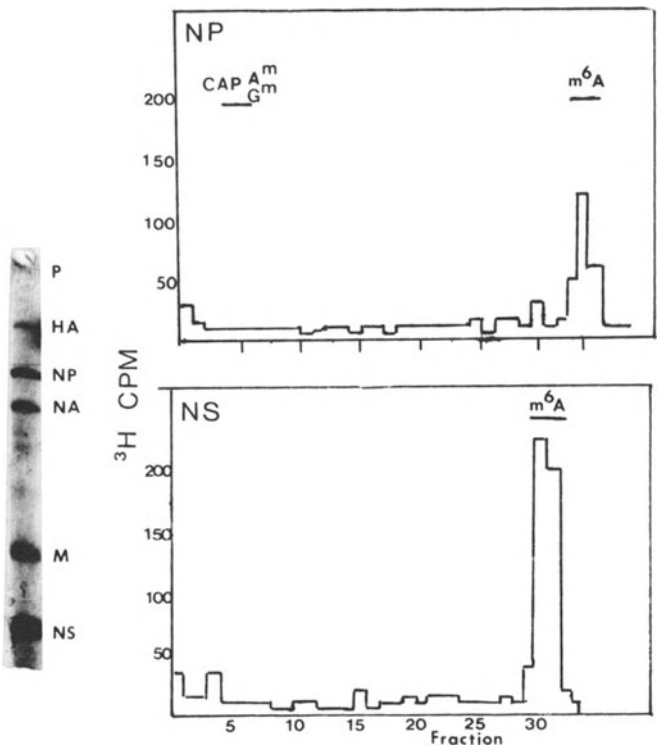

FIGURE 5. Determination of Which Viral mRNAs Contain m^6A. (A) Infected cells were labeled with ^3H-methyl methionine as previously described (25,26), and the poly(A)$^+$ RNA was annealed to an excess of vRNA. The hybrids were digested with T2 RNase and electrophoresed on a 5% nondenaturing acrylamide gel. After treatment with Autofluor, the gel was exposed to X-ray film. With longer exposure of the X-ray film, the ^3H-label in the P bands are more apparent. (B). ^3H-methyl labeled poly(A)$^+$ RNA from infected cells was mixed with ^{32}P-labeled viral mRNA synthesized in vitro, and the mixture was processed as described in (A). The positions of the double-strands in the 5% gel were localized by virtue of their ^{32}P content by exposure to X-ray film. The bands were eluted, digested with nuclease P1 and calf intestine alkaline phosphatase and analyzed by thin layer chromatography on CEL 300 CM plates in isopropanol: H_2O:NH_4OH - 7:2:1. Individual strips were cut out of the chromatogram, eluted in H_2O and counted in aqueous scintillation liquid. One representative sample from an unspliced mRNA (NP) and from a mRNA that is subsequently spliced (NS equivalent to NS1) is shown. From (39) with permission.

Figure 5 (39), essentially all the ^{3}H-radioactivity in both viral mRNAs in the viral-coded regions was in m^{6}A (39). In fact, all of the viral mRNAs were analyzed this way and all were found to contain m^{6}A. To determine the sequences in which these m^{6}A residues were found, the vRNA-mRNA hybrids were eluted from the gel, and after heat-denaturation were hydrolyzed with pancreatic RNase or a mixture of pancreatic and T1 RNases (39). Analysis by paper electrophoresis showed that pancreatic RNase yielded ^{3}H-label in AAC and the mixture of pancreatic and T1 RNases yielded AC and AAC. Based on the known specificities of these two RNases, these results indicated that the m^{6}A residues are in AAC and GAC (39), the same sequences that m^{6}As are found in mRNAs (both cellular and viral) synthesized by polymerase II (40-42). Consequently, the same nuclear enzyme that methylates internal A residues of cellular mRNAs most likely also methylates the A residues in the virus-coded region of the influenza viral mRNAs.

These results certainly indicate that any role that m^{6}A residues may play in splicing must be in conjunction with other sequences, e.g., appropriate splice junctions. Recent experiments in other systems using inhibitors of methylation (cycloleucine and S-tubercidinyhomocystein) have indicated that polymerase II transcripts with diminished amounts of m^{6}A are inefficiently transported to the nucleus (29,30). These results were interpreted as suggesting a role for internal methylation in processing and/or transport. Because we have found internal m^{6}A residues in influenza viral mRNAs that do not undergo any processing steps, it is more likely that m^{6}As are involved in transport rather than in processing. Finally, it should be pointed out that the role of the m^{6}A residue apparently present in the host-donated region of influenza viral mRNAs is also not known (25).

Synthesis of the Templates for Virion RNA Replication In Vitro. Because the synthesis of both the viral mRNAs and the templates for virion RNA replication occurs in the nucleus of infected cells (16), we determined whether infected cell nuclei are active in the synthesis of these two types of transcripts in vitro. Preparations of infected cell nuclei which exhibited the highest activity for viral RNA transcription in vitro were obtained by avoiding the use of

detergents (33). The nuclear pellet obtained after Dounce homogenization of the cells in hypotonic buffer was washed with a large volume of this buffer to remove trapped cytoplasmic components. Phase contrast microscopy indicated that the nuclei were free of essentially all debris, presumably cytoplasmic in origin, but that some of the nuclei contained small fragments of cytoplasmic material attached to the nuclear membrane. This low level of cytoplasmic contamination was tolerated to preserve optimal transcription activity.

Because of the requirement for ongoing protein synthesis for the synthesis of template RNAs in vivo (31,32), the in vitro assays were carried out under conditions in which concomitant viral protein synthesis occurred (33). Under these conditions, our nuclear preparations synthesized a small amount of viral proteins, presumably becuase of the presence of residual cytoplasm. To assay for the synthesis of template RNAs, the labeled RNA products were separated into poly(A)$^+$ and poly(A)$^-$ fractions on oligo(dT) cellulose. To distinguish between viral mRNAs lacking poly(A) and template RNAs, the poly(A)$^-$ RNA was annealed to an M13 single-stranded DNA specific for transcripts copied off the NS vRNA. After digestion with RNase T2 (which will remove the poly A and the 5' primer-donated region from the mRNA), the specific hybrids were bound to nitrocellulose filters. The NS in vitro transcripts were eluted by heat-denaturation and analyzed by gel electrophoresis. As shown in Figure 6, this procedure allows the analysis of NS1 mRNA and NS template RNA free from a background of incomplete transcripts of larger vRNA segments (33). The poly(A)$^+$ RNA synthesized by nuclei isolated at 3 hours after infection contains a predominant band corresponding to the NS1 mRNA (lane 1), and the poly(A)$^-$ RNA synthesized by these nuclei contains an additional predominant band, larger in size, corresponding to the NS template RNA (lane 2). A small amount of template RNA-synthesizing activity was detected in the nuclei as early as at 1 hour after infection. This nuclear activity reached its maximum at 3-4 hours, and then declined. Only very low activity was detected in the cytoplasmic fraction (lane 7).

To determine whether concomitant protein synthesis was actually required for the synthesis of template RNAs in vitro, we tested the

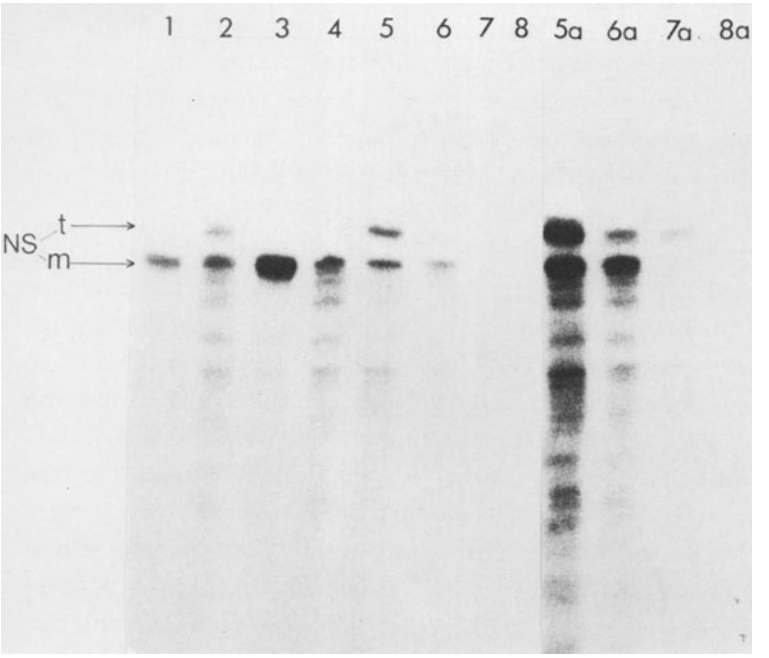

FIGURE 6. In Vitro Synthesis of Template RNAs by Infected Cell Nuclei and the Ability of Infected Cell Cytoplasmic Extracts and its High Speed Supernatant to Confer Template RNA-Synthesizing Activity on Nuclei from Anisomycin-Treated Cells. BHK-21 cells at 3 hr after infection were fractionated into nuclei and cytoplasm (14,23). The nuclei (10 µ l, corresponding to approximately 10^6 nuclei) were incubated for 60 min at 31C in a final volume of 100µl containing the four ribonucleoside triphosphates (UTP labeled with ^{32}P), 5mM $Mg(OAc)_2$, creatine phosphate and creatine phosphokinase, dithiothreitol, and the 20 amino acids. The RNA products were separated into poly(A)$^+$ (lane 1) and poly(A)$^-$ (lane 2) fractions by oligo (dT) cellulose chromatography. Both fractions were annealed to a vRNA-sense M13 DNA clone of the NS segment (prepared by Geoffrey Shapiro of this laboratory) and were processed as described in the text. Infected cells treated with 100µM anisomycin at 2.5 hr post infection were collected at 3 hr, and the isolated nuclei were incubated in the absence (lanes 3 and 4: poly(A)$^+$ and poly(A)$^-$, respectively) or the presence of 35 µl of infected cell cytoplasmic extract (lane 5, poly(A)$^-$ only) or 35 µl of its high speed supernatant (1.5 hr at 65,000 rpm) (lane 6, poly(A)$^-$ only) obtained from cells at 3 hr after infection. Lanes 7 and 8 show the poly(A)$^-$ NS RNAs synthesized in reactions containing the infected cell cytoplasmic extract and its high supernatant, respectively, in the absence of nuclei. Lanes 5a, 6a, 7a and 8a are longer exposures of lanes 5-8. t designates NS template RNA; and m designates NS1 mRNA. From (33) with permission.

effect of inhibitors of protein synthesis on template RNA synthesis. Neither cycloheximide nor anisomycin inhibited either viral mRNA or template RNA synthesis catalyzed by the nuclear preparations (33). This lack of dependence on ongoing protein synthesis was also observed using unfractionated infected cells (permeabilized with lysolecithin) to catalyze template RNA synthesis.

These results suggest that influenza virus-infected cells contain a pool of protein(s) that is capable of supporting the synthesis of template RNAs in vitro and that a sufficient amount of these proteins remains associated with the nuclear preparations. To remove these proteins, we treated the infected cells with anisomycin at 2.5 hours prior to the isolation of the nuclei at 3 hours (33). Nuclei isolated from these cells do not catalyze template RNA synthesis in vitro (Figure 6, lane 4), whereas they retain their activity in viral mRNA synthesis (lane 3). Thus, the protein(s) needed for template RNA synthesis can be depleted from the nuclei by inhibiting protein synthesis in the infected cells prior to isolation of the nuclei.

We determined whether we could restore the activity for template RNA synthesis to the nuclei from anisomycin-treated cells (33). We added to these nuclei either the cytoplasmic extract from infected cells (3 hours after infection) or the fraction prepared by high-speed centrifugation of this extract. The cytoplasmic extract effectively restored the activity for template RNA synthesis (Figure 6, compare lanes 4 and 5). The amount of template RNA-synthesizing activity in this reconstituted system was much higher than that seen in the cytoplasmic extract alone (compare lanes 5 and 5a to lanes 7 and 7a). In addition, the high-speed supernatant fraction from the cytoplasmic extract, which contains no detectable template-synthesizing activity by itself (lanes 8 and 8a), restores template RNA-synthesizing activity to the nuclei (lanes 6 and 6a). Anisomycin had no effect on template RNA synthesis in this reconstituted system, and cytoplasmic extracts from uninfected cells did not restore activity.

Consequently, we have shown that infected cell nuclei exhibit a relatively high activity for template RNA synthesis in vitro (33). In fact, some nuclear preparations synthesized more template RNA than

viral mRNA. In contrast, template RNA synthesis in the infected cell is only about 5% of viral mRNA synthesis (7,16). In vitro and in vivo template RNA synthesis also differ in their need for concomitant protein synthesis. Protein synthesis inhibitors block template RNA synthesis in vivo (31,32), but not in vitro catalyzed either by nuclei or lysolecithin-permeabilized whole cells (33), as summarized here. Thus, the infected cells contain a pool of protein(s) sufficient for template RNA synthesis in vitro, whereas these protein(s) are apparently rate-limiting for this synthesis in vivo. This probably explains why template RNA synthesis (relative to viral mRNA synthesis) is higher in vitro than in vivo.

The existence of a pool of soluble protein(s) needed for template RNA synthesis in infected cell extracts enabled us to establish a reconstituted system that catalyzes template RNA synthesis in vitro (33). This protein pool can be depleted by treating infected cells with a protein synthesis inhibitor prior to the isolation of the nuclei, and as a consequence the viral transcriptase complexes in the nuclei are rendered inactive in template RNA synthesis in vitro. However, they regain this synthetic capacity when supplemented with the cytoplasmic extract from infected cells or its high-speed supernatant fraction (33). These results indicate that the protein(s) and/or other components in the cytoplasmic extract of infected cells enable the viral transcriptase complex in the nucleus to antiterminate i.e., to continue copying the vRNA template past the site at which termination occurs during viral mRNA synthesis (17 to 22 nucleotides from the 5' ends of the vRNAs). As discussed previously, the available evidence suggests that during viral mRNA synthesis "stuttering" or reiterative copying of the uridine tracts at this position on the vRNAs occurs, resulting in poly A synthesis and the termination of transcription of the ensuing 5' terminal vRNA sequences (9). Factors present in the cytoplasmic extract of infected cells clearly inhibit this "stuttering" and termination step. A similar type of antitermination may also occur during the switch from mRNA to template RNA synthesis with other negative-strand RNA viruses, for example, with vesicular stomatitis virus (VSV), where an antitermination mechanism may account for the switch from the synthesis of multiple mRNAs to a single full-length

template RNA (43). In fact, a reconstituted system capable of antitermination *in vitro* has recently been described for VSV (44-46). Template RNA synthesis with influenza virus, but not with VSV, has an added complication in that it requires a change from the primer RNA-dependent initiation occurring during viral mRNA synthesis to a primer-independent initiation. The cytoplasmic factors from influenza virus-infected cells also probably confer the capability for primer-independent initiation on the viral transcriptase complexes in the nuclei. As verification, however, it will be necessary to establish that the template RNAs synthesized *in vitro* contain labeled di- or tri-phosphorylated 5' termini. These experiments will be feasible when the efficiency of template RNA synthesis in the reconstituted system is increased 10-fold or more.

As described previously in this report, during capped RNA-primed viral mRNA synthesis the three P proteins as a complex move down from the 3' ends of the vRNA templates until the site of poly(A) addition and chain termination is reached (15). The inhibition of this "stuttering" and termination step by the cytoplasmic factors from infected cells most likely results from modification of one or more of the three P proteins and/or by the addition of one or more virus-coded proteins to the P protein complex. Similarly, it is likely that the P protein complex would have to be altered to allow initiation of chains without a primer. Another important question to be addressed is whether the P protein complexes associated with the nucleocapsids in the nuclei of infected cells are identical with the P protein complexes associated with the virion nucleocapsids. We have found that the addition of infected cell cytoplasmic fractions to virion nucleocapsids results in template RNA-synthesizing activity in some, but not all, experiments, suggesting that with virion nucleocapsids additional factors are required for the switch to template RNA synthesis. It should be possible to resolve this issue and to determine the mechanism of the switch from viral mRNA synthesis to template RNA synthesis - the first step in vRNA replication - by the identification and purification of the cytoplasmic factors that confer template RNA-synthesizing activity to nuclear and virion nucleocapsids.

Translational Controls in Influenza Virus-Infected Cells. The

nuclear degradation of newly synthesized polymerase II transcripts in infected cells does not explain the shutoff of host cell protein synthesis. Initial analysis of the steady-state levels of cytoplasmic ALV and β-actin mRNAs in CEFs indicate that these two mRNAs do not decrease in amount until after three hours post infection. Our results indicate that the shutoff of host (and ALV) protein synthesis occurs prior to three hours. These results suggest that the shutoff is mediated by a viral translational system that selectively translates influenza viral and not host mRNAs.

The existence of this control system was established by experiments in which adenovirus-infected cells at late times post infection were superinfected with influenza virus. At late times of infection adenovirus effectively blocks the expression of host cell mRNAs. The transport of newly synthesized host cell mRNAs from the nucleus to the cytoplasm is greatly reduced, or essentially eliminated (47,49), and the translation of preexisting host cell mRNAs is drastically inhibited (47,48,50,51). It has been found that VAI RNA, an adenovirus-encoded polymerase III product (52), is required for the translation of adenovirus major late mRNAs, and that in the absence of VAI RNA (using an adenovirus mutant, dl331, which does not synthesize VAI RNA) neither host nor adenovirus late mRNAs are translated (53). These results indicate that adenovirus infection alters the translational machinery so that neither host nor adenovirus mRNAs can be translated and the VAI RNA then imparts to the translational machinery the capability for selectively translating late adenovirus mRNAs. To determine whether influenza viral mRNA overcomes these blocks, we superinfected HeLa cells with influenza virus at early (4 hours) and late (16 hours) times after adenovirus 2 infection (36). Four hours later the cells were labeled with ^{35}S-methionine for 1 hour, and the labeled proteins were analyzed by gel electrophoresis (Figure 7). As compared to HeLa cells infected with only influenza virus (lane 1), the levels of synthesis of the most readily detectable influenza virus-specific proteins, NP (nucleocapsid protein), M1 (membrane protein) and NS1 (nonstructural protein 1), were not significantly reduced at early (lane 3) and at late (lane 5) times of adenovirus infection. At late times of adenovirus infection, little or no host cell proteins are

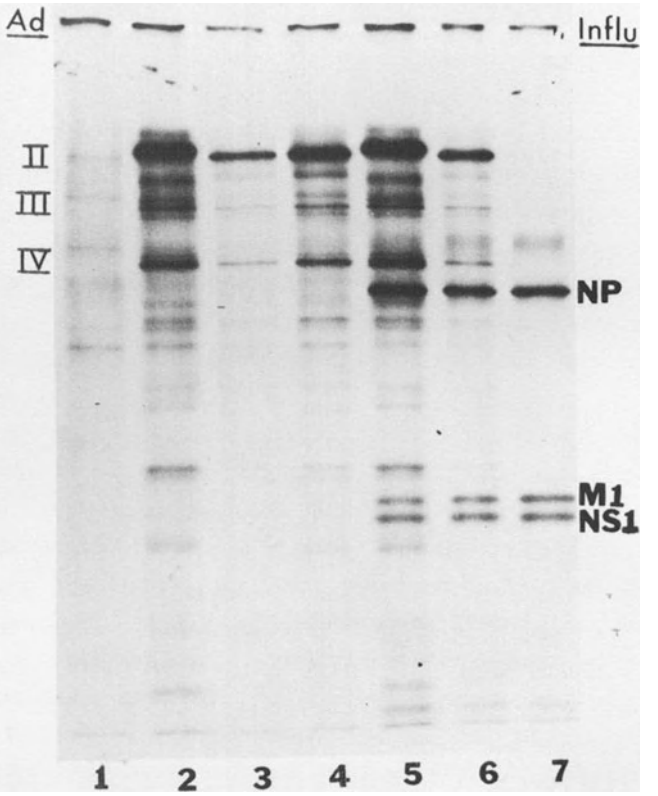

FIGURE 7. Influenza Virus-Specific Protein Synthesis in Adenovirus-Infected Cells. Suspension HeLa cells were infected with adenovirus 2 (50 PFU/cell) for 4 hr (lanes 2 and 3) or for 16 hr (lanes 4 and 5), and superinfected with influenza virus (100 PFU/cell) (lanes 3 and 5) or mock-superinfected (lanes 2 and 4). The cells were incubated an additional 4 hr and then labeled with ^{35}S-methionine for 1 hr. As a control, cells were infected with influenza virus alone for 4 hr, and then labeled for 1 hr (lane 1). The labeled proteins were analyzed by electrophoresis on a 14% acrylamide gel. Equivalent amounts of each sample were applied to the gel. The positions of representative influenza virus- and adenovirus-specific proteins are shown on the left and right, respectively. From (36) with permission.

---

synthesized (compare lanes 2 and 4). Thus, even at this late time after adenovirus infection, influenza viral mRNAs synthesized in the nucleus are efficiently transported to the cytoplasm and translated (36). Immunofluorescence studies performed with anti-hexon antiserum

demonstrated that all the HeLa cells examined were infected with adenovirus, indicating that influenza virus-specific protein synthesis is occurring in adenovirus-infected cells. In addition, though influenza virus inhibits host cell protein synthesis (e.g., compare lanes 1 and 3 to lane 2), the synthesis of late adenovirus proteins, such as the hexon (polypeptide II), the nonstructural 100K, penton base (III) and fiber (IV), remained at high levels after superinfection with influenza virus (compare lanes 4 and 5). The same pattern of synthesis of influenza virus and adenovirus proteins was seen at 5 and 6 hours after influenza virus superinfection.

As a consequence of capped RNA-primed initiation of transcription, the influenza virus mRNAs synthesized in late adenovirus 2-infected cells might be expected to contain the 5' ends of the major late adenovirus transcripts. The presence of these adenovirus 5' ends might explain the ability of influenza viral mRNAs to be transported from the nucleus and to be translated. Consequently, we sequenced the 5' ends of a representative influenza viral mRNA - the NS1 mRNA - synthesized in adenovirus-infected cells (36). An 18 base-pair restriction fragment obtained from cloned NS1 DNA (labeled at the 5' end of the strand complementary to viral mRNA) (5) was hybridized to the cytoplasmic poly(A)$^+$ RNA from doubly-infected HeLa cells or from HeLa cells infected with influenza virus alone (36). The 5' end(s) of the NS1 mRNAs were copied using reverse transcriptase, and the extension products were separated from the primer by gel electrophoresis (Figure 8A). The extension products from the cells infected with influenza virus alone formed a broad band in the range of 60 to 63 nucleotides in length (lane 2). In contrast, the extension products from the doubly-infected cells contained two discrete bands, 63 and 60 nucleotides in length, against a background (lane 1).

Sequence analysis of the extended products from the cells infected with influenza virus alone confirmed that the NS1 mRNAs in these cells contained heterogeneous 5' ends (Figure 8D) (36). This reflects the usage of a heterogeneous population of capped RNA primers, with a preference for CA-terminated fragments (GT in the DNA copy), as shown previously (5,54). In contrast, sequence analysis of the two discrete bands in the extended products from the

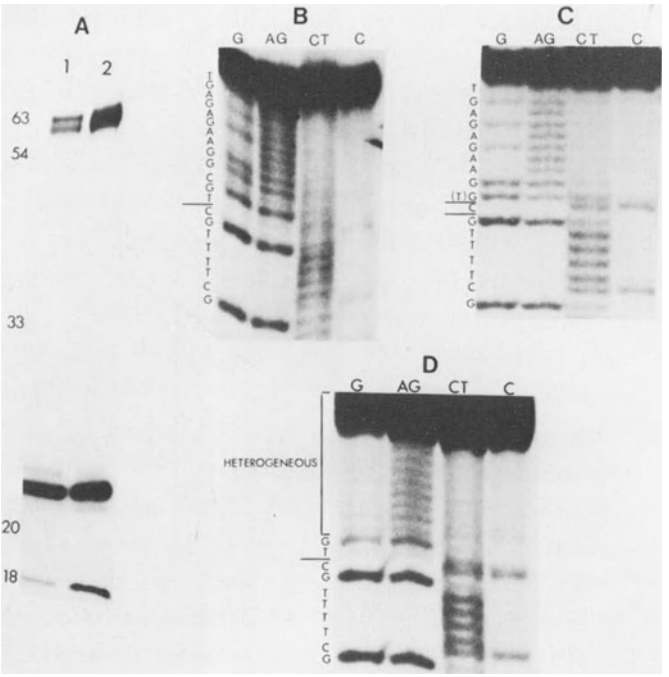

FIGURE 8. Sequence Analysis of the 5' Terminal Sequences of Influenza Viral NS1 mRNA in the Cytoplasm. Panel A: Poly(A)$^+$ cytoplasmic RNA (50 μg) obtained from HeLa cells infected with influenza virus alone for 4 hr (lane 2) and from HeLa cells infected with adenovirus 2 for 16 hr followed by superinfection with influenza virus for 4 hr (lane 1) were each hybridized to a 5' end - labeled Hind II-Bam HI 18 base-pair fragment obtained from the NS-10 DNA clone (5). Primer extension was carried out using reverse transcriptase, and the extended products (60-63 nucleotides in length) were separated from the primer on a 20% denaturing polyacrylamide gel containing 8M urea. The labeled band of 22 nucleotides represents primer which was filled in by the reverse transcriptase to form blunt ends (as established by sequence analysis). The position of marker DNA bands 20, 33 and 54 nucleotides in length are also shown on the left. Panel B: DNA sequencing gel of the 63 nucleotide-long extension product copied off the NS1 mRNA synthesized in doubly-infected cells (lane 1 of Panel A). Panel C: DNA sequencing gel of the 60 nucleotide-long extension product copied off the NS1 mRNA synthesized in doubly-infected cells (lane 1 of Panel A). Panel D: DNA sequencing gel of the extension products copied off the NS1 mRNA synthesized in cells infected by influenza virus alone (lane 2 of Panel A). The horizontal line on the left of panels B-D marks the boundary between adenovirus or cell-derived sequences and influenza virus-encoded sequences. From (36) with permission.

doubly-infected cells indicated that some of the NS1 mRNA molecules in these cells contained adenovirus 5' ends (Figures 8B and 8C) (36). On the basis of the sequence of the discrete band of 63 nucleotides, it is evident that the first 13 nucleotides from the 5' end of the major late adenovirus transcript, (m^7Gppp)ACUCUCUUCCGCA (55,56), are joined to the NS1 mRNA, with G as the first influenza virus-encoded nucleotide (36). From the sequence of the band of 60 nucleotides, it can be concluded that the first 10 or 11 nucleotides from the 5' end of the major late adenovirus transcript are linked to the NS1 mRNA. Some of the NS1 mRNA molecules from the doubly-infected cells do not contain 5' ends from the major late adenovirus transcript. This is apparent from the sequence analysis of the 60 nucleotide-long extension product (Figure 8C), where the tenth 5' terminal nucleotide is evidently a mixture of A and C (T and G in the DNA copy). This indicates the presence on some NS1 viral mRNAs of CA-terminated primer fragments, 10 nucleotides in length, which are not derived from the adenovirus major late transcript. In addition, these non-adenovirus ends can be identified in the reverse transcriptase-catalyzed extension products which migrate between the 63 and 60 nucleotide-long discrete bands (Figure 8A, lane 1). The sequence in these DNAs corresponding to the 5' ends of the NS1 mRNAs is heterogeneous. From these results, however, it is not possible to obtain a precise estimate of the percentage of the NS1 mRNAs in the cytoplasm that contain adenovirus 5' ends.

Consequently, we can postulate two ways for the influenza viral mRNAs to overcome the adenovirus-imposed blocks against mRNA expression. In one, the influenza viral mRNAs, because they contain adenovirus sequences at their 5' ends, would mimic adenovirus mRNAs and would utilize the adenovirus-specified systems for nuclear transport and translation. Alternatively, influenza virus would establish its own systems for transport and translation that are independent of adenovirus. The first postulate would predict that the NS1 mRNAs containing adenovirus 5' ends are selectively transported from the nucleus to the cytoplasm. To determine whether such selective transport occurred, we compared the 5' ends of nuclear and cytoplasmic NS1 mRNAs, using primer extension and sequence analysis (36). We found that the amount of the extension products

and hence the amount of NS1 mRNA in the nucleus was about 10-20% of that in the cytoplasm. A similar fraction of the nuclear and cytoplasmic NS1 mRNAs were shown to contain adenovirus 5' ends, indicating that the transport of NS1 mRNAs from the nucleus is independent of the presence of adenovirus 5' ends. These results indicate that: (i) the system for the transport of influenza viral mRNAs from the nucleus is distinct from that for the nuclear transport of cellular mRNA sequences; and (ii) adenovirus infection does not affect the influenza virus-specific transport system.

If the presence of adenovirus 5' ends explains the ability of influenza viral mRNAs to be translated in late adenovirus-infected cells, then the influenza viral mRNAs associated with polyribosomes would be expected to contain almost exclusively adenovirus 5' ends, or at least, to be enriched in these ends in comparison to the total cytoplasmic influenza viral mRNAs. To examine this possibility, polyribosomes were isolated from late adenovirus-infected cells at 4 hours after superinfection with influenza virus (36). Figure 9A shows the sedimentation profile of the polyribosomes. The polyribosomes were found in the pellet (designated sample number 1) and in the gradient region (sample 2) which sedimented faster than the monosomes (80S region; sample 3). When the cells were labeled with ^{3}H-uridine for 90 minutes prior to collection of the cells, the resulting radioactivity profile was essentially coincident with the A_{260} profile shown in Figure 9A, except that an extra peak was observed in the 30 to 60S region (sample 4), presumably representing messenger ribonucleoproteins and/or influenza viral nucleocapsids. Northern blot analysis was carried out on cell-equivalent amounts of poly(A)$^{+}$ RNA from samples 1 through 5 using as probe either ^{32}P-labeled adenovirus DNA or a mixture of ^{32}P-labeled influenza virus PB2, NP and NS cloned DNAs. The results showed that the majority of the adenovirus late mRNAs and of the influenza viral NP and NS1 mRNAs were associated with polyribosomes (Figures 9B and 9C) (36). Some of these two influenza viral mRNAs, as well as the majority of the PB2 mRNAs, were associated with the monosome region (Figure 9C, sample 3). A portion of all three influenza viral mRNAs were found in the 30 to 60S region (sample 4), where messenger ribonucleoproteins and/or viral nucleocapsids would be expected to be

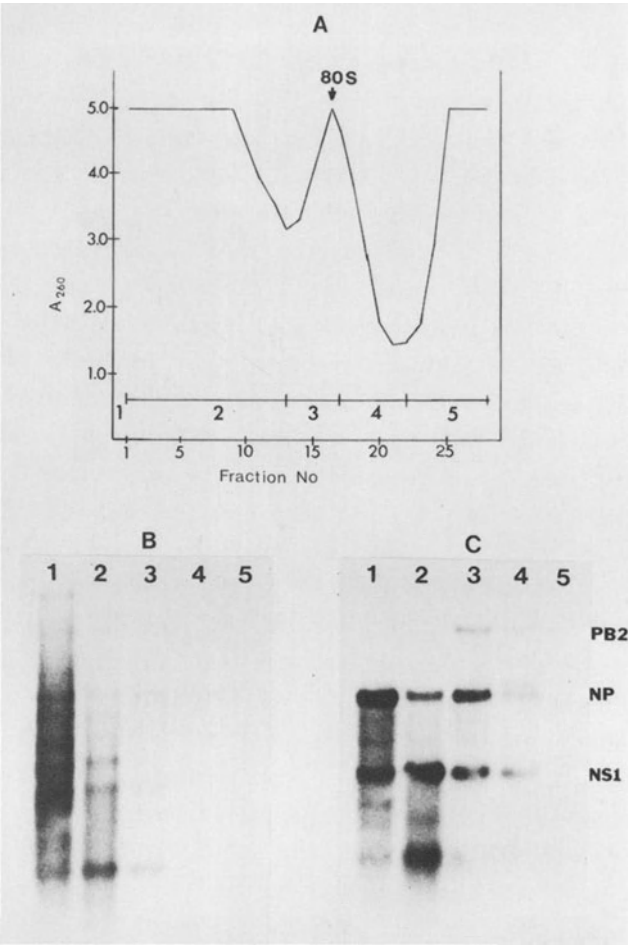

FIGURE 9. Identification of Polyribosomes Containing Adenovirus and Influenza Virus mRNAs from Doubly-Infected Cells. Panel A: Sedimentation profile of polyribosomes from doubly-infected cells. A cytoplasmic extract from $3x10^8$ HeLa cells infected with both adenovirus and influenza virus was layered on a 10-50% sucrose gradient in 50mM KCl, 5mM $Mg(OAc)_2$, 20mM Tris-HCl, pH 7.5, and centrifuged in a SW28 rotor for 4 hr at 27,000 rpm. After dilution with an equal volume of TE buffer (10mM Tris-HCl, pH 7.5; 1mM EDTA), the absorbancy at 260 nm (A_{260}) of the fractions from the gradient was determined. The maximum A_{260} that could be measured without further dilution of the fractions was 5.0, so that the A_{260} of fractions 1-8 and 25-27 is actually 5.0 or greater. The monosome peak sediments at 80S. The pellet was taken up in TE buffer and designated sample number 1. The gradient fractions were pooled as

found. Little or no viral mRNAs were detected at the top of the gradient (sample 5). To examine the 5' ends of the NS1 mRNAs associated with the different samples from the polyribosome gradient, primer extension and sequence analysis was carried out (36). The results indicated that the NS1 mRNAs with adenovirus 5' ends are not preferentially associated with the polyribosomes.

From the analysis of the polyribosome-associated NS1 mRNA, it was likely that influenza viral mRNA translation is independent of the positive regulation apparently exerted on the translation of late adenovirus mRNAs by VAI RNA. To obtain direct evidence for this conclusion, we employed dl331 virus, an adenovirus 5 deletion mutant defective in the synthesis of VAI RNA (55). Adenovirus 5-transformed 293 cells were infected with either wild-type adenovirus 2 or the mutant dl331 virus for 16 hours and then superinfected with influenza virus for 4 hours (36). As controls, cells were infected with either influenza virus or adenovirus alone. The cells were labeled for 30 minutes with ^{35}S-methionine, and the labeled proteins were analyzed by gel electrophoresis (Figure 10). In the absence of influenza virus superinfection (lanes 2-4), approximately 8-10-fold less adenovirus late proteins were synthesized at 20 hours after dl331 infection (lane 3) than after wild-type adenovirus 2 (lane 2) or wild-type adenovirus 5 (lane 4) infection, as observed previously by others (48,53). On the other hand, the dl331 virus was at least as efficient as the wild-type adenoviruses in suppressing host cell protein synthesis (36). In fact, the inhibition of host cell protein synthesis is more apparent in dl331-infected cells (lane 3) than in wild-type adenovirus-infected cells (lanes 2 and 4) because of the paucity of comigrating adenovirus-specific proteins. As shown in lanes 5 and 6, influenza virus superinfection of dl331-infected cells

---

shown, yielding samples 2 through 5. The RNA from these 5 samples was extracted, and poly(A)$^+$ RNA was selected by oligo (dT)-cellulose chromatography. Panel B: Northern blot analysis of the poly(A)$^+$ RNAs in samples 1-5 of the polyribosome gradient shown in panel A, using ^{32}P-labeleld adenovirus virus 2 DNA as probe. Panel C: Northern blot analysis of the poly(A)$^+$ RNAs in samples 1-5 of the polyribosome gradient in panel A, using ^{32}P-labeled influenza virus PB2, NP and NS DNA clones as probes. Cell-equivalent amounts of the poly(A)$^+$ RNA in samples 1-5 were used for the analyses shown in panels B and C. From (36) with permission.

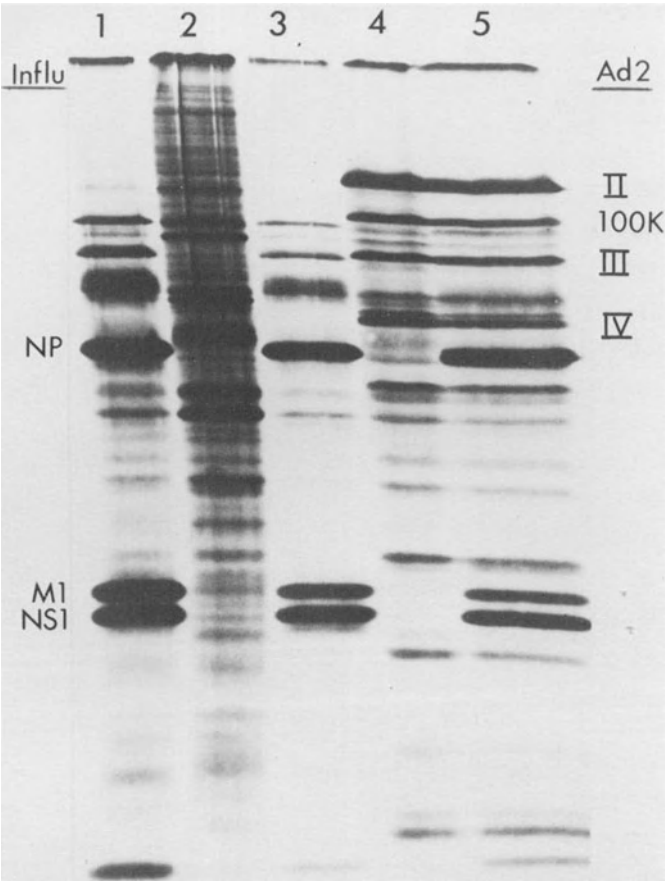

FIGURE 10. The Synthesis of Influenza Virus-Specific Proteins in Cells Infected by Adenovirus dl331, Which Does Not Synthesize VAI RNA. Monolayers of 293 cells were mock-infected (lane 1); or were infected with 10 PFU/ml of adenovirus 2 (lane 2), dl331 (lane 3), or adenovirus 5 (lane 4) for 20 hr; or were infected with 50 PFU/ml of influenza virus for 4 hr (lane 7). Another set of 293 monolayers were infected with adenovirus 2 (lane 5) or dl331 (lane 6) for 16 hr, and were then superinfected with influenza virus for an additional 4 hr. Each set of cells was subsequently labeled with ^{35}S methionine for 30 min, and cell-equivalent amounts of the labeled proteins were analyzed by electrophoresis on a 14% acrylamide gel. The positions of representative adenovirus- and influenza virus-specific proteins are shown on the left and right, respectively. From (36) with permission.

resulted in the synthesis of the same amount of the influenza virus-specific M1 and NS1 proteins as after superinfection of wild-type adenovirus 2-infected cells (36). The relative amounts of the NP protein synthesized in the two sets of cells is difficult to ascertain because the wild-type adenovirus 2-infected cells contain one or more adenovirus-specific proteins that migrate close to the influenza viral NP protein. Also, the amount of the influenza virus NP, M1 and NS1 proteins synthesized in dl331-infected cells (lane 6) is essentially the same as that synthesized in cells not infected with adenovirus (lane 7). Thus, we can conclude that the translation of influenza viral mRNAs is VAI RNA-independent in late adenovirus-infected cells.

One of the defects in translation in dl331-infected cells has recently been identified: the initiation of translation is reduced apparently because the 43S preinitiation complex (the 40S ribosome containing a ternary complex consisting of met-tRNA, GTP and initiation factor eIF-2) is unable to interact with mRNA to form a 48S preinitiation complex (57). This block in the initiation of translation must result from some modification of the translational machinery occurring during adenovirus infection. VAI RNA can overcome this block, but only for adenovirus mRNAs (53). The mechanism by which VAI RNA accomplishes this has not be established, but it has been postulated that VAI RNA somehow interacts with both the ribosome and the tripartite leader of late adenovirus mRNAs (58). Clearly, because the translation of the NP, M1 and NS1 influenza viral mRNAs occurs at essentially normal levels in dl331-infected cells (36), influenza virus can overcome the block in translation initiation without VAI RNA. This suggests that an influenza virus-specified product, possibly a small RNA analogous to VAI RNA, enables the 43S preinitiation complex to interact with influenza viral mRNAs, but not with adenovirus or cellular mRNAs, to form a 48S preinitiation complex. It will be of great interest to identify this postulated influenza virus-specified product and to determine the sequences and/or other features common to the various influenza viral mRNAs that is recognized by, and interacts with, this postulated viral-specified product.

In cells infected by wild-type adenovirus, the expression of

both adenovirus and influenza viral mRNAs proceeds efficiently, with apparently minimal effects on each other. Our results suggest that this is accomplished at the level of translation by the establishment of two kinds of polyribosomes: one kind, presumably containing VAI RNA, that specifically translates late adenovirus mRNA; and another kind, presumably containing an influenza virus-specified product, that specifically translates influenza viral mRNAs. In contrast, cellular mRNAs are displaced from large polyribosomes in cells infected by adenovirus alone (48), and this displacement would be expected to be increased in cells which are also infected with influenza virus. Because both the adenovirus mRNAs and the influenza viral mRNAs (particularly the major NP and NS1 mRNAs) are predominantly associated with polyribosomes (as shown by the Northern analyses of Figure 9) (36), a sufficient number of functional ribosomes exist to accomodate the majority of both viral mRNAs.

We can conclude that an influenza virus-specific translational system also mediates the shutoff of host cell translation observed in the absence of adenovirus infection. It is not known whether the establishment of an influenza virus-specific translational system is preceeded by a modification of this system that blocks all translation, analagous to the events occurring in adenovirus-infected cells. Translational control via the selective binding of specific mRNAs to ribosomes occurs in several other virus systems (59-62) and also in cellular mRNA translation (e.g., after heat shock) (63-65), but the mechanism of this selectivity has not been definitively established in any of these systems. Consequently, it is of considerable general interest to elucidate the mechanism of the translational controls operating in influenza virus-infected cells.

REFERENCES

1. Bouloy, M., Plotch, S.J. and Krug, R.M. Proc. Natl. Acad. Sci. USA 75: 4886-4890, 1978.
2. Plotch, S.J., Bouloy, M. and Krug, R.M. Proc. Natl. Acad. Sci. USA 76: 1618-1622, 1979.
3. Krug, R.M., Broni, B.B. and Bouloy, M. Cell 18: 329-334, 1979.
4. Plotch, S.J., Bouloy, M., Ulmanen, I. and Krug, R.M. Cell 23: 847-858, 1981.
5. Beaton, A.R. and Krug, R.M. Nucleic Acids Res. 9: 4423-4436, 1981.
6. Krug, R.M. Current Topics in Microbiol. and Immunol. 93:125-150, 1981.

7. Hay, A.J., Lomniczi, B., Bellamy, A.H. and Skehel, J.J. Virology 83: 337-355, 1977.
8. Hay, A.J., Abraham, G., Skehel, J.J., Smith, J.C. and Fellner, P. Nucleic Acids Res. 4: 4179-4209, 1977.
9. Robertson, J.S., Schubert, M. and Lazzarini, R.A. J. Virol. 38: 157-163, 1981.
10. Inglis, S.C., Carroll, A.R., Lamb, R.A. and Mahy, B.W.J. Virology 74: 489-503, 1976.
11. Ulmanen, I., Broni, B.A. and Krug, R.M. Proc. Natl. Acad. Sci. USA 78: 7355-7359, 1981.
12. Horisberger, M.A. Virology 107: 302-305, 1980.
13. Blaas, D., Patzelt, E. and Keuchler, E. Nucleic Acids Res. 10: 4803-4812, 1982.
14. Ulmanen, I., Broni, B.A. and Krug, R.M. J. Virol. 45: 27-35, 1983.
15. Braam, J., Ulmanen, I. and Krug, R.M. Cell 34: 609-618, 1983.
16. Herz, C., Stavnezer, E., Krug, R.M. and Gurney, T., Jr. Cell 26: 391-400, 1981.
17. Jackson, D.A., Caton, A.J., McCready, S.J. and Cook, P.R. Nature 296: 366-368, 1982.
18. Lamb, R.A., Briedis, D.J., Lai, C.-J. and Choppin, P.W. In: Genetic Variation Among Influenza Viruses. (Ed. Nayak, D.P.) Academic Press, 1981, pp. 141-158.
19. Breidis, D.J. and Lamb, R.A. J. Virol. 42: 186-193, 1982.
20. Lamb, R.A. and Lai, C.-J. Cell 21: 475-485, 1980.
21. Lamb, R.A., Lai, C.-J. and Choppin, P.W. Proc. Natl. Acad. Sci. USA 78: 4170-4174, 1981.
22. Lamb, R.A. and Lai, C.-J. Virology 123: 237-256, 1982.
23. Krug, R.M. In: Genetics of Influenza Viruses (Eds. Palese, P. and Kingsbury, D.W.) Springer-Verlag, Wien, New York, 1983, pp. 70-98.
24. Lamb, R.A. and Choppin, P.W. Ann. Rev. Biochem. 52: 467-506, 1983.
25. Krug, R.M., Morgan, M.M. and Shatkin, A.J. J. Virol. 20: 45-53, 1976.
26. Krug, R.M., Broni, B.A. and Bouloy, M. Cell 18: 329-334, 1979.
27. Chen-Kiang, S., Nevins, J.R. and Darnell, J.E. J. Mol. Biol. 135: 733-752, 1979.
28. Stoltzfus, C.M. and Dane, R.W. J. Virol. 42: 918-931, 1982.
29. Finkel, D. and Groner, Y. Virology 131: 409-425, 1983.
30. Camper, S.A., Albers, R.J., Coward, J.K. and Rottman, T.M. Mol. and Cell. Biol. 4: 538-543, 1984.
31. Hay, A.J., Skehel, J.J. and McCauley, J. Virology 116: 517-522, 1982.
32. Barrett, T., Wolstenholme, A.J. and Mahy, B.W.J. Virology 98: 211-225, 1979.
33. Beaton, A.R. and Krug, R.M. Proc. Natl. Acad. Sci. USA, 1984, in press.
34. Skehel, J.J. Virology 49: 23-36, 1972.
35. Lazarowitz, S.G., Compans, R.W. and Choppin, P.W. Virology 46: 830-843, 1971.
36. Katze, M.G., Chen, Y.-T. and Krug, R.M. Cell, 1984, in press.
37. Robertson, J.S. Nucleic Acids Res. 6: 3745-3757, 1979.
38. Romanos, M.A. and Hay, A.J. Virology 132: 110-117, 1984.

39. Beaton, A.R., Ph.D., Dissertation (Sloan-Kettering Division, Graduate School of Medical Sciences, Cornell University), 1984.
40. Wolstenholme, A.J., Barrett, T., Nichol, S.T. and Mahy, B.W.J. J. Virol. 35: 1-7, 1980.
41. Dimock, K. and Stoltzfus, C.M. Biochemistry 16: 471-478, 1977.
42. Canaani, D., Kahana, C., Lavi, S. and Groner, Y. Nucleic Acids Res. 6: 2879-2899, 1979.
43. Banerjee, A.K., Abraham, G. and Colonno, R.J. J. Gen. Virol. 34: 1-8, 1977.
44. Davis, N.L. and Wertz, G.W. J. Virol. 41: 821-832, 1982.
45. Patton, J.T., Davis, N.L. and Wertz, G.W. J. Virol. 49: 303-309, 1984.
46. Peluso, R.W. and Moyer, S.A. Proc. Natl. Acad. Sci. USA 80: 3198-3202, 1983.
47. Beltz, G.A. and Flint, S.J. J. Mol. Biol. 131: 353-373, 1979.
48. Babich, A., Feldman, L., Nevins, J., Darnell, J. and Weinberger, C. Mol. Cell Biol. 3: 1212-1221, 1983.
49. Flint, S.J., Beltz, G.A. and Linzer, D.I.H. J. Mol. Biol. 167: 335-359, 1983.
50. Anderson, C.W., Baum, P.R. and Gesteland, R.F. J. Virol. 12: 241-252, 1973.
51. Bello, J. and Ginsberg, H.S. J. Virol. 1: 843-850, 1976.
52. Reich, P.R., Rose, J., Forget, B. and Weissman, S.M. J. Mol. Biol. 17: 428-439, 1976.
53. Thimmappaya, B., Weinberger, C., Schneider, R. and Shenk, T. Cell 31: 543-551, 1982.
54. Lamb, R.A., Lai, C.-J. and Choppin, P.W. Proc. Natl. Acad. Sci. USA 78: 4170-4174, 1981.
55. Akusjärvi, G. and Pettersson, U. J. Mol. Biol. 134: 143-158, 1979.
56. Zain, S., Sambrook, J., Roberts, R., Keller, W., Fried, M. and Dunn, A.R. Cell 16: 851-861, 1979.
57. Schneider, R.J., Weinberger, C and Shenk, T. Cell 37: 291-298, 1984.
58. Logan, J. and Shenk, T. Proc. Natl. Acad. Sci. USA, 1984, in press.
59. Bhat, R.A. and Thimmappaya, B. Proc. Natl. Acad. Sci. USA 80: 4789-4793, 1983.
60. Dunigan, D. and Lucas-Lenard, J. J. Virol. 45: 618-626, 1983.
61. Trachsel, J., Sonenberg, N., Shatkin, A.J., Rose, J., Leong, K., Bergmann, J.E., Gordon, J. and Baltimore, D. Proc. Natl. Acad. Sci. USA 77: 770-774, 1980.
62. Ehrenfeld, E. Cell 28: 435-436, 1982.
63. Ashburner, M. and Bonner, J.J. Cell 17: 241-254, 1979.
64. Scott, M.P. and Pardue, M.L. Proc. Natl. Acad. Sci. USA 78: 3353-3357, 1981.
65. Beinz, M. and Gurdon, J.B. Cell 29: 811-819, 1982.

12

TRANSCRIPTION BY LA CROSSE VIRUS

Jean L. PATTERSON and Daniel KOLAKOFSKY

INTRODUCTION

The genome of La Crosse (LAC) virus, a member of the California encephalitis serogroup of the insect-transmitted Bunayviridae (1), consists of 3 segments of single-stranded RNA of negative polarity (-) each contained within a separate nucleocapsid labelled small (S), medium (M), and large (L). Genetic and molecular studies have led to the following gene assignments: the S segment codes for the N protein, the M segment codes for the two surface glycoproteins, and the L segment, by elimination, codes for the L protein which is located internally and thought to be part of the viral polymerase (2,3,4). In addition to the structural proteins, bunyavirus-infected cells also contain at least two nonstructural (NS) proteins, NS_s and NS_m, coded for by the S and M genome segments respectively (5). The complete nucleotide sequences of the S genome of LAC and snowshoe hare (SSH) viruses have recently been determined (6,7). In both viruses, the S segment has been shown to contain two overlapping open reading frames (ORF).

Recently, the use of synthetic oligonucleotide primers, end-labelled

Y. Becker (ed.), VIRAL MESSENGER RNA.

DNA and genome RNA probes have revealed that transcription at the beginning of the 3′ end of the LAC S (-) genome segment is more complicated than expected (8). Both primer extension and S1 nuclease mapping studies have detected the 5′ end of the major message-sized transcript near the precise 3′ end of the (-) genome template, and at least 7 other minor transcripts, also putative mRNAs, which start internally within a 75 nucleotide-long stretch near this end. Further, using 3′ end-labelled, S- genome RNA as a probe, three leader RNAs complementary to the exact 3′ end and extending for 75, 95, 115 nucleotides were also demonstrated. The presence of these leader RNAs suggested that the putative mRNAs which start internally might be the result of reinitiation of the viral polymerase after termination of leader RNA synthesis.

We describe here the mapping of the 3′ end of the major S mRNA. We also demonstrate that the major LAC S genome mRNA starts on a host-derived primer. Due to this unexpected finding that LAC transcription resembles that of influenza virus, we further characterized the in vitro transcription mechanism of LAC and found stimulation of transcription with oligonucleotides such as (A)nG, cap analogues, and natural mRNAs such as alfalfa mosaic virus (ALMV) 4 RNA. In the case of (A)nG and ALMV 4 RNAs, stimulation occurs by acting as primers for viral transcription while the cap analogues appear to stimulate via an alternate mechanism. We also report that LAC virions contain a cap- dependent endonuclease activity similar to that reported for influenza virus.

MATERIALS AND METHODS

Preparation of intracellular RNA: BHK cells were infected and harvested (4) as modified by Lindsey-Regnary (9). Cytoplasmic extracts of uninfected and LAC- infected BHK-21 cells were prepared and fractionated into CsCl pellet and banded material as previously described (8). The CsCl pellet RNA was adjusted to a concentration of 200 OD/ml.

S1 analysis of LAC S clone DNA and LAC CsCl pellet mRNA:

The 520 bp Ava II restriction fragment corresponding to nucleotide 600-982 of the LAC S genome fused to a poly C tract and to nucleotide 3616-3505 of pBR322 (see Fig. 3) was isolated from 10 µg of LAC S plasmid 4C-26 (6). This fragment labelled with α-^{32}P-CTP and the Klenow fragment of DNA polymerase 1 was either left double-stranded or strand separated as described in the figure legends. The labelled (-) strand or the double-stranded fragments were eluted from the preparative gel, recovered by ethanol precipitation, and mixed with varying concentrations of uninfected or LAC- infected CsCl pellet RNA (as described in the figure legends) and ethanol precipitated. For the single-stranded probe, the recovered nucleic acids were raised to a volume of 20 µl of 40 mM PIPES, pH 6.4, 2 mM EDTA, 0.4 M NaCl and hybridized at 55°C overnight as described in text. The double-stranded fragments were raised to a volume of 20 µl of 40 mM PIPES, pH 6.4, 1 mM EDTA, 0.4 M NaCl, 80% formamide and hybridized at 50°C. After hybridization, 0.2 ml of ice- cold S1 nuclease buffer (0.28 M NaCl, 0.05 M sodium acetate pH 4.6, 4.5 mM ZnCl, 2 mM EDTA and 10 µg/ml tRNA) containing 500 units/ml of S1 nuclease was added and the reactions were incubated for 45 min at 37°. The reactions were then phenol extracted and the remaining nucleic acids were recovered by ethanol precipitation and electrophoresed on 7 M urea, 8% polyacrylamide gel (10).

Oligodeoxynucleotide primer extension (ODPE) analysis

Oligodeoxynucleotide primers which correspond to nucleotides 38 to 50 and 136 to 147 from the 3′ end of the S (-) genome were gifts of J. Obijeski (Genetech, Inc.). The 5′ end of the primers were labelled with gamma ^{32}P-ATP and T4 polynucleotide kinase. The RNAs to be tested were mixed with 25 ng of labelled primer and ethanol precipitated. The reaction was then resuspended in 7 µl of H_2O and heat denatured at 90° for two minutes. The reactions were then placed at 43° and the reaction conditions were changed to 50 mM Tris (pH 8.3), 8 mM MgCl, 0.08 M NaCl and 16 units of avian myoblastosis virus (AMV) reverse transcriptase in a total volume of 20 µl and incubated for 90 min. The reactions were then

phenol extracted and ethanol precipitated and analyzed on an 8 or 12% polyacrylamide gel (10).

Preparation of purified LAC virions

BHK cells were infected with LAC virus at a multiplicity of 0.01 pfu/cell and the infection was carried out for 48 hours at 37°. The cellular supernatant was then harvested, cleared by low speed centrifugation (10,000 x g for 20 min) and the virus was then precipitated by the addition of 2.3 gm of NaCl and 7 gm of polyethylene glycol 6000/ml, stirred for 4 hours at 0° and collected by centrifuging for 20 min at 10,000 x g. The pellet was then resuspended in 10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 1 mM EDTA (TNE) and the virus banded in a glycerol-tartrate gradient (4) by centifugation for 4 hours at 35,000 rpm in SW 41 rotor. The visible virus band was then removed, diluted with TNE, pelleted through a 30% glycerol cushion and resuspended in 10 mM Tris-HCl, pH 7.4, 1 mM EDTA at a concentration of 0.5 to 1 mg/ml. The virus was stored at -20° and appeared stable to freezing and thawing.

LAC virion in vitro polymerase and endonuclease reactions

The in vitro reaction conditions were essentially those of Plotch and Krug (11) and contained 50 mM HEPES, pH 8.1, 100 mM KCl, 6 mM $MgCl_2$, 1 mM DTT, 0.2% NP40, 1mM ATP, GTP and CTP, 50 μM UTP containing 5 μC of ^{32}P-α-UTP, 40 μg/ml of rat liver ribonuclease inhibitor and purified LAC virions at 100 to 200 μg /ml. The reaction was incubated at 30° and incorporation was measured by TCA precipitation onto Whatman G/FC filters which were washed with 5% TCA containing 50 mM $Na_2P_2O_7$. For non-radioactive reactions the UTP was raised to 1 mM and the reaction products of 100 μl reactions were recovered by centrifugation through a 20 to 40% CsCl gradient after disruption of the virions with four volumes of 0.5M NaCl, 1% NP40 and 10 mM EDTA. The pelleted RNA was redissolved, recovered by ethanol precipitation and dissolved in 50 μl of 1 mM EDTA, 10 mM Tris-HCl, pH 7.4 (ET). For the endonuclease reactions LAC virions

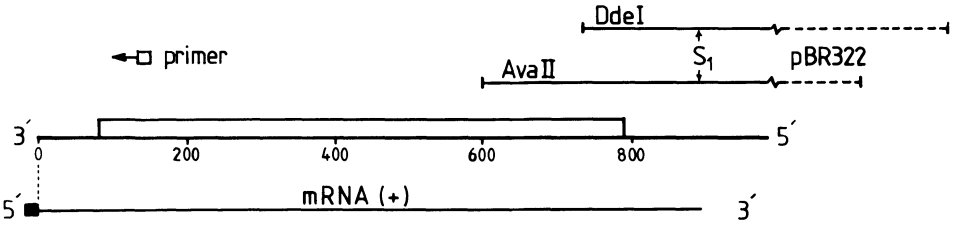

Fig. 1. Organization of the major LAC S genome transcript.
The left side of the top line shows the position of the synthetic oligonucleotide used for primer extension, and the right shows the position of the restriction fragments used to map the 3′ end of the S mRNA with S1. The middle line shows the S(-) genome with its 3′ end at position 0: the bar above the line denotes the N protein-coding sequence. The bottom line shows the structure of the major LAC S transcript deduced from these studies. The box at the left denotes the presumed host primer at the 5′ end of the mRNA.

were incubated with ^{32}P-cap-labelled ALMV 4 RNA in the same reaction conditions as the polymerase reaction but without triphosphates and analyzed on 12.5% polyacrylamide gels.

RESULTS

Location of the 3′ end of the LAC S mRNA

Annealing studies using radiolabelled virion (-) genome RNA and total intracellular viral RNA have demonstrated both full-length and almost full-length transcripts of the S genome (9,12). The full-length transcript was found entirely in the nucleocapsid structures which band in CsCl density gradients, whereas the almost full-length transcript was found in the pellet of the gradient (9). The LAC S genome segment is 982 nucleotides long. To map the 3′ end of the S mRNA, 3′- end labelled DNA restriction fragments which began at position 600 and 736 in the viral DNA sequences and extended past the end of the S genome into the pBR322 vector sequences were used to examine the mRNA (CsCl pellet) from both uninfected and virus- infected cells (see Fig. 1). The results of these experiments using both ds and ss DNA starting at position 600 as probes show that a single 286 bp band is obtained only when infected cell, CsCl pellet RNA was used (Fig. 2 lanes 4,5,and 9). The probe starting at position 736 confirmed this result, placing the end of the S mRNA 150 nucleotides away from the start of this probe (data not shown). These results therefore place the 3′ end of the S mRNA at approximately position 886 on the S genome segment.

A noteworthy feature of the sequence at position 886 is the polyuridine tract 3′(-) GUUUUU (886-892) which is similar to other (-) RNA virus termination-polyadenylation signals which also contain a U run 5 to 7 nucleotides long (13,14,15). It is therefore curious that LAC S mRNA is apparently not polyadenlylated (16,17). It should be noted however, that the absence of a poly(A) tail on the S mRNA has not yet been examined directly, therefore the inability of bunyavirus mRNA to bind to oligo dT columns may be due to the shortness of the poly(A) tail rather than its complete absence. On overexposure of similar experiments there

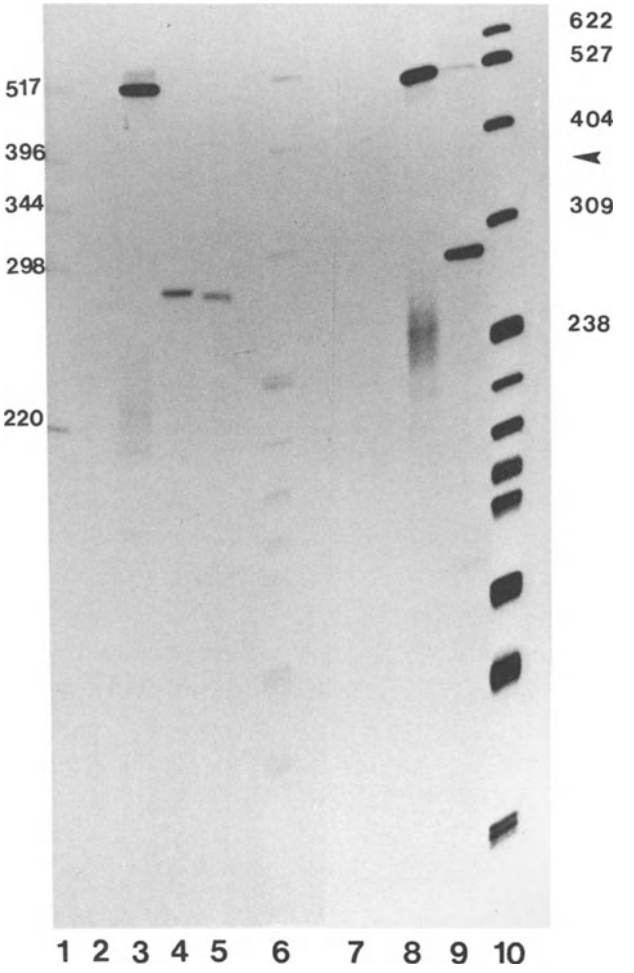

Fig. 2. S1 nuclease mapping of the 3′ end of the LAC S mRNA.

The 520 bp Ava II restriction fragment (see Fig. 1) was annealed to 2 or 1 optical density units of LAC-infected, CsCl-pellet RNA (lanes 4 and 5, respectively) and 2 optical density units of uninfected CsCl pellet RNA (lane 2) overnight in 80% formamide at 50°C and digested with S1 nuclease as described in the text. Lane 3 shows the undigested restriction fragment. The same Ava II restriction fragment was also strand separated, and the (-) strand was annealed to 2 optical density units of LAC-infected, CsCl-pellet RNA (lane 9) or 2 optical density units of uninfected CsCl-pellet (lane 7) at 55°C overnight without formamide. Lane 8 shows the undigested single-stranded DNA fragment. Lane 12 shows Hinf 1 digestion of pBR322, and lanes 6 and 10 are Msp 1 digestion of pBR322. The arrow on the right refers to the region where a full-length transcript would be expected.

appears to be a back up stop at position 925 marking a GUUUUUU sequence on the (-) genome. Little if any full-length anti-genome RNA appears in the pellet of the CsCl gradient as shown by the lack of S1- detectable RNA at position 982 (arrow on right hand margin of Fig. 2).

Examination of the precise 5′ end of the major S mRNA

The curious and unexpected finding that on-going protein synthesis in bunyavirus-infected cells is apparently required for transcription (16,17) as well as genome replication led us to reexamine the 5′ ends of the LAC mRNA more precisely. As mentioned above we have examined the 5′ ends of the S genome mRNAs by both S1 mapping and primer extension. Both techniques detected the major transcript at or near the precise 3′ end of the S genome template. The primer extension experiment was also carried out in the presence of ddNTPs so that the nucleotide sequence could be read directly from the gel. These experiments (cf. Fig. 4b, ref. 8) precisely mapped the 5′ ends of the minor transcripts at position 74 to 123 with reference to the cloned S genome, but the precise 5′ end of the major transcript could not be determined because the sequence within the first 30 to 40 nucleotides of the major transcript was too weak to read. Furthermore, relative to restriction fragment markers, the major transcript appeared to be slightly longer (156-160 bp) than predicted from the S genome clone (147 bp).

We have therefore repeated this experiment using the same 5′ end labelled primer (position 147 to 136) except that more radioactivity was used and the sequence gel was electrophoresed longer to better display this region (Fig.3). Note that the extended primer does not stop at position 1 (the precise 3′ end of the (-) genome template) but continues beyond the end of the (-) template and stops heterogeneously at position -10 to -14 (just above 154 bp, Hinf 1 marker). The sequence between position +1 and -10 is uninterpretable; the A and G lanes contain a band at every position, wherease the C and T lanes are too weak to read. Such a result is consistent with a heterogeneous sequence at position +1 to -10. These results suggest that the first 10 to 14 nucleotides at the 5′ end of this mRNA cannot be coded for by the virus genome, and that they

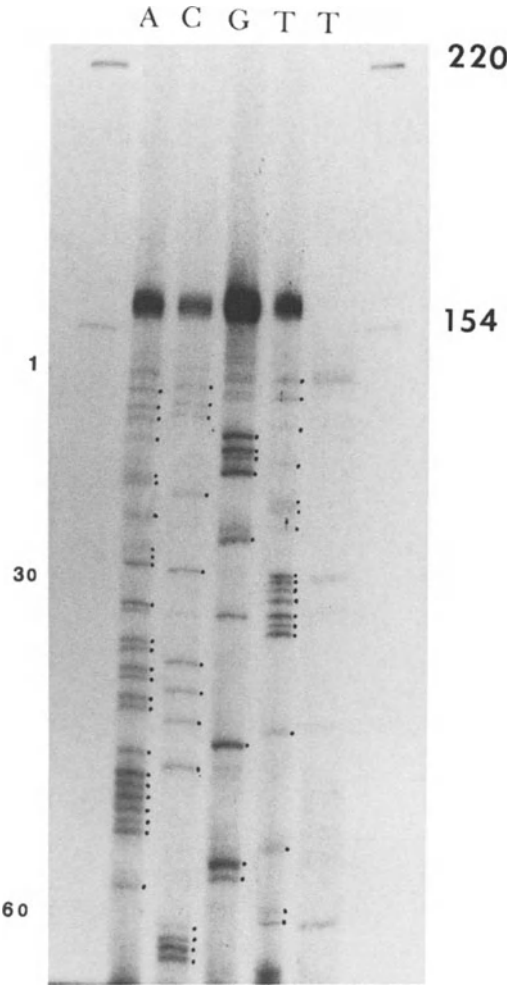

Fig. 3. Dideoxy sequence of the 5′ end of the major S mRNA.
The 5′ end-labelled duodecamer (see text) was primer extended on uninfected (right hand T lane) or LAC-infected, CsCl-pellet RNA (four left hand lanes) as described in the presence of individual dideoxynucleotide triphosphate. Lanes surrounding the sequence are Hinf 1 digestion of pBR322. The numbers on the right-hand margin refer to the lengths of the restriction fragment markers. The sequence lanes are marked with their appropriate dideoxynucleotide added, and the dots to the right of the ladder bands show the sequence expected at the 5′ end of the major S mRNA. The numbers on the left-hand margin refer to the nucleotide position from the 3′ end of the S (-) genome.

are heterogeneous both in sequence and length similar to the 5′ ends of influenza virus mRNAs (18-20). It therefore appears that the major LAC S mRNA initiates on a 10 to 14 nucleotide-long primer, presumably of host origin, and continues to position 886 on the (-) genome (Fig. 1).

Effect of oligonucleotides and cap analogues on LAC virion polymerase activity

Because of the obvious similarities between the structures at the 5′ ends of bunyavirus and influenza virus mRNAs mentioned above, and the large body of information aready availible for how influenza virus initiates its mRNA synthesis, we have chosen influenza virus as a model in examining LAC virions for polymerase activity. McGeoch and Kitron (21) and Plotch and Krug (11) have demonstrated that guanosine nucleosides, nucleotides and oligonucleotides stimulate the influenza virion polymerase reaction. The dinucleotide ApG can stimulate the influenza polymerase reaction as much as 100-fold and has been shown to act as a primer presumably by base pairing with the 3′ end terminal OH-UpC of the influenza (-) genome segments (11,21). Since the LAC genome segments contain a similar 3′ terminal sequence, purified LAC virions were incubated as described in Materials and Methods, plus and minus the oligonucleotides shown in Table 1 and TCA precipitable incorporation was determined. LAC virions alone did incorporate a low level of ^{32}P-UTP into acid-precipitable material. The dinucleotide ApG (but not GpA or ApA, not shown) and the trinucleotide ApApG stimulated the polymerase activity 1.4 to 4.1-fold and 1.2 to 13.2-fold respectively (Table 1). The extent of stimulation varied from one experiment to another for unknown reasons, but the stimulation by these short oligonucleotides appeared greatest when the endogenous polymerase activity of the virions was the lowest. Penn and Mahy (22) have demonstrated that cap analogues such as ^{m}GpppAm or ^{m}GpppA$_{m}$ also stimulate the influenza virion polymerase reaction. However, unlike dinucleotides such as ApG, the cap analogues are not incorporated into the transcripts but presumably act through an allosteric mechanism since the cap analogue and ApG stimulations are additive. Table 1 also demonstrates that these cap analogues similarly

TABLE I Stimulation of LaC Virion Polymerase

	Additions	cpm x 10^{-3} / 5 µl	Fold stimulation	
Experiment 1	none	3.0	-	
	AG	4.2	1.4	
	AAG	7.1	2.4	
	$^{m}GpppA^{m}_{m}$	13.6	4.5	
	AG + $^{m}GpppA^{m}_{m}$	19.6	6.5	(5.9)
	AAG + $^{m}GpppA^{m}_{m}$	22.2	7.4	(6.9)
Experiment 2	none	6.5	-	
	AAG	8.0	1.2	
	$^{m}GpppA^{m}_{m}$	16.7	2.6	
	AAG + $^{m}GpppA^{m}_{m}$	22.1	3.4	(3.8)
	ALMV4	46.6	7.2	
	ALMV4 + $^{m}GpppA^{m}_{m}$	31.2	4.8	(9.8)
Experiment 3	none	5.0	-	
	+ $^{m}GpppA_{m}$	18.8	3.8	
	+ $^{m}GpppA^{m}_{m}$	21.5	4.3	
	+ ALMV4	64.5	12.9	
Experiment 4	none	2.1	-	
	+ AAG	9.2	4.5	
	none	1.3	-	
	+ AAG	5.2	4.0	
Experiment 5	none	.97	-	
	AG	3.98	4.1	
	AAG	12.78	13.2	

stimulate the LAC virion polymerase reaction from 2.6 to 4.5 fold, and when added together with ApG or ApApG, the stimulations are similarly additive.

By far the strongest stimulators of the influenza polymerase *in vitro* are natural mRNAs which contain at their 5′ end the cap 1 structure, since these RNAs stimulate at molarities hundreds of times lower than oligonucleotides such as ApG (23,24). Krug and co-workers have shown that this stimulation is due to a virus-coded endonuclease contained within the the influenza polymerase complex which generates an oligonucleotide primer containing the 5′ capped end of the mRNA and the first 12 to 14 nucleotides, which is then incorporated into the influenza transcripts (25,26). LAC virions were therefore incubated with ALMV 4 RNA which contains the partially methylated cap structure ^{m}GpppG at its 5′ end (27). In two experiments 1.4 μM ALMV 4 RNA stimulated the LAC polymerase 7.2 and 12.9 times. Interestingly, when ALMV 4 RNA was added together with a cap analogue, these stimulations were antagonistic rather than additive (Table 1).

Purified LAC virions thus contain an endogenous RNA polymerase which is stimulated by the same oligo and polynucleotides as the influenza polymerase *in vitro*. To examine the products of the LAC virion polymerase, samples of the unstimulated and ApG-stimulated (2.5-fold) reaction products were isolated after phenol extraction together with the virion (-) genome RNAs. After self-annealing, ssRNAs were digested with RNAse A in high salt and the remaining dsRNAs were examined by PAGE (Fig. 4). A smear rather than discrete bands, from 100 to 1000 bp in length were seen in both the unstimulated and ApG-stimulated reactions. The proportion of transcript from each of the three LAC genome segments was not determined.

ApG stimulated S genome polymerase products begin at the precise 3′ end of the S (-) genome template

To determine where the ApG-stimulated reaction products had initiated RNA synthesis, the position of the 5′ end of the *in vitro* transcripts relative to the 3′ end of the S(-) genome template was

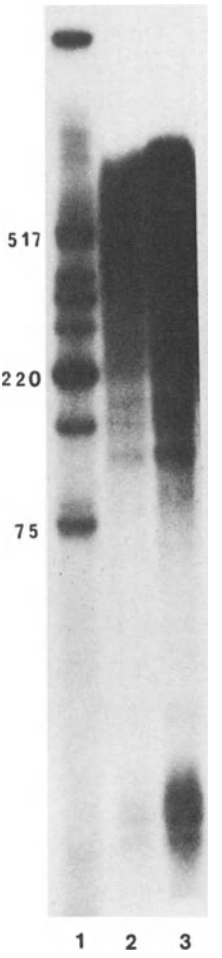

Fig. 4. Examination of the chain length of the LAC in vitro transcripts. The total RNAs from both an unstimulated (lane 2) and ApG-stimulated (lane 3) 4 hr polymerase reactions were isolated by phenol extraction following incubation of the reactions with 500 µg/ml of proteinase K in 0.2% SDS and 10 mM EDTA for 5 min at 60°C. The ethanol precipitated RNAs were then digested with 10 µg/ml RNase A in 2.5 x buffer A (375 mM NaCl, 25 mM Tris, pH 7.4, 2.5 mM EDTA) for 15 min at 25°C followed by proteinase K digestion as above for a further 15 min. The remaining dsRNAs were recovered by ethanol precipitation and electrophoresed on an (40:1 acrylamide:bis-acrylamide) non-denaturing gel along with an end-labelled Hinf 1 digestion of pBR322 as markers.

measured. The LAC polymerase products were first isolated free of the virion (-) genomes by CsCl density gradient centrifugation under conditions where the (-) genome RNAs which are contained in helical nucleocapsids band at a buoyant density of 1.3 (8,9) but the unencapsidated reaction products pellet through the gradient. To determine the position of the 5′ end of the transcript made in vitro, a synthetic oligodeoxynucleotide corresponding to nucleotides 38 to 50 of the (-) S genome RNA was used as a primer and extended on the in vitro (+) polymerase products using reverse transcriptase (see Fig. 5). To avoid subsequent ambiguities, this reaction will be referred to as ODPE (oligodeoxynucleotide primer extension). An example of such an experiment is shown in Fig. 6. Lanes 2 and 3 of Fig. 6a show the results of the ODPE analysis on CsCl pellet RNA from uninfected and LAC infected BHK cells respectively as a control. Note that whereas no primer extension occurred on uninfected RNA, on LAC-infected CsCl pellet RNA the primer was extended both to a single band approximately 50 bases long relative to the restriction fragment markers, as well as a group of 4 bands 12-15 bases longer. The band approximately 50 bases long was determined to represent extension to the precise 3′ end of (-) genome (position 1) since in a parallel experiment the primer was also extended on (+) antigenome RNA isolated from intracellular nucleocapsids. This extension yielded a single band which co-migrated with the 50 base band (Fig. 3b, lanes 1 and 2). Since CsCl gradients appear to cleanly separate genome and antigenome nucleocapsids from other virus-specific RNAs, the band at position 1 in lane 3, Fig.6a probably represents leader RNAs or other (+) transcripts rather than antigenome RNA. The bands at -12 to -15 represent mRNAs with non-templated 5′ extensions previously described.

The deoxyoligonucleotide primer was also extended on the reaction products of both a mock virion polymerase reaction, i.e., a reaction to which no rNTPs were added (lane 1) and an ApG-stimulated polymerase reaction (lane 4, Fig. 5a). Interestingly, ODPE analysis of the mock polymerase products yielded a pattern identical to that of LAC-infected CsCl pellet RNA, whereas primer extension on the ApG-stimuulated products yielded a greatly enhanced band at position 1 and bands around position 1 without any stimulation of the bands at -12 to -15. The finding of identical primer extensions on both mock polymerase products and bona

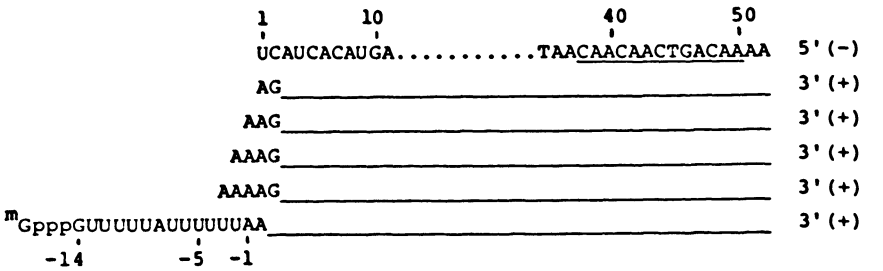

Fig. 5. A schematic representation of the LAC in vitro transcripts.
The top line shows the sequence at the 3′ end of the S (-) genome with the precise 3′ end at the left and numbered at position 1. The synthetic deoxyoligonucleotide used for ODPE analysis (position 38 to 50) is underlined. The middle four lanes show the 5′ ends of the the (A)nG-stimulated reaction products (on the left) as deduced from this work. The bottom line shows the 5′ end of the ALMV RNA 4 stimulated transcript assuming that the 5′ end of the ALMV RNA 4 was transferred to the in vitro product. The minus numbers denote extension beyond the 3′ end of the S (-) genome.

fide intracellular mRNAs (LAC-infected, CsCl-pellet RNA) is not as surprising as might appear at first glance. LAC virus does not produce high yields in tissue culture and the virus used in the mock polymerase reaction represent the harvest from 12 plates of cells, wherease the LAC-infected, CsCl-pellet RNA used was the yield from 4/10 of a single plate. Since the method used to determine the 5′ end of the transcripts is indirect and cannot differentiate between de novo transcripts and pre-existing ones, it seems likeley that the transcripts present in the mock poylmerase reaction are due to contaminating intracellular RNAs, as other pleomorphic (-) RNA viruses such as Sendai virus also contain considerable amounts of cellular contaminants (28). By subtracting the pattern in lane 1 from that of lane 4, we conclude that the major de novo transcripts of the ApG-stimulated reaction began at position 1 or the precise 3′ end of the S genome, with minor transcripts on either side of position 1. This can be seen more clearly in Fig 6b, lane 2, which is a lesser exposure of the ODPE analysis of the ApG-stimulated <u>in vitro</u> reaction.

<u>(A)nG and ALMV 4 RNA act as primers for LAC transcription in vitro</u>

The question of whether in the (A)nG-stimulated reactions the oligonucleotides were acting as primers, was also examined by ODPE analysis. If indeed the ApG stimulated reactions began transcription at the precise 3′ end of the S genome template with the 5′ ApG base pairing with the 3′ terminal HO-UpC and acting as a primer, then analysis of the AAG-stimulated reaction products as described above should lead to transcripts whose 5′ ends started at position -1, the AAAG-stimulated reactions at position -2 etc (see Fig. 5). LAC polymerase reactions were therefore carried out as described in Fig. 7a with either an additional 20 mM EDTA to inhibit de novo RNA synthesis (lane 1), no additions (lane 2), ApG (lane 3), $(A)_2G$ (lane 4), $(A)_{3-5}$(lane 5), $(A)_{3-9}G$ (lane 6) and the reaction products were isolated as above and analyzed by ODPE. The pattern of primer extensions on intracellular CsCl pellet RNA in a parallel reaction is shown in lane 7. Note that in lane 7 the same pattern of bands at -12 to -15 was seen as before but a group of doublet

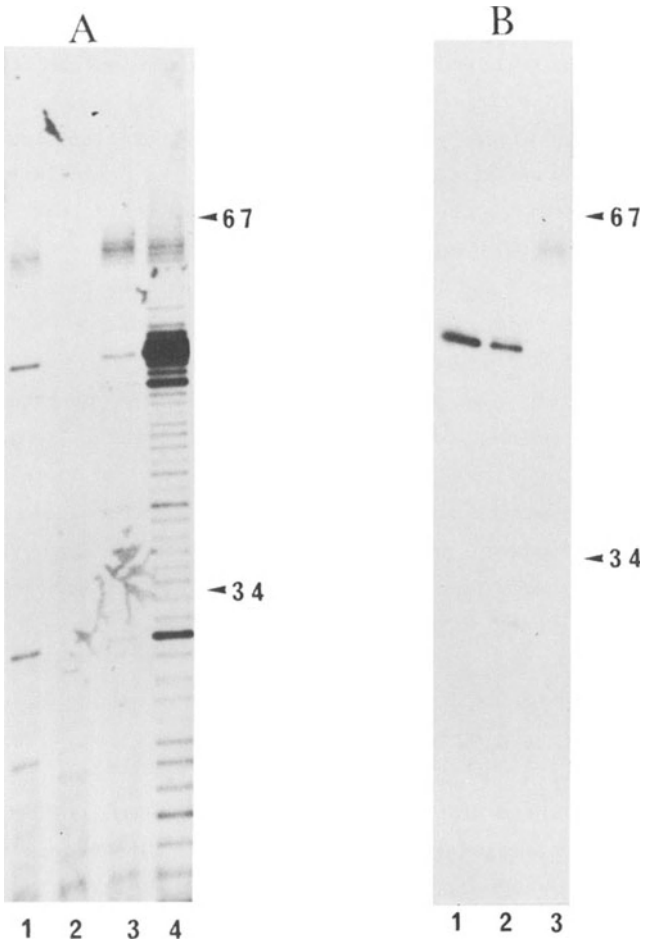

Fig. 6. ODPE analysis of the unstimulated and ApG- stimulated polymerase products.
Panel A The products of non-radioactive 100 µl reactions were pelleted through CsCl density gradients and used for primer extension of end-labelled deoxyoligonucleotides as described in Materials and Methods. Lanes 2 and 3 show the result of ODPE analysis of 1 OD U of uninfected and LAC-infected, intracellular CsCl-pellet RNA respectively. Lanes 1 and 4; ODPE analysis of 40% of the unstimulated and ApG-stimulated reaction products respectively. The numbers at the right refer to the position of the MSP 1 digestion of pBR322 as markers.

Panel B ODPE analysis of RNA extracted from intracellular nucleocapsids banded in CsCl density gradients (lane 1), 20% of ApG-stimulated reaction products (lane 2) and 1 OD U of LAC-infected CsCl-pellet RNA (lane 3). The tetramer band at position -12 to -15 is just visible below the 67 b marker in lane 3.

bands around position +1 was detected rather than one prominant band as in lane 3 of Fig. 5a. The reasons for this variation are not clear, but the band representing the precise 3′ end of the S genome was again determined by ODPE analysis on antigenome RNA and found to be the lower band of the middle doublet (arrow Fig. 7a). When the primer was extended on the reaction products of the unstimulated and (A)nG reactions, the unstimulated and the ApG-stimulated reaction yielded mostly a single band at position +1 (lanes 2 and 3), the AAG-stimulated reaction yielded an additional strong band at position -1 (lane 4), and the $(A)_{3-5}G$ and $(A)_{3-9}G$-stimulated reaction yielded bands at -2 to -3 and -2 to -6 respectively (lanes 5 and 6). Also note that at this exposure, no bands in this region were observed when excess EDTA was added to the reaction before the virus (lane 1). It therefore appears that the unstimulated LAC virion polymerase reaction initiates RNA synthesis at the precise 3′ end of the (-) genome template, presumably with ATP, and the $(A)_nG$-stimulated reaction initiates RNA synthesis by the ApG portion of each oligonucleotide base pairing with the 3′ end of the (-) genome and acting as primers.

We next examined the position of the 5′ ends of the RNAs made in ALMV 4 RNA-stimulated reaction. Krug and co-workers (25,29) have shown that natural mRNAs such as ALMV 4 RNA stimulate influenza virion transcription by first being cut by a virion endonuclease activity to generate a primer containing the capped 5′ end of the RNA and the first 12 to 14 nucleotides, which is then incorporated into the in vitro transcripts. When the 5′ end of the ALMV 4 RNA-stimulated LAC polymerase products were similarly examined by ODPE (lane 1, Fig. 7b), the majority of transcripts were found to begin at -12 to -14 similar to intracellular mRNAs (cf lanes 5 and 6, Fig. 7b) but with somewhat less heterogeneity in this region. The LAC virion polymerase thus appears to be stimulated by natural mRNAs in a manner analogous to the influenza virion polymerase. In the same experiment we examined the effect of adding the cap analogue $^{m}GpppAm$ to the $(A)_2G$ and the $(A)_{3-5}$-stimulated reactions. The results of this experiment are shown in lanes 3 and 2 of Fig. 7b respectively. Note that the presence of the cap analogue does not alter the pattern of the 5′ end of the in vitro transcripts relative to the transcripts made in their absence (cf lanes 4 and 5 of panel a with lanes 3 and 2 of panel

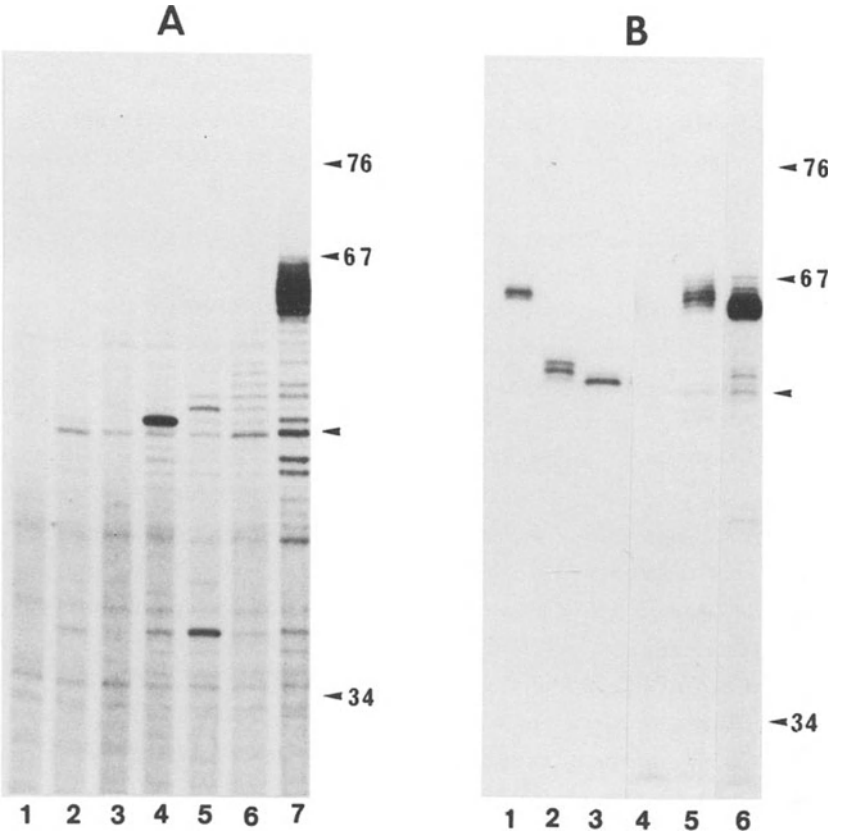

Fig. 7. ODPE analysis of the (A)nG and ALMV RNA 4 stimulated reactions

Panel A The products of 100 µl reaction containing an additional 20 mM EDTA (lane 1) , no addition (lane 2), 0.4mM $(A)_1G$,$(A)_2G$, $(A)_{3-5}G$ and $(A)_{3-9}G$ (lanes 3 to 6 respectively) were recovered through CsCl gradients and 40% of each reaction analyzed by ODPE. Lane 7 shows ODPE analysis of 1 OD U of intracellular LAC CsCl-pellet RNA. The arrow shows position 1 as determined in Fig. 6b (see text).

Panel B ODPE analysis of 40% of the ALMV RNA stimulated reaction products (lane 1), 40% of the $(A)_{3-5}G$ and $(A)_2G$ plus $^{m}GpppA_{m}$ stimulated reaction products (lane 1 and 3, respectively) and 1 OD U of intracellular uninfected and LAC infected, CsCl pellet RNA (lanes 4 and 5 respectively). Lane 6 is a longer exposure of lane 1.

b). Since we have shown in Table 1 that the cap analogue and $(A)_nG$ stimulations are additive, it seems likely that the cap analogues stimulate the LAC polymerase reaction differently than (A)nG, possibly through an allosteric mechanism as suggested for influenza virions.

LAC virions contain a cap-dependent endonuclease

In order to test whether like influenza virus, LAC contains a cap-dependent endonuclease we examined the endonuclease activity of LAC virions by analyzing the cleavage of ^{32}P-cap-labelled ALMV 4 RNA. The original cap group of ALMV 4 RNA was removed by β-elimination and replaced with α-^{32}P-GTP and SAM to give a ^{32}P-labelled cap 1 structure RNA ($^{m}GpppG_{m}$) or without SAM generating a ^{32}P-labelled unmethylated cap (GpppG, cap 0). Reactions were carried out as described in Materials and Methods. Fig. 8b shows the results of the in vitro endonuclease activity with either cap 0 or cap 1 RNA. The cap 0 RNA has no apparent change after incubation with LAC virions as compared to the minus virus control (lanes 1 and 2)), while the cap 1 RNA shows the generation of 2 major bands approximately 13 and 14 bases long (marked with arrows, lane 3) and several minor bands surrounding the major bands. The reaction is relatively fast and is complete after 10 min (Fig. 8b). LAC virus thus contains a methylated cap-dependent endonuclease which cleaves ALMV 4 RNA at specific positions near the 5′ end, exactly as influenza virus. In addition, the cleavage products of the endonuclease reaction can be extended by the virion polymerase. Lanes 5 and 6 of Fig. 7b show the result of adding cold ALMV 4 RNA to a LAC polymerase reaction which contains small amounts of highly radioactive GTP as the only triphosphate present. Under these conditions, one to nine nucleotides appear to have been added to ALMV RNA 4 primer.

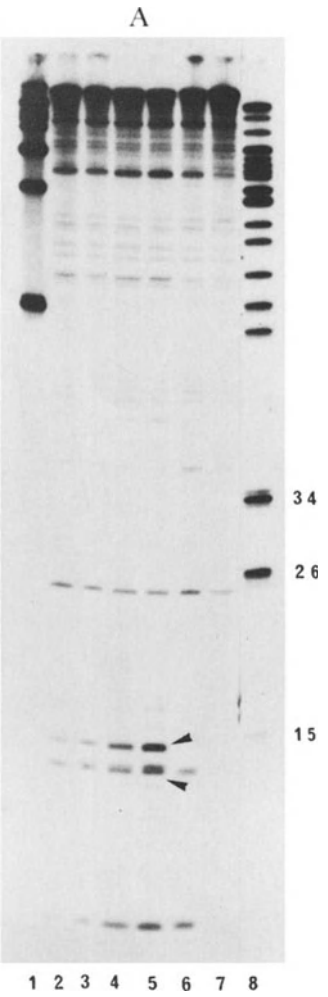

Fig. 8. Characterization of the LAC virion endonuclease activity
Panel A Samples of the methylated cap-labelled ALMV RNA 4 containing 20,000 cpm (approximately 0.5 pMoles) were incubated at 30°C in 20 µl of 50 mM HEPES pH 8.1, 100 mM KCl, 6 mM $MgCl_2$, 1 mM DTT, 0.2% NP40, 40 µg/ml of rat liver ribonuclease inhibitor and 3 µg of purified LAC virions for 0.5, 1, 2, and 5 min (lanes 2 to 5 respectively). As controls, identical reactions were incubated for 20 min with either an additional 40 mM EDTA (lane 6) or without LAC virions (lane 7). The reactions were then phenol extracted, recovered by ethanol precipitation in the presence of 20 µg of carrier tRNA, and electrophoresed on a 7 M urea 12.5% polyacrylamide sequencing gel. Lanes 1 and 8 show Hinf 1 and Msp 1 digest of pBR322 as markers. The numbers on the right refer to the lengths of the smaller restriction fragments markers.

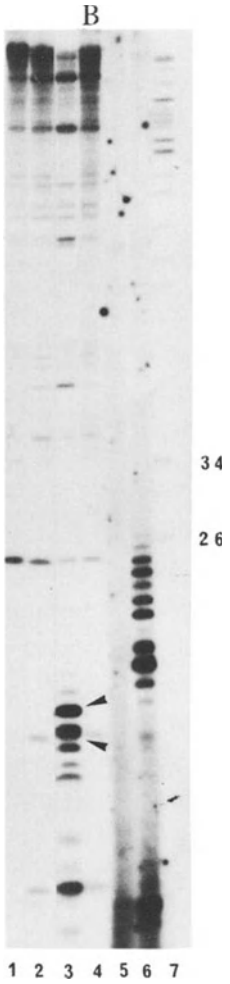

Fig. 8. Characterization of the LAC virion endonuclease activity
Panel B Samples containing 20,000 cpm of either unmethylated (lanes 1 and 2) or methylated cap-labelled ALMV RNA 4 (lanes 3 and 4) were incubated for 20 min as described above either in the presence (lanes 1 and 3) or absence of purified LAC virions (lanes 2 and 4). In a separate experiment, LAC virions were incubated in 25 µl reactions as above with the addition of 10 µM ^{32}P-α-GTP (400 C/mM) either in the presence (lane 6) or absence (lane 5) or 10 µg of unmodified ALMV RNA 4. The reactions were then treated as described above.

DISCUSSION

The major transcript of the LAC S genome segment *in vivo* starts on a 10 to 15 nucleotide long primer, presumably of host origin, and continues to approximately position 886, 100 nucleotides short of the 5′ end of the (-) genome template. *In vitro*, purified LAC virions contain a polymerase which begins the RNA chains at position +1 (the precise 3′ end) and this polymerase activity is stimulated by oligonucleotides such as $(A)_nG$ which act as primers and cap analogues by an alternate mechanism. When a natural mRNA such as ALMV 4 RNA is added to purified LAC virions, this RNA is cleaved at approximately 13 and 14 nucleotides from the cap group, the polymerase activity is greatly stimulated, and the 5′ ends of the *in vitro* transcripts now extend 12 to 14 nucleotides beyond the 3′ end of the (-) genome template. The LAC virion endonuclease activity only cleaves ALMV 4 RNA which contains a methylated capping group. The above evidence suggests that LAC uses a mechanism similar to that of influenza virus to synthesize mRNA, namely, the capture of host cell mRNA which is then used as a source of capped oligonucleotide to prime transcription on the (-) genome template. Thus in contrast to non-segmented (-) RNA viruses such as VSV and Sendai, both families of segmented (-) RNA viruses for which we have information to date appear to use host cell mRNAs as a source of the cap group required for efficient translation of mRNAs in eucaryotic cells, rather than containing their own virus-coded guanylyl and methyl transferases. It is however interesting to note that non-segmented (-) RNA viruses also do not cap transcripts which are initiated at the 3′ end of their polycistronic genome templates but only the mRNAs which are initiated internally, eg Sendai (unpublished observations) and vesicular stomatitis virus (VSV) (30) leader RNAs contain a (p)ppA at their 5′ ends. The absence of virus-coded capping activities among segmented (-) RNA viruses thus appears to reflect the monocistronic nature of their genome segments.

Although influenza and bunyaviruses appear to initiate their mRNA synthesis in a similar manner mechanistically, clear differences between influenza and bunyaviruses transcription remain. For example, bunyavirus transcription is thought to take place in the cytoplasm rather than the nucleus (31), and unlike influenza transcription, it is not sensitive to

drugs such as actinomycin D which inhibit host mRNA synthesis. In contrast, LAC transcription appears to be sensitive to drugs such as cycloheximide and puromycin which inhibit protein synthesis (16,17), a situation without precedent among other (-) RNA viruses. The reasons why LAC transcription requires on-going protein synthesis *in vivo* is indeed intriguing since transcription can take place in purified virions as demonstrated above. Further studies on the nature of the LAC *in vitro* transcripts will be required to hopefully deal with this dilemma.

LITERATURE CITED

1. Bishop, D.H.L., Calisher, C.H., Casals, J., Chumakov, M.T., Gaidamovich, S.Y.A., Hannoun, C., Lvov, D.K., Marshall, I.D., Oker-Blom, N., Pettersson, R.F., Porterfield, J.S. , Russell, P.K., Shope, R.E. and Westaway, E.G. Intervirology 14:125-143, 1980.
2. Gentsch, J.R. and Bishop, D.H.L. J. Virol. 28:417-419, 1978.
3. Gentsch, J.R. and Bishop, D.H.L. J. Virol. 30:767-770, 1979.
4. Obijeski, J.F., Bishop, D.H.L., Murphy, F.A. and Palmer, E.L. J. Virol. 19:985-997, 1976.
5. Fuller, F. and Bishop, D.H.L. J. Virol. 41:643-648, 1982.
6. Cabradilla, C., Holloway, B. and Obijeski, J. Virology 128:463-468, 1983.
7. Akashi, H. and Bishop, D.H.L. J. Virol. 45:1155-1158, 1983.
8. Patterson, J.L., Cabradilla, C., Holloway, B.P., Obijeski, J.F. and Kolakofsky, D. Cell 33:791-799, 1983.
9. Lindsey-Regnary, H. Ph.D. Thesis. Emory University, Atlanta, Georgia, 1983.
10. Sanger, F. and Coulson, A.R. FEBS Letters 87:107-110, 1978.
11. Plotch, S.J. and Krug, R.M. J. Virol. 21:24-35, 1977.
12. Cash, P., Vezza, A.C., Gentsch, J.R. and Bishop, D.H.L. J. Virol. 31:685-694, 1979.
13. Gupta, K.C. and Kingsbury, D.W. Virology 120:518-523, 1982.
14. Robertson, J.W. Nucleic Acids Res. 6:3737-3745,1979.
15. Robertson, J.S., Schubert, M. and Lazzarini, R.A. J. Virol. 38:157-163, 1981.
16. Abraham, G. and Pattnaik, A.K. J. Gen. Virol. 64:1277-1290, 1983.
17. Pattnaik, A.K. and Abraham, G. J. Virol. 1983, in press.
18. Caton, A.J. and Robertson, J.S. Nucleic Acids Res. 12:2591-2603, 1980.
19. Lamb, R.A. and Lai, J. Cell 21:475-485, 1980.
20. Plotch, S.J., Bouloy, M. and Krug, R.M. Proc. Natl. Acad. Sci. USA 76:1618-1622, 1979.
21. McGeoch, D. and Kitron, N. J. Virol. 15:686-695, 1975.
22. Penn, C.R. and Mahy, B.W.J. Virus Res., 1984, in press.
23. Bouloy, M., Plotch, S.J. and Krug, R.M. Proc. Natl. Acad. Sci. USA 75:4886-4890, 1978.

24. Bouloy, M., Plotch, S.J. and Krug, R.M. Proc. Natl. Acad. Sci. USA 77:3952-3956, 1980.
25. Plotch, S.J., Bouloy, M. Ulmanen, L and Krug, R.M. Cell 23:847-858, 1981.
26. Ulmanen, I., Broni, B. and Krug, R.M. J. Virol. 45:27-35, 1983.
27. Koper-Zwarthoff, E.C., Lockard, R.E., Alzner-DeWeerd, B., RajBhandary, U.L. and Bol,J.F.Proc. Natl. Acad. Sci. USA 74:5504-5508, 1977.
28. Kolakofsky, D. J. Virol. 10:555-559, 1972.
29. Krug, R.M. Curr. Top. Microbiol. Immunol. 93:125-149, 1981.
30. Banerjee, A.K., Abraham, G. and Colonno, R.J. J. Gen. Virol. 34:1-8, 1977.
31. Goldman, N., Presser, I. and Sreevalson, T. Virology 76:352-364, 1977.

13

mRNAs OF UUKUNIEMI VIRUS, A BUNYAVIRUS

RALF F. PETTERSSON, ESA KUISMANEN, RAGNA RÖNNHOLM AND ISMO ULMANEN

1. SUMMARY

Bunyaviruses are enveloped arthropod-borne viruses containing a tripartite single-stranded RNA genome of negative polarity. We have studied the strategy of gene expression of one member of this large family, namely Uukuniemi virus, the prototype of the genus Uukuvirus. Three mRNA species, which do not bind to oligo(dT)-cellulose, are found associated with polysomes of infected cells. Sucrose gradient and gel analyses, in vitro translation and molecular cloning have shown that: (i) The M virion RNA (vRNA) segment is transcribed into an mRNA of roughly the same size (Mr 1.1×10^6) and encodes a precursor (p110) to the two glycoproteins G1 and G2. (ii) Sequence analysis of the 3' end of the M vRNA, which is complementary to the 5' end of the M mRNA, indicates an open reading frame starting from an AUG codon 18 nucleotides from the end of the mRNA. (iii) The S vRNA segment (Mr 0.5×10^6) is transcribed into a full-length plus-strand RNA (antigenome) and two small mRNA species (Mr about 0.3×10^6) that encode a nonstructural (NS_S) protein (Mr 30,000) and the nucleocapsid (N) protein (Mr 25,000). Whether the two mRNA species contain overlapping nucleotide sequences is not known. By elimination, the L mRNA, transcribed from the L vRNA segment (Mr 2.4×10^6), codes for the L protein (Mr 200,000) the putative RNA polymerase. The results obtained with Uukuniemi virus and other bunyaviruses indicate that the strategy of gene expression of these viruses in some aspects is different from that of other known RNA viruses.

2. INTRODUCTION

The Bunyaviridae family of arboviruses comprises more than 200 different viruses grouped into four genera, called Bunyavirus, Nairovirus, Phlebovirus and Uukuvirus (1). In addition, a large number of possible members are still unclassified. The bunyaviruses have been grouped together because of similar structural properties and mode of maturation (1, 2, 3). All characterized members have a lipoprotein envelope containing two glycoproteins, G1 and G2, and an internal, probably helical nucleoprotein

Y. Becker (ed.), VIRAL MESSENGER RNA.

consisting of three single-stranded RNA segments, designated L, M and S, of negative polarity to which multiple copies of the N protein and a few copies of the L protein are associated. The size of the structural proteins and the RNA segments varies between viruses of different genera (2, 3). Both the nucleoproteins (4, 5, 6) and the protein-free RNA segments (7, 8)have a circular structure due to base-pairing of short inverted complementary sequences at the 3' and 5' ends of the RNA segments (7, 9, 10, 11). In viruses of the same genus, about 10-13 nucleotides at the ends of each RNA segment are conserved (12), suggesting some important role of the ends in replication.

As a model, we have studied one bunyavirus, Uukuniemi virus, the prototype of the Uukuvirus genus (1). This virus has two envelope glycoproteins, G1 (Mr 70,000) and G2 (Mr 65,000) (13, 14). The oligosaccharide sequences of the glycans attached to the proteins have been determined (14). The N protein (Mr 25,000) (13, 15) is associated with the RNA segments forming the nucleoproteins (4). The RNA-associated L protein (Mr about 200,000) may represent the RNA-dependent RNA polymerase detected in virions (16, 17). The three circular RNA segments (7) have Mr's of about 2.4×10^6 (L), 1.1×10^6 (M) and 0.5×10^6 (S) (ref. 18) and have a negative polarity (16, 18). In addition to the above mentioned structural proteins, a nonstructural protein (NS) (Mr 30,000) is found in infected cells (17).

A unique feature of the bunyavirus morphogenesis is that the virus particles are formed intracellularly by a budding process at smooth-surface vesicles in the Golgi area (2, 19, 20). Using Uukuniemi virus as a model we have shown by immunofluorescence and immunoelectron microscopy techniques that the site of maturation in fact is the Golgi complex (20). Virus particles are thought to be expelled from the cells in large vesicles, which fuse with the plasma membrane.

Here we summarize our results on the characterization of the mRNAs found in Uukuniemi virus-specific cells and present a model for the general strategy of gene expression of this virus. A part of these results has been published previously (17).

3. RESULTS

3.1. RNA species in Uukuniemi virus-infected cells

3.1.1. Total virus-specific cytoplasmic RNA. The logarithmic growth phase of Uukuniemi virus starts about 5-6 hrs and ends about 16-18 hrs postinfection (pi) (ref. 21). To isolate total cytoplasmic viral RNA enriched for viral mRNAs, infected chick embryo fibroblasts were labeled between 4 and 8 hrs pi in the presence of actinomycin D and the RNAs fractionated on a sucrose gradient (Fig. 1A). In addition to the radioactivity remaining

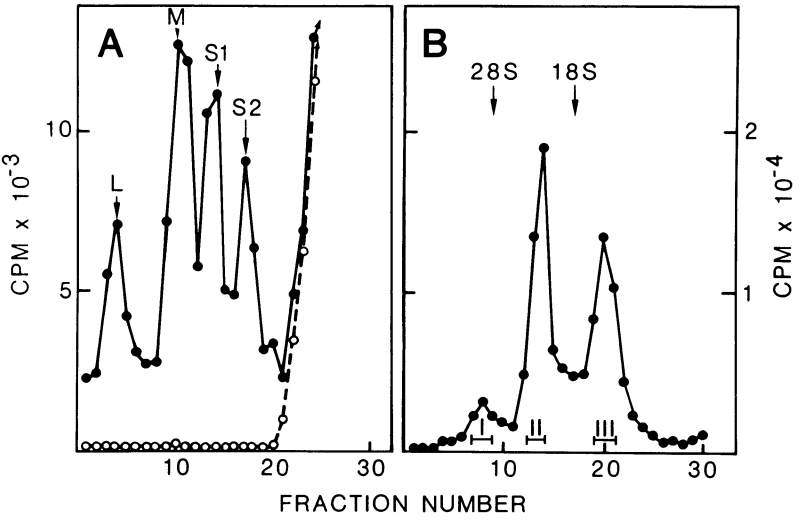

FIGURE 1. Sucrose gradient analysis of Uukuniemi virus-specific total cytoplasmic and polysomal RNA. (A) Infected cells were labeled at 4-8 hrs pi with ^{3}H-uridine in the presence of actinomycin D (2 µg/ml) and the SDS-treated cytoplasmic RNA was fractionated on a 15-30% (wt/wt) sucrose gradient (17 hr, 25,000 rpm, 23°C, SW27 rotor). RNA from mock-infected cells was analyzed on a separate gradient. L, M and S1 indicate the positions of the virion RNA segments. S2 is the 12S mRNA transcribed from the S RNA segment. (B) Analysis of Uukuniemi virus-specific polysomal RNAs. A cytoplasmic extract from cells infected and labeled as above was prepared and the polysomes collected onto a 60% sucrose cushion by sedimentation through a 15-40% gradient. After concentration and treatment with SDS, the RNA was fractionated on a 15-30% sucrose gradient as above. Pools I, II and III correspond to the L, M and S2 mRNA species. The positions of the 28S and 18S rRNAs are indicated. RNA from infected cells (●——●), and mock-infected cells (o----o).

at the top of the gradient, four RNA peaks sedimenting at 29S, 23S, 17S and 12S were observed. We have designated these RNA species L, M, S1 and S2, respectively (17). No radioactivity corresponding to these peaks were recovered from mock-infected cells. The four RNA species were single-stranded since more than 95% of the radioactivity in each peak could be rendered

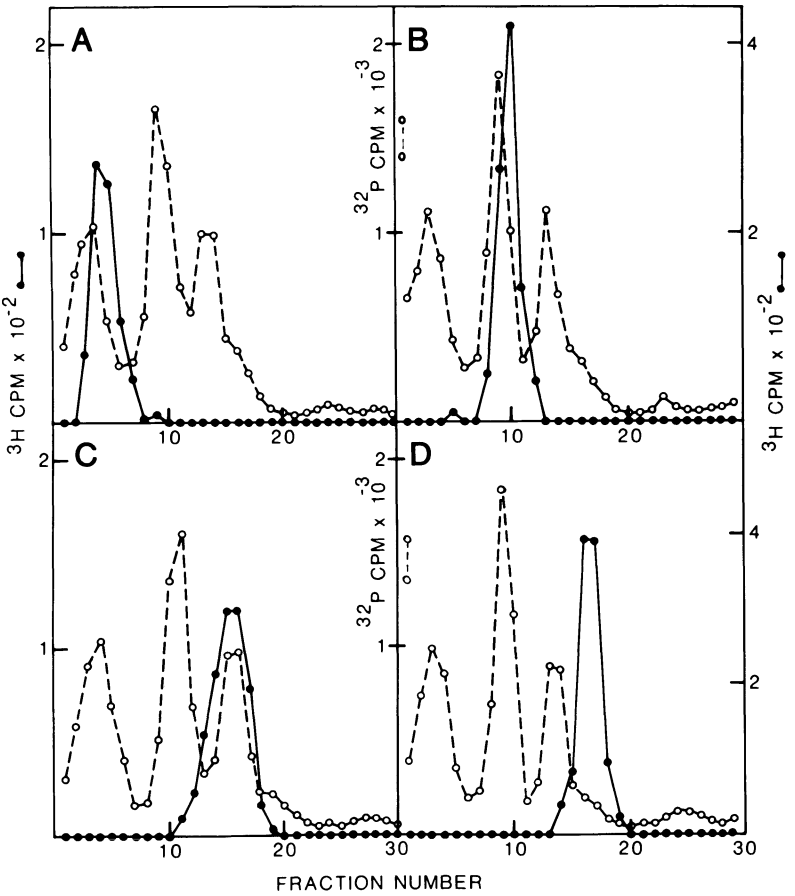

FIGURE 2. Cosedimentation of Uukuniemi virion RNAs with individual polysomal RNA species. ^{32}P-labeled virion RNAs and individual ^{3}H-uridine labeled polysomal RNAs, and the cytoplasmic S1 RNA (Fig. 1A) were mixed and fractionated on a 15-30% (wt/wt) sucrose gradient as described in the legend to Fig. 1. Virion RNAs together with RNA from pool I in Fig. 1B (A), pool II (B), pool III (D) and from the S1 peak fraction (C). ^{32}P-labeled virion RNA (o----o), and ^{3}H-labeled polysomal and S1 RNA (●——●).

acid-soluble by ribonuclease A treatment. Early in infection (3-5 hrs), the 12S RNA is the most abundant species, whereas later (10-20 hrs) it constitutes only a minor species (data not shown). The L, M and S virion RNAs (vRNAs) cosedimented with the three largest cytoplasmic RNAs, whereas no RNA corresponding to the cytoplasmic 12S RNA was found in virions.

3.1.2. Virus-specific polysomal RNA. To study which of the above RNA fractions contained RNA species actively engaged in translation, polysomes were isolated from infected and mock-infected cells labeled at 4-8 hrs pi with ^{3}H-uridine in the presence of actinomycin D. Polysomes, collected onto a sucrose cushion, yielded upon solubilization with SDS three RNA species sedimenting at 29S, 23S and 12S (Fig. 1B) (ref. 17). Resedimentation of each RNA species (pools I, II and III) together with total ^{32}P-labeled vRNAs indicated that the 29S and 23S RNA species sedimented slightly slower than the L and M vRNA segments, respectively (Fig. 2A and B). The 17S RNA (S1) recovered from the total cytoplasmic RNA cosedimented with the virion S RNA (Fig. 2C). As mentioned above, no virion RNA segment cosedimented with the polysomal 12S RNA (Fig. 2D), which we have designated S2.

3.1.3. The S vRNA segment specifies two 12S mRNA species. As reported below and previously (17), the 12S mRNA (S2) fraction directs the *in vitro* synthesis of both the nucleoprotein (N) and a nonstructural protein (NS). We therefore wanted to study whether this RNA species could be resolved into two RNA species that would code for the two proteins. Total ^{32}P-labeled cytoplasmic RNA was prepared from infected cells as described above, denatured with glyoxal and dimethyl sulfoxide (22) and analyzed by polyacrylamide gel electrophoresis (Fig. 3, lane 2). ^{32}P-labeled virion RNA was used as a control (lane 1). In the latter, the L RNA, because of its large size, did not enter the gel and is therefore not seen in the autoradiogram.

Two RNA species (designated S2a and S2b) with Mr of about 0.3×10^{6} that were not present in virions were resolved (Fig. 3, lane 2). The cytoplasmic M and S1 RNA species roughly comigrated with the virion M and S segments.

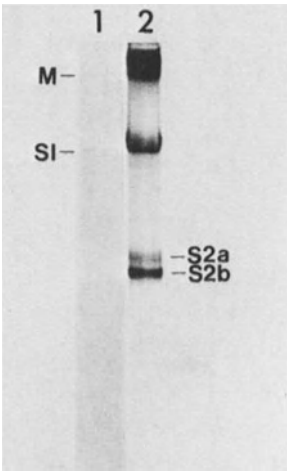

FIGURE 3. Polyacrylamide gel electrophoresis of Uukuniemi virus-specific cytoplasmic RNA. Virus-specific RNA was labeled at 4-6 hr pi with ^{32}P in the presence of actinomycin D and the RNA isolated from a cytoplasmic extract by phenol extraction. The ethanol-precipitated RNA was denatured with 1 M glyoxal and 50% dimethyl sulfoxide at 50° for 1 hr (22) and run on a 2.5-5% gradient polyacrylamide gel (23) (lane 2). ^{32}P-labeled vRNA was run as a control (lane 1). The L RNA did not enter the gel but remained in the stacking gel and is not shown in the figure. The positions of the M and S (S1) RNA species are as indicated. S2a and 2b represent the two small mRNA species transcribed from the virion S RNA segment.

When the cytoplasmic RNA from infected cells was fractionated on a 1% agarose gel, transferred to nitrocellulose paper and hybridized to ^{32}P-labeled purified virion S RNA (24) (Northern blot analysis), a diffuse band with an Mr of about $0.3x10^6$ was found to hybridize (Fig. 4, lane 3). This band, which corresponds in size to the S2a and S2b RNA species, thus contained RNA complementary to the virion S RNA segment. In addition, a band

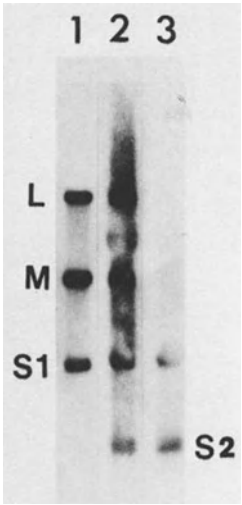

FIGURE 4. Northern blot analysis of Uukuniemi virus-specific complementary RNA. Total RNA was extracted from infected cells at 8 hr pi, denatured with glyoxal and dimethyl sulfoxide (see legend to Fig. 3) and fractionated on a 1% agarose gel. The RNA was transferred to nitrocellulose paper (24) and hybridized to ^{32}P-labeled total virion RNA (lane 2) or isolated S virion RNA (lane 3). ^{32}P-labeled vRNA was run as a control (lane 1).

comigrating with the S vRNA hybridized weakly to the probe. This band probably represents the full-length plus strand RNA species (antigenome) complementary to the virion S RNA.

When total ^{32}P-labeled vRNA was used as probe, RNA species comigrating with virion L, M and S RNA segments were detected, in addition to S2a/2b (Fig. 4, lanes 1 and 2).

We thus conclude that the virion S RNA is transcribed into two small mRNA species, which we call S2a and S2b. In addition, all three segments are copied into apparently full-length complementary RNAs.

3.2. Lack of binding of Uukuniemi virus mRNAs to oligo(dT)-cellulose

We have previously noted that none of the virus-specific RNA species obtained from total cytoplasmic RNA could be bound at high salt to oligo(dT)-cellulose (17). Thus, this method could not be used for enrichment of mRNAs. To study whether the polysomal RNAs could be recovered by oligo(dT)-cellulose chromatography, the individual RNAs (Fig.1, pools I, II and III) were passed several times over an oligo(dT)-cellulose column at 0.3 M NaCl and the bound material eluted with low salt. As shown in Table 1, less than 1% of each mRNA species was retained by the column

Table 1. Binding of Uukuniemi virus polysomal mRNAs to oligo(dT)-cellulose.

RNA	Radioactivity (cpm) Not bound	Bound	(%)
Uukuniemi mRNA			
L	1,208	10	(0.8)
M	8,694	61	(0.7)
S2	7,094	14	(0.2)
Uukuniemi vRNA	7,700	29	(0.4)
SFV 18S DI RNA	850	17,550	(95.4)

^{3}H-uridine-labeled RNAs were bound to oligo(dT)-cellulose (PL-Biochemicals) at 0.3 M NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% SDS and the columns were washed extensively with the same buffer. RNAs were eluted with 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.1% SDS. Radioactivity was determined from eluted fractions by counting in a xylene-Triton X100 scintillation fluid. SFV DI RNA = Semliki Forest virus defective interfering RNA (25).

under conditions where more than 95% of a poly(A)-containing Semliki Forest virus 18S defective interfering RNA (25) was bound. No virion RNA was bound either. These results have been obtained reproducibly also using poly(U)-Sepharose (unpublished results). Thus, it appears that Uukuniemi virus mRNAs either lack a poly(A)-tract, or that it is too short (less than 10-15 residues) to allow binding to these homopolymers.

3.3. In vitro translation of Uukuniemi virus-specific mRNAs

To find out which polypeptides are encoded by each RNA segment the peak fractions from a sucrose gradient similar to the one shown in Fig. 1A were pooled separately and the RNAs translated in a micrococcal nuclease-treated rabbit reticulocyte lysate (17). Corresponding fractions obtained from mock-infected cells were used as controls (referred to as mock-infected RNA).

The 29S mRNA, transcribed from the L RNA segment, did not give rise to any clearly detectable virus-specific polypeptide (Fig. 5A, lane 3) as compared to lysates programmed with the corresponding mock-infected RNA (lane 2).

The 23S mRNA, transcribed from the virion M RNA segment, gave rise to a 110,000-dalton protein (designated p110) (lane 5, arrow) that was not present in translates from mock-infected RNA (lane 4). This polypeptide could be immunoprecipitated with antiserum directed against purified virion polypeptides (Fig. 5B, lane 1). When translation was carried out in the presence of dog pancreas microsomes, which allow posttranslational modifications such as glycosylation and proteolytic cleavages, no p110 was observed. Instead, an immunoprecipitable polypeptide, migrating at the position of the virion glycoproteins G1/G2 (Mr 65,000-70,000) was observed (Fig. 5B, lane 2). Based on these and previous (17) results, we therefore conclude that p110 represents the precursor of G1 and G2, which is cleaved roughly in the middle when translation is carried out in the presence of microsomal membranes.

The 17S cytoplasmic RNA, which was not present on polysomes (Fig. 1A and B), did not give rise to any virus-specific polypeptides (Fig 5A, lane 7), apart from some p110, N and NS proteins

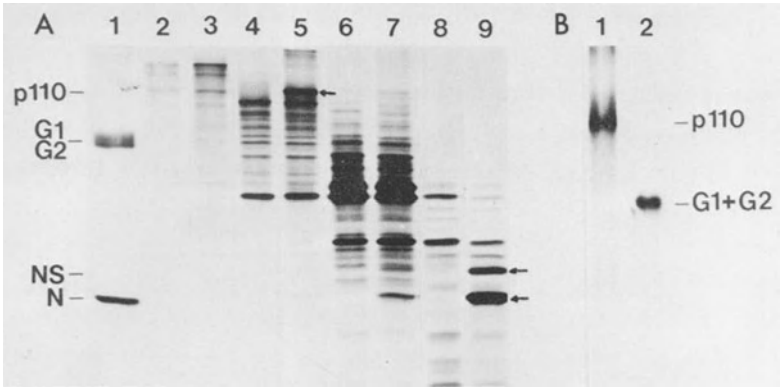

FIGURE 5. In vitro translation of Uukuniemi virus-specific cytoplasmic RNAs. The RNA species from the peak fractions of a sucrose gradient similar to the one in Fig. 1A (pools L, M, S1 and S2) were ethanol precipitated and translated in a micrococcal nuclease-treated rabbit reticulocyte lysate in the presence of ^{35}S-methionine (17). Corresponding RNA fractions from mock-infected cells were analyzed as controls. (A) Virion proteins (lane 1); Translation of the L RNA fraction of infected (lane 3) and uninfected (lane 2) cells; M RNA fraction of infected (lane 5) and uninfected (lane 4) cells; S1 fraction of infected (lane 7) and uninfected (lane 6) cells; S2 fraction of infected (lane 9) and uninfected (lane 8) cells. Arrows indicate the positions of the 110,000-dalton glycoprotein precursor (p110) (lane 5), and the NS (upper arrow) and N (lower arrow) proteins (lane 9). (B) Translation of the M RNA fraction from infected cells in the absence (lane 1) and presence (lane 2) of dog pancreas microsomes (a generous gift from Dr. B. Dobberstein, EMBL, Heidelberg, FRG). The products were immunoprecipitated with antiserum against whole virion proteins. The translation products were analyzed on a 12% (A) and 10% (B) polyacrylamide gel (17) and an autoradiogram prepared.

synthesized as a result of contamination of the 17S RNA fraction with the 23S and 12S RNA species (see above and below).

The 12S mRNA, containing the S2a and S2b mRNA species, derived from the virion S RNA segment, gave rise to two polypeptides: the nucleoprotein N (Mr 25,000) and a 30,000-dalton protein not found in virions (Fig. 5A, lane 9, arrows). These proteins were not translated from mock-infected RNA (lane 8). We have previously shown by immunoprecipitation (17) that the 25K

protein represents the N protein, whereas the 30K protein is a nonstructural protein (hence it is designated NS), since it is not precipitable with antiserum against virion proteins.

Results similar to those presented above have also been obtained by in vitro translation of the individual polysomal mRNA species shown in Fig. 1B (unpublished results).

3.4. Molecular cloning and sequencing of the virion M RNA

With the intention to study the primary structure of the virion M RNA segment and to eventually be able to express a full-length cDNA copy in animal cells, we have initiated the cloning and sequencing of this RNA species. The cloning procedure was carried out as described in the legend to Fig. 6. So far only the sequence of the 3' end of the virion M RNA has been obtained. This sequence should correspond to the 5' end of the mRNA encoding the signal peptide and the N-terminus of the G1 glycoprotein. We have recently established that the gene coding for G1 is N-terminal to that of G2 (E. Kuismanen, unpublished results). It has previously been shown that the nucleotide sequence of the extreme 3' end of each Uukuniemi virus RNA segment is 3' UGUGUUUC 5' (ref. 26). One of our cDNA clones, which is 750 bp long, and hybridizes only to the virion M RNA starts with this sequence. Thus, this cDNA clone extends to the last nucleotide at the 3' end of the M segment. The complementary sequence (5' end of mRNA) is shown in Fig. 6 and reveals an open reading frame starting from an AUG initiation codon at nucleotide 18-20. There are two other AUG codons in the same reading frame (residues 84-86 and 180-182) that also could serve as initiator codons. The only other AUG codon within this short cDNA stretch is located at nucleotides 88-90 and is immediately followed by a UGA termination codon. The N-terminal amino acid sequence deduced from the open reading frame starting at nucleotide 18 displays a stretch of nonpolar, mostly hydrophobic residues between amino acids 4 and 19 (including six consecutive leucine residues) typical of signal peptide sequences. Since the other two in frame AUG codons do not give such a nonpolar N-terminus (Fig. 6 and unpublished sequence data), we believe

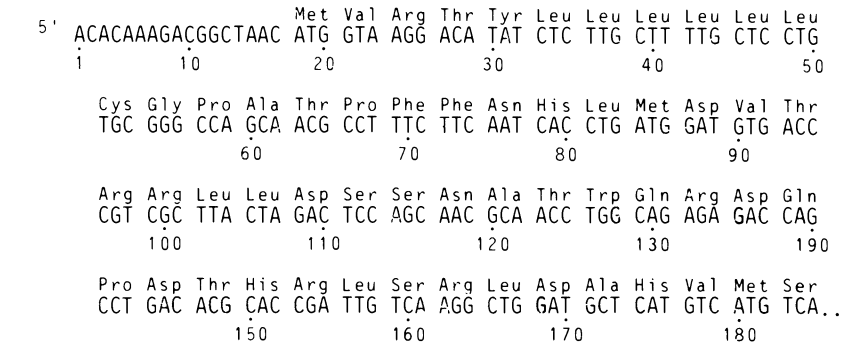

FIGURE 6. Nucleotide sequence of the 5' end of the complementary M RNA as deduced from the 3' end sequence of the virion M RNA segment. Purified M vRNA was tailed at its 3' end with A-residues using E. coli poly(A) polymerase (Bethesda Research Lab.) (ref. 27) and cDNA synthesized using AMV reverse transcriptase (Life Sciences Inc., Florida), and oligo(dT)$_{12-18}$ (PL-Biochemicals) as primer (28). The second strand was also made with reverse transcriptase. Following S1 nuclease (Calbiochem) treatment and tailing with dCTP and terminal transferase (BRL), the cDNA was cloned into the dG-tailed PstI site of pBR322 and the hybrid plasmids transformed into E. coli (28). Transformants containing M vRNA-specific inserts were identified by hybridization using ^{32}P-labeled virion M RNA as a probe. One cDNA clone containing a 750 base pair insert was characterized and partially sequenced by the Maxam-Gilbert method (29). The sequence complementary to the extreme 3' end of the virion M RNA is shown. This sequence should represent the 5' end of the M mRNA. The deduced amino acid sequence of an open reading frame starting with an ATG at nucleotide 18 is also shown.

the AUG at nucleotide 18 to be the initiatior codon for the signal peptide of G1. Since the amino acid sequence at the N-terminus of the virion G1 is not known, the possible cleavage site between the signal peptide and the mature G1 is at present unknown.

4. DISCUSSION

Based on the results presented here and previously, we have schematically summarized our current view on the strategy of Uukuniemi virus gene expression in Fig. 7. Virions contain three single-stranded RNA segments, L, M and S, of negative polarity. Due to hydrogen bonding between inverted complementary sequences at their 3' and 5' ends, they have a circular structure (7). The circularization of the segments is likely to play a central, but as yet undefined role in replication and/or encapsidation of the vRNA (30).

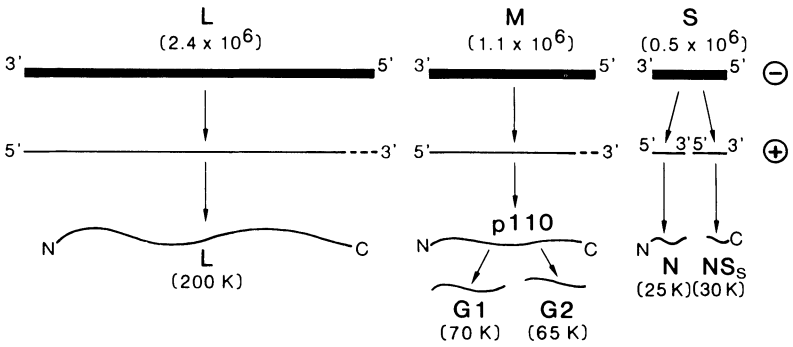

FIGURE 7. General strategy of gene expression of Uukuniemi virus. For details, see Discussion.

The L and M vRNA segments are each transcribed into mRNAs of roughly the same size as the templates. Whether the mRNAs are identical in size or slightly shorter than the corresponding vRNA segments remains to be determined. In the light of the results obtained for viruses of the Bunyavirus genus, such as Germiston (31), La Crosse (Patterson and Kolakofsky, pers.comm.) and Akabane (32) viruses, which indicate that the S mRNA lacks about 100-150 nucleotides at its 3' end (corresponding to the 5' end of the S vRNA), we have tentatively drawn a broken line at the 3' ends of the L and M mRNAs (Fig. 7) to indicate that the transcripts may be truncated at this end. Although there is no

direct evidence as yet for such a hypothesis, we find it plausible, since such molecules would not be able to form circular structures, which may be crucial for replication and/or encapsidation. Such molecules would therefore serve only as mRNAs and not participate in the replication cycle. We found here that the polysomal L and M RNAs sedimented slightly slower than the corresponding vRNA segment, whereas their mobilities in gels were roughly the same. This may suggest that their size is similar, but their conformation different, i.e. the vRNA L and M segments would be circular structures (7), whereas the corresponding mRNAs could be linear molecules. The 5' end of the S mRNA of La Crosse (33) and Germiston (31) viruses is apparently complementary to the very 3' end of the virion S RNA. That the 5' end of the Uukuniemi virus M mRNA indeed may extend to the very end of the M vRNA is supported by preliminary sequence data for the 3' end of the M vRNA. The sequence complementary to the very 3' end of the M vRNA which should correspond to the 5' end of the M mRNA, contains a potential AUG initiation codon 18 nucleotides from the end that results in an open reading frame for at least 190 amino acid residues (Fig. 6 and unpublished data).

In contrast to the mRNAs, full-length transcripts of virion L, M and S RNA segments (antigenomes) would have inverted complementary ends and would thus be able to form circles. These molecules would therefore serve as templates for new genomic segments.

The Uukuniemi virus M RNA segment encodes a precursor (p110) to the two glycoproteins G1 and G2, as shown by *in vitro* translation of M mRNA. In the presence of dog pancreas microsomes, p110 was quantitatively processed to G1/G2 by proteolytic cleavage roughly in the middle. We have recently shown that the unglycosylated forms of G1 and G2 both have Mr's of about 55,000 in good agreement with the Mr of the unglycosylated precursor (110,000). The gene coding for G1 is located N-terminal to that of G2 (E. Kuismanen, unpublished results). So far we have not identified an M vRNA-encoded nonstructural protein (NS) either *in vivo* (17) or *in vitro* (Fig. 5A, lanes 6 and 7) similar to the one reported for La Crosse and snowshoe hare viruses (34).

That the L vRNA segment of Uukuniemi virus and other bunyaviruses encodes the L protein (Mr about 180,000-200,000), the putative RNA polymerase (16, 35), is based on elimination and on the fact that this RNA species, at least in the case of Uukuniemi virus (Fig. 7), is the only one sufficiently large to encode that protein.

The Uukuniemi virus S vRNA is transcribed into a full-length plus strand RNA (antigenome) (S1) and two small mRNA species, which we have designated S2a and S2b*. These RNA species, after denaturation, clearly separated on a gradient polyacrylamide gel, but comigrated on agarose gels and cosedimented on sucrose gradients. The S2 fraction from the sucrose gradient directed the *in vitro* synthesis of the nucleoprotein N (Mr 25,000) and the nonstructural protein NS (Mr 30,000). Using the nomenclature adopted for the S vRNA encoded nonstructural protein of the *Bunyavirus* genus (34, 36), we designate this protein NS_S. Both *in vivo* (17) and *in vitro*, the Uukuniemi virus N protein is the predominant one relative to NS_S. Since the S2b RNA was the predominant mRNA species (Fig. 3), it is likely that this mRNA codes for the N protein, a conclusion that also fits the size determinations of the mRNAs and the proteins. We do not yet know the order of the N and NS_S genes on the S vRNA segment. Thus, the order in Fig. 7 is arbitrary. Neither do we know whether S2a and S2b mRNAs contain any overlapping nucleotide sequences. In the case of the members of the *Bunyavirus* genus, the N and NS_S proteins are coded for by completely overlapping nucleotide sequences translated in two different reading frames (36, 37, see below). To answer this question for Uukuniemi virus, hybridization experiments and direct sequencing of the cloned mRNAs have to be carried out.

We have repeatedly been unable to bind Uukuniemi virus mRNAs to oligo(dT)-cellulose of poly(U)-Sepharose. This suggests that the mRNAs either lack a poly(A)-tract, or that the poly(A) is too short to allow binding. These results are in agreement with

---

*In our nomenclature, 1 (one) (e.g. S1) stands for full-length complementary RNA (antigenome) and 2 (e.g. S2a and S2b) for complementary mRNA.

those obtained by others for different bunyaviruses (31, 38). However, that the bunyavirus mRNAs indeed lack poly(A) has in no case been yet confirmed by chemical analysis or sequencing. A sequence, GUUUUU, about 100 nucleotides from the 5' end of virion S RNA has been identified in La Crosse and snowshoe hare viruses (37, 38) and suggested to be a transcription termination and polyadenylation signal (Patterson and Kolakofsky, pers.comm.) similar to the one present in influenza virus RNA segments (40). If this sequence indeed is a termination signal, it would suggest, analogously to e.g. influenza and vesicular stomatitits virus (40, 41, 42) that at least the S mRNA could terminate in a poly-(A) tract, although perhaps a short one.

Uukuniemi virus is a member of the genus Uukuvirus. Thus, the details of the strategy of gene expression may not be identical to those of viruses belonging to the other Bunyaviridae genera. Based on data available at present, there does not, however, seem to be major differences. So far the presence of two small mRNAs derived from the S vRNA of Uukuniemi virus, and only one from the S vRNA of several members of the Bunyavirus genus indicate that some differences exist. In the case of viruses of the Bunyavirus genus, the S vRNA appears to be transcribed into only one (or a nested set of) near full-length mRNA (31, 38, Patterson and Kolakofsky, pers.comm.), which is translated into both the N and NS_S proteins by employing two different overlapping reading frames (36, 37, 38, 43).

At present, only the S vRNA of the Bunyavirus members has been studied in detail. The S segment of La Crosse and snowshoe hare viruses have been cloned and fully sequenced (36, 37, 39). It appears that the transcription of the S vRNA into the mRNA(s) terminates at about 100-150 bases from the 5' end of the vRNA. The 5' end of the mRNA contains a heterogeneous segment of 10-14 nucleotides, apparently derived from host cell mRNAs (33, Patterson and Kolakofsky, pers.comm.), a situation resembling that of influenza virus mRNAs (44, 45). Since the replication of bunyaviruses is insensitive to α-amanitin and actinomycin D (46), and the bunyaviruses appear to be able to grow in enucleated cells (47), this suggests that the addition of the host-derived

5' ends to bunyavirus mRNAs takes place in the cytoplasm. Interestingly, it has been shown recently that both primary and secondary transcription, as well as replication, require continuous protein synthesis (32, 38, Patterson and Kolakofsky, pers.comm.). Whether this means that the continuous synthesis of a labile host cell protein is required, or that only host mRNAs actively engaged in protein synthesis can deliver their 5' ends to the mRNAs, remains to be elucidated. Although the role of the 5' terminal extension is unknown, it is likely to serve a primer function in initiation of transcription similarly to the influenza virus system (44, 45). It appears that a long awaited explanation for the very poor in vitro activity of the bunyavirus virion-associated RNA polymerase (16, 35) has at last been offered.

Kolakofsky and coworkers have recently shown that a set of short leader RNA sequences are transcribed from the 3' end of the S vRNA and probably also from the L and M vRNAs(48). They interpreted their results to mean that a nested set of mRNAs with common 3' ends and different 5' ends are synthesized in La Crosse virus-infected cells from the S segment. According to their hypothesis, one set of mRNAs, using the most 5' located initiator AUG codon, would give rise to the N protein, while another set of mRNAs, slightly shorter and lacking that AUG codon, would give rise to the NS_S protein by using a downstream located AUG codon and a different reading frame. Another interpretation would be that the leader RNAs of the bunyaviruses serve a function similar to that of the leader RNAs in vesicular stomatitis virus and Sendai virus -infected cells (49). These leader RNAs, together with the availability of the nucleocapsid protein, have been shown to regulate the replication-transcription pathways.

As is evident from the data summarized above, the mechanism of transcription of bunyaviruses is not yet fully understood. It is, however, clear that these viruses have developed a very interesting and unique strategy for gene expression, which has ingredients from e.g. the orthomyxo-, paramyxo- and rhabdo-viruses.

ACKNOWLEDGEMENTS

We wish to thank Dr. D. Kolakofsky for communicating results from his laboratory prior to publication. This work was supported by grants from the Sigrid Jusélius Foundation.

5. REFERENCES

1. Bishop, D.H.L., Calisher, C.H., Casals, J., Chumakov, M.T., Gaidamovich, S.Ya., Hannoun, C., Lvov, D.K., Marshall, I.D., Oker-Blom, N., Pettersson, R.F., Porterfield, J.S., Russell, P.K., Shope, R.E., and Westaway, E.G. Intervirology 14: 125-243, 1980.
2. Bishop, D.H.L. and Shope, R.E. In: Comprehensive Virology, vol. III (Eds. H. Fraenkel-Conrat and R. Wagner). Plenum Press, pp. 1-156, 1979.
3. Bishop, D.H.L. Curr.Topics Microbiol.Immunol. 68:1-33, 1979.
4. Pettersson, R.F. and von Bonsdorff, C.-H. J.Virol. 15: 386-392, 1975.
5. Samso, A., Bouloy, M. and Hannoun, C. C.R.Acad.Sci. (Paris) 280D:779-782, 1975.
6. Obijeski, J.F., Bishop, D.H.L., Palmer, E.L. and Murphy, F.A. J. Virol. 20:664-675, 1976.
7. Hewlett, M.J., Pettersson, R.F. and Baltimore, D. J.Virol. 21:1085-1093, 1977.
8. Samso, A., Bouloy, M. and Hannoun, C. C.R.Acad.Sci. (Paris) 282D:1653-1655, 1976.
9. Obijeski, J.F., McCaulel, J. and Skehel, J.J. Nucl.Acids Res. 8:2431-2438, 1980.
10. Pardigon, N., Vialat, P., Girard, M. and Bouloy, M. Virology 122:191-197, 1982.
11. Patterson, J.L., Kolakofsky, D., Holloway, B.P. and Obijeski, J.F. J.Virol. 45:882-884, 1983.
12. Clerx-van Haaster, C.M., Clerx, J.P.M., Ushijima, H., Akashi, H., Fuller, F. and Bishop, D.H.L. J.gen.Virol. 61:289-292,1982
13. von Bonsdorff, C.-H. and Pettersson, R.F. J.Virol. 16: 1296-1307, 1975.
14. Pesonen, M., Kuismanen, E. and Pettersson, R.F. J.Virol. 41:390-400, 1982.
15. Pettersson, R., Kääriäinen, L., von Bonsdorff, C.-H. and Oker-Blom, N. Virology 46:721-729, 1971.
16. Ranki, M. and Pettersson, R.F. J.Virol. 16:1420-1425, 1975.
17. Ulmanen, I., Seppälä, P. and Pettersson, R.F. J.Virol. 37:72-79, 1981.
18. Pettersson, R.F., Hewlett, M.J., Baltimore, D. and Coffin, J.M. Cell 11:51-63, 1977.
19. Murphy, F.A., Harrison, A.K. and Whitfield, S.G. Intervirology 1:297-316, 1973.
20. Kuismanen, E., Hedman, K., Saraste, J. and Pettersson, R.F. Mol.Cell Biol. 2:1444-1458, 1982.
21. Pettersson, R. Med.Biol. 52:90-97, 1974.
22. McMaster, G.K. and Carmichael, G.G. Proc.Natl.Acad.Sci. USA 74:4835-4838, 1977.
23. Jeppesen, P.G.N. Methods Enzymol. 65:305-319, 1980.

24. Thomas, P.S. Proc.Natl.Acad.Sci. USA 77:5201-5205, 1980.
25. Kääriäinen, L., Pettersson, R.F., Keränen, S., Lehtovaara, P, Söderlund, H. and Ukkonen, P. Virology 113:686-697, 1981.
26. Parker, M.D. and Hewlett, M.J. In: The replication of negative strand viruses. (Eds. Bishop, D.H.L. and Compans, R.W.) Elsevier North-Holland, Inc. pp. 125-133, 1983.
27. Sippel, A.E. Eur.J.Biochem. 37:31-40, 1973.
28. Söderlund, H., Keränen, S., Lehtovaara, P., Palva, I., Pettersson, R.F. and Kääriäinen, L. Nucl.Acids Res. 9: 3403-3417, 1981.
29. Maxam, A.M. and Gilbert, W. Methods Enzymol. 65:499-560, 1980.
30. Baltimore, D., Pettersson, R.F., Flanegan, J.B., Hewlett,M.J., Rose, J. and Ambros, V. In: Perspectives in Virology, vol. 10 (Ed. M. Pollard), Raven Press, New York, pp.1-12. 1978.
31. Bouloy, M., Vialat, P., Girard, M. and Pardigon, N. J.Virol., in press.
32. Pattnik, A.K. and Abraham, G. J.Virol. 47:452-462, 1983.
33. Bishop, D.H.L., Gay, M.E. and Matsuoko, Y. Nucl.Acids Res. 11:6409-6418, 1983.
34. Fuller, F. and Bishop, D.H.L. J.Virol. 41:643-648, 1982.
35. Bouloy, M. and Hannoun, C. Virology 69:258-264, 1976.
36. Fuller, F., Bhown, A.S. and Bishop, D.H.L. J.gen.Virol. 64:1705-1714, 1983.
37. Akashi, H. and Bishop, D.H.L. J.Virol. 45:1155-1158, 1983.
38. Abraham, G. and Pattnik, A.K. J.gen.Virol. 64:1277-1290, 1983.
39. Cabradilla, C.D., Jr, Holloway, B.P. and Obijeski, J.F. Virology 128:463-468, 1983.
40. Robertson, J.S., Schubert, M. and Lazzarini, R.A. J.Virol. 38:157-163, 1981.
41. Gupta, K.C. and Kingsbury, D.W. Virology 120:518-523, 1982.
42. Rose, J.K. Cell 19:415-421, 1980.
43. Bishop, D.H.L., Could, K.G., Akashi, H. and Clerx-van Haaster, C.M. Nucl.Acids Res. 10:3703-3713, 1982.
44. Krug, R.M. Curr. Topics Microbiol.Immunol. 93:125-150, 1981.
45. Braam, J., Ulmanen, I. and Krug, R.M. Cell 34:609-618, 1983.
46. Vezza, A.C., Repik, P.M., Cash, P. and Bishop, D.H.L. J.Virol. 31:426-436, 1979.
47. Goldman, N., Presser, I. and Sreevalsan, T. Virology 76: 352-364, 1977.
48. Patterson, J.L., Cabradilla, C., Holloway, B.P., Obijeski, J.F. and Kolakofsky, D. Cell 33:791-799, 1983.
49. Leppert, M.L., Rittenhouse, L., Perrault, J., Summers, D.F. and Kolakofsky, D. Cell 18:735-749, 1979.

14

FOOT-AND-MOUTH DISEASE VIRUS RNA

JOSE L. LA TORRE AND EDUARDO A. SCODELLER

SUMMARY

In this work we report results obtained in our laboratory concerning the existence in foot-and-mouth disease virus (FMDV) of a virion-associated endoribonuclease, the conditions used for its activation and the steps taken toward the development of an experimental vaccine based on these observations.

In addition, we present data related to the biochemical characterization of the FMDV genome and its biological and epidemiological relevance.

INTRODUCTION

Foot-and-Mouth Disease is considered the most important disease affecting all domestic and wild cloven-hoofed animals. The disease is highly contagious and usually produces lesions on feet, mouth, snouts and teats that result in a loss of about 25% in productivity of farm animals (1). Indirect economical losses are also very high due to the fact that international markets are restricted or closed to animal products from infected countries (1, 2).

The causative agent of the disease is an acid-sensitive virus belonging to the aphthovirus genus of the animal picornaviruses (3). The virions are naked icosahedral particles of about 25 nm in diameter (Fig. 1 A) (4, 5) having a buoyant density of about 1.425 g/cm^3 and a sedimentation coefficient of 145S (1, 4, 6).

The viral capsid is composed of 60 copies each of four coat proteins VP_1, VP_2, VP_3 and VP_4. One or two copies of other polypeptides designated as VP_0 (the precursor of the capsid proteins VP_2 and $VP4$) and P_{56} (the viral RNA polymerase) are also present in the mature virions (7). The viral structural polypeptides VP_1, VP_2 and VP_3 are arranged in 12 pentameric subunits each of which forms one of the 12 vertices of the icosahedron (Fig. 1 B) (5). VP_1 is the only structural polypeptide capable of inducing

viral neutralizing antibodies (8, 9, 10) and is presumably clustered at the apices of the virions (11). Viral polypeptide VP_4 appears to be internally located (12).

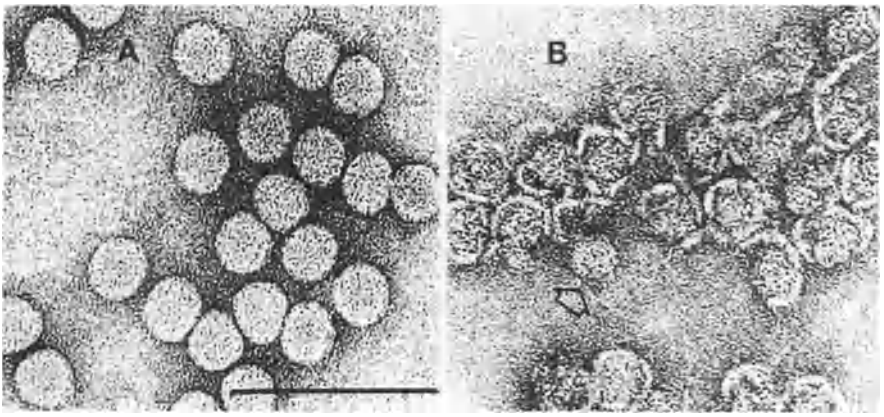

FIGURE 1. FMDV virus type C_3 strain Resende was replicated in BHK cells, purified by means of a sucrose gradient and observed directly (A) or after spreading onto a subphase of 0.1 M ammonium acetate (pH 6.2) (B). Arrow in B shows a core particle. Magnification x 360,000. The bar represents 0.1 µm. For details see ref.18. Reprinted with permission of Academic Press.

The viral genome is an infectious single-stranded RNA of about 8,000 nucleotides long (6) and of positive polarity. This RNA has a poly (A) tract very variable in length (13) and located at the 3' end of the molecule. Another characteristic of the FMDV RNA is the presence of a poly (C) tract 80 to 200 nucleotides long internally located at about 400 nucleotides from the 5' end (14). As in the rest of the picornaviruses, a protein of about 3,000 daltons (VPg) is covalently linked to the 5' end of the genome (15, 16, 17). The genomic RNA is buried within the capsid probably folded in a compact spherical configuration, forming the viral core (18) (Fig. 1 B). This RNA is easily degraded into subgenomic fragments when FMDV virions are incubated at 37° leading to a progessive loss of the virus infectivity (19). The origin of this fragmentation is the presence within intact virions of a ribonuclease which, under appropiate physicochemical conditions, is capable of promoting a complete degradation of the virion RNA without affecting the integrity of the particles (see below).

Another relevant property of FMDV is its antigenic variation probably

promoted, as in the rest of the RNA viruses, by a high mutational rate (20). Recombination between the genomes of similar viruses has also been postulated as a mechanism for virus variability (21).

We have chosen as the subject of this chapter these two properties of FMDV. The work was separated into two parts. One describes all the studies we have performed to ascertain the origin of the degradation of FMDV genomic RNA and how we have used this observation to develop an experimental vaccine for FMDV. The other part is dedicated to describe the use of biochemical techniques to study virus variability and other biological and biochemical properties of the FMDV genome. The perspectives of future work in both areas are also discussed.

MATERIALS AND METHODS

Cell cultures and virus preparations. Different FMDV types were used through this work. The origin and conditions of replication in BHK21 clone 13 cells are given in each particular experiment.

Virus purification. In all cases, infections were carried out at high multiplicity (higher than 10 plaque forming units per cell (pfu/cell)). Adsorption period was 45 min at 37°. Virus was then purified by two different procedures.

Method A. At the end of the infection (6 to 10 hr), medium was centrifuged at 10,000 g for 30 min, and the virus was then concentrated from the resulting supernatant with 6% polyethylene glycol (PEG 6000) or by centrifugation at 105,000 g for 1 hr. The precipitated virions were resuspended in ET buffer (0.001 M EDTA, 0.030 M Tris-HCl, pH 7.4) with gentle magnetic agitation, for at least 8 hr at 4°, and were purified by means of a CsCl density gradient (18). The virus-containing fractions were pooled and dialyzed overnight in the cold (4°) against ET buffer. Sodium deoxycholate was added to the dialyzate to 0.5% final concentration and the sample was layered on top of a 15 to 45% (wt/vol) linear sucrose gradient made in ET buffer. Gradients were centrifuged at 27,000 rpm for 4 hr at 2-3° in an SW 27 rotor of a Beckman ultracentrifuge. At the end of the run, fractions of 0.5 ml each were collected by puncturing the bottom of the tubes. The OD_{260nm} was measured and the radioactivity of each fraction was assayed; finally those fractions containing virus were pooled.

Method B. In order to obtain freshly produced virus, medium was never

harvested at periods later than 4 hr post-labeling. If at that time cultures were not completely degenerated their medium was replaced at 4-hr intervals; collected fluids were kept cold (0°) until processed. Usually two harvests were pooled and centrifuged at 10,000 g for 10 min and the supernatant was centrifuged for 80 min at 123,000 g in order to pellet the virus. Virion-containing pellets were resuspended in 2 ml of NET buffer (0.100 M NaCl; 0.001 M EDTA; 0.050 M Tris-HCl, pH 7.4) and after adding sodium deoxycholate (0.5% final concentration) the mixture was vigorously stirred for 5 min in the cold. Finally the sample was layered at top of a 15-45% (wt/vol) linear sucrose gradient made in NET buffer and analysed as in Method A.

To obtain virus devoid of external nucleases the pooled peaks were brought to 0.5% Sarkosyl and repurified in an additional sucrose gradient.

Preparation of viral and cytoplasmic RNAs. It was performed as described in ref. 32. Briefly: samples of purified virions, suspended in NET buffer (pH 7.4) were made 1% (final concentration) with Sarkosyl NL 97 and were treated twice with an equal volume of chloroform-phenol. RNA was recovered from the aqueous phase by precipitation with 2.5 vol of ethanol at -20° overnight, centrifuged at 10,000 g for 15 min, and dissolved in the appropiate buffer. Alternatively, RNA was obtained from virions by release at low pH: 0.1 ml of the viral suspension was mixed with 1.9 ml of citrate buffer (0.005 M sodium citrate; 0.100 M NaCl, pH 5.4), 0.5% (final concentration) Sarkosyl was added, and the mixture was incubated at 37° for 10 min.

Virus purification and RNA analysis of viral suspensions incubated in inactivation medium. At 5 hr pi ^{32}P-labeled cells and the medium were frozen and then thawed in order to disrupt the cells. Nuclei and cellular debris were removed by centrifugation (3,000 g, 5 min) and the supernatant used as ^{32}P-labeled viral suspension. These suspensions were then incubated at different pHs and ionic conditions, as described in Results. After incubation, 2 ml of the viral suspension were brought to 0.2% Sarkosyl 20 mM EDTA and layered on top of a 35 ml 10-30% (wt/vol) sucrose gradient made in NET buffer and centrifuged 27,000 rpm, for 3.5 hr in an SW27 rotor (Beckman), at 4°. One ml aliquots were collected from the bottom using a peristaltic pump. Fractions containing the 140S viral peak were pooled, made 0.5% SDS and their RNAs extracted twice with phenol-chloroform, ethanol precipitated and analysed as described in Results.

Acrylamide gel electrophoresis. RNA samples were analysed on 2.7% acrylamide gels as described in ref.28.

Labeling of the RNA with (^{32}P) orthophosphoric acid and RNase T_1 fingerprinting. Conditions used were as described in ref.38.

Characterization of the poly (C) tract by affinity chromatography. In order to isolate the poly (C) tract(s), a fraction of the RNase T_1 digest was diluted with 500 mM NaCl, 10 mM Tris-HCl (pH 7.5) and 0.2% SDS (binding buffer) and chromatographed twice at room temperature on oligo-(dG)-cellulose microcolumns which were previously equilibrated with binding buffer. Fractions containing unbound material, poly (C) were discarded. The poly (C)+ fractions were eluted with elution buffer (10 mM Tris HCl plus 0.2% SDS, pH 7.5) and also precipitated with ethanol at -20° overnight and analysed on one-dimensional gels (38).

RESULTS

PART A

Degradation of the FMDV RNA by a virion-associated ribonuclease.

As it was stated above, one of the surprising biochemical characteristics of FMDV is the fragility of the genomic RNA when the virions are incubated at 37°. Analysis of the viral RNA by means of several conventional procedures indicates that, besides the 35S peak belonging to full-length molecules, variable amounts of subgenomic fragments were always detected (22, 23, 24). Several mechanisms have been proposed to explain the heterogeneous nature of the viral RNA, such as the presence of weak linkages within the molecule (25), cationic attack (19), effect of heat (26) and degradation by nucleases (27). In this regard, it was important to determine whether this heterogeneity was an inherent property of FMDV (i.e. fragmented genome) or if it was the result of RNA degradation during purification of the virus particles.

Supecting that the heterogeneity of the RNA was introduced by prolonged handling of the sample during virus purification, we devised two different purification protocols (28). Method A. It was a conventional procedure which required a prolonged maintenance of virus at 4°(i.e. 8-10 hr stirring of a polyethylene glycol precipitated viral pellet, overnight dialysis and CsCl gradients. Method B. It was a shortened purification procedure where virus was never harvested later than 5 hr pi. The harvests were quickly clarified at low speed centrifugation, the virus was pelleted,vigorously resuspended in NET buffer containing detergents,

and layered on top of an appropiate sucrose gradient. Prolonged stirring of virus pellets, CsCl gradients and overnight dialysis were thus avoided. With this method, purified virus can be obtained in a few hours. As can be seen in Fig. 2 A and 2 B, the RNA obtained from virus purified by Method A was composed of a main 35S component with a high proportion of low molecular weight (m.w.) RNA species. However, viral RNA isolated by Method B migrated as a single homogeneous peak when analysed on sucrose gradients or acrylamide gels (Fig. 2 C and 2 D).

From these data we concluded that when FMDV was purified by Method A we were not able to obtain homogeneous RNA. Nevertheless, we have shown that FMDV RNA can be isolated as a single homogeneous molecule using Method B for viral purification. Moreover, RNA obtained by this procedure was very stable when submitted to different denaturation procedures (28). By using calibrated acrylamide gels, we have estimated that the m.w. of the viral RNA is 2.7×10^6 (28). The sedimentation coefficient of this RNA was estimated to be 35S (28), a value in agreement with those previously reported. Finally, since other causes of RNA heterogeneity were ruled out, we thought that a nuclease could be the cause of FMDV RNA breakdown. This was further supported by the results described above which show that the viral RNA can be obtained unbroken when incubation of virions at 37° was avoided during virus production, harvesting and purification. On the other hand, it has been shown that incubation of FMDV at 37° for different lengths of time led to an inactivation of the virus with conservation of its physical and immunological properties. This inactivation was correlated with a simultaneous degradation of the 35S genomic RNA into fragments of smaller size (19). Similar results were obtained with rhinovirus type 14 (RV14) (29) and it was suggested that inactivation of picornaviruses at low temperature also involved a mechanism directed against the viral RNA (30). Since degradation of RV14-RNA during incubation occurred "in situ" within the viral particle, Gauntt (29) postulated the presence of a nuclease located inside the virions.

For this reason, it was decided to study the degradation of the FMDV RNA in highly purified virions to find out whether or not they contained a nucleolytic activity responsible for the degradation of the viral genome. Virus purified by Method B was resuspended in 1% Sarkosyl in NET buffer and further purified in a sucrose gradient to eliminate any residual contamination with external nucleases (31). An aliquot of the treated

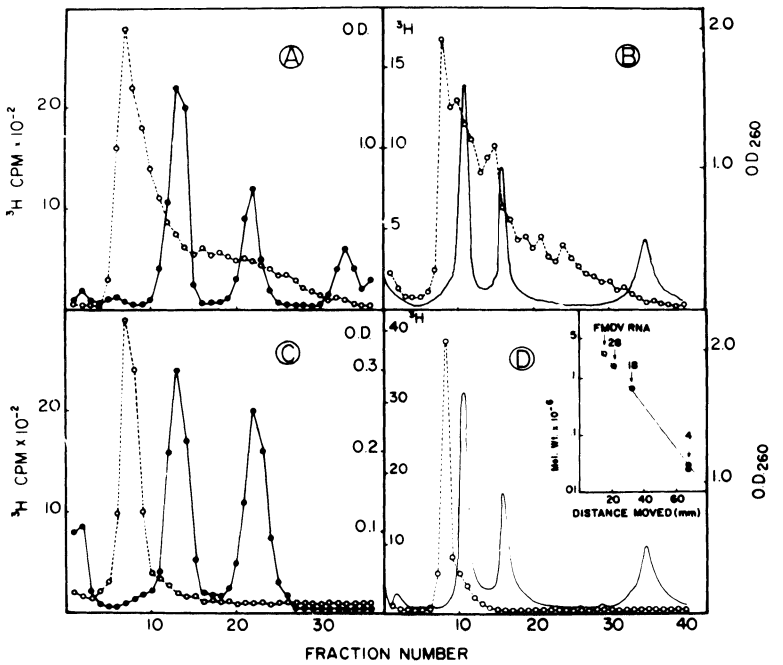

FIGURE 2. Characterization of ^{3}H-labeled FMDV-RNA. (A and B) FMDV-RNA extracted with phenol-chloroform from virions purified by Method A and mixed with unlabeled chick embryo cytoplasmic RNA. The RNA samples were layered over 36 ml 10-30% (wt/vol) linear sucrose gradients made in NET buffer plus 0.5% Sarkosyl and were centrifuged (Beckman SW27 rotor, 27,000 rpm, 15 hr, 2° (A); or were analysed on a 2.4% acrylamide gel (4 hr, 5 mA/gel) (B). (C and D) FMDV-RNA extracted from virions purified by Method B and mixed with unlabeled ribosomal RNA; in (C) viral RNA was obtained by disruption of virions at low pH. Conditions for analysis were the same as those described in Fig. 2 A and 2 B, respectively. (●——●) OD_{260nm} of cytoplasmic or ribosomal RNA. (o——o) ^{3}H-labeled FMDV-RNA (for gradients TCA-insoluble counts per minute were plotted). Insert of Fig. 1 D: relationship between electrophoretic mobilities and the log of the molecular weights; a straight line could be drawn between the 28, 18, and 4S marker RNA (●) ; FMDV RNA (o) was inserted according to its electrophoretic mobility. Reprinted with permission of Academic Press (reference 28).

---

virus was mixed with ribosomal RNA and held at 0° for 15 hr in NET buffer; another aliquot was incubated at 37° under the same conditions. As can be

seen in Fig. 3 A, the RNA extracted from virions incubated at 0°(control) and the ribosomal RNAs sedimented as homogeneous peaks. Fig. 3 B shows the profile of the viral RNA extracted from Sarkosyl-treated virions that were incubated at 37° for 15 hr. As can be appreciated, the viral RNA was degraded during the incubation and very few molecules were found sedimenting at the 35S position. On the contrary, ribosomal RNA coincubated with the detergent-treated virions showed no degradation, indicating the absence of nucleolytic activity outside virions. However, this activity was not specific for the viral RNA since ribosomal RNA coincubated with pH-disrupted virions was also degraded (data not shown). The same result can be obtained using CsCl repurified virions.

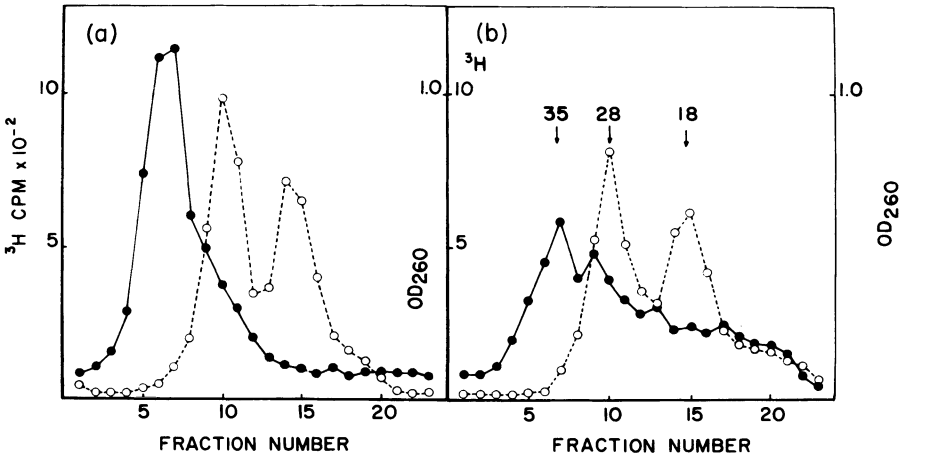

FIGURE 3. Sedimentation analysis of ^{3}H-FMDV RNA extracted from Sarkosyl treated virions that were incubated at 0°and 37°, respectively, together with rat liver ribosomal RNA. RNA samples were layered on 11 ml, 15-30% (wt/vol) linear sucrose gradients made in NET buffer containing 0.5% SDS and centrifuged (Beckman SW41 rotor, 40,000 rpm, 4.5 hr, 20°. (●——●) ^{3}H-FMDV RNA; (o——o) unlabeled 28S and 18S RNAs; RNA from virions incubated 15 hr at 0° and pH 7.4 (a) or at 37° (b). Reprinted with permission of Springer-Verlag (reference 31).

Based on this result, we concluded that there is a nucleolytic activity tightly associated with highly purified virions, probably located inside the viral particles and we think that it could be the enzyme responsible for the genomic FMDV RNA breakdown. Since we have demonstrated that the viral capsid is not permeable to a commercial RNase (32), it is probable that this enzyme is incorporated into virions during the intracellular

virus assembly process and could be considered a constitutive component of the virions. Under the precedent conditions, the enzymatic activity was very low and long incubation periods were required to detect an appreciable degradation of the genomic RNA (32); hence we decided to test whether or not these conditions could be improved. As can be seen in Fig. 4, by addition of 5 mM of Mg^{2+} to the incubation medium and by increasing the pH only one unit (Fig. 4 B) the enzymatic activity is dramatically enhanced since the pattern of RNA degradation after 1 hr incubation under these conditions was similar to that obtained after 20 hr incubation in NET buffer (Fig 4A). Ca^{2+} can also be used as divalent cation with the same result (not shown). The optimal concentration of both ions is around 5 mM. Apparently pHs higher than 8.5 enhance even more the enzymatic activity; however, these conditions promote also some alkaline degradation of the RNA (not shown).

The RNA degradation apparently correlates with a loss of the infectivity since, as shown by Brown et al (19), purified virions lose one log of the infectivity each 8 hr incubation at 37° in phosphate buffer (immunogenic properties remained intact upon this treatment). Hence we decided to try the inactivation of FMDV by activation of the enzyme. However, conditions showed in Fig. 4 appeared not to be so efficient when applied to whole particles. Therefore, we decided to look for conditions capable of activating the enzyme inside non-disrupted virions.

Ward had reported (33) that poliovirus inactivation can be promoted by incubating the virus in 0.5 M NH_4Cl (pH 9.5); this inactivation was correlated with the breakdown of viral RNA. However, the author did not provide a definite explanation for this observation and several hypotheses were given. Thinking that the poliovirus inactivation could be mediated by a polio—associated enzyme similar to the FMDV one, we decided to test the effect of these conditions on FMDV enzyme. Tentative experiments in this direction (not shown) demonstrated that these conditions strongly activate the enzymatic activity of FMDV. Encouraged by these results, we began to test systematically the effect of pH and higher concentrations of NH_4^+ (and related ions) on the enzymatic activity. Fig. 5 shows the effect of pH on the enzyme activity; an important increase of activity was obtained by increasing the pH only one unit. Viral naked RNA was not affected upon incubation at the highest pH used (8.5); pH values lower than 7.5 were not tested because of the known lability of FMDV to pHs under 7.0 (1, 3). On

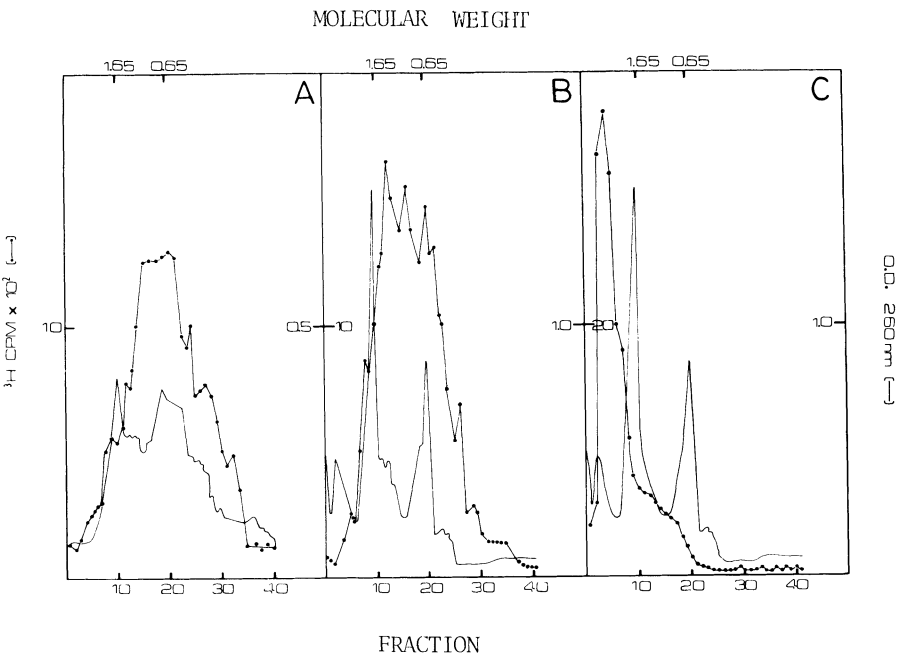

FIGURE 4. Three µg of unlabeled purified virions (free of external nucleases) were heated for 3 min at 60° in 20 µl of NET buffer-10% glycerol and incubated with 50 µg of unlabeled BHK cell cytoplasmic RNA plus ^{3}H-uridine-labeled viral 35S RNA. Samples were diluted to 1 ml with NET buffer and incubated for 20 hr at 37°(A) or diluted to 1 ml with 100 mM Tris-HCl; 100 mM KCl; 5 mM $MgCl_2$, pH 8.5, and incubated for 1 hr at 37°(B). ^{3}H-uridine viral RNA plus BHK cell cytoplasmic RNA were also incubated for 20 hr at 37° (control) (C). After incubation, samples were brought to 1% SDS - 20 mM EDTA, heated for 3 min at 60° and ethanol precipitated. The pellets were resuspended in electrophoresis buffer (28) and analysed on 2.7% acrylamide cylindrical gels. (———) A_{260nm} of unlabeled BHK cells RNA. (●———●) ^{3}H-FMDV RNA.

---

the other hand, pH values higher than 8.5 promoted some alkaline degradation on naked RNA controls. When Tris-HCl (100 mM final concentration) was used instead of NaOH, a similar activation of the enzyme was obtained. However, in this case, the RNA degradation was more extensive that that obtained when NaOH was used, indicating a possible potentiating effect of the Tris-buffer on the enzymatic activity.

The effect of different types of cations and ionic concentrations on the enzymatic activity was also studied. Fig. 6 shows the effect of increasing the NaH_4Cl concentration on viral RNA degradation, indicating that RNA degradation increases proportionally to the ammonium concentration, reaching a maximum between 0.50 and 0.75 M NH_4Cl (Fig. 6 C and 6 D).

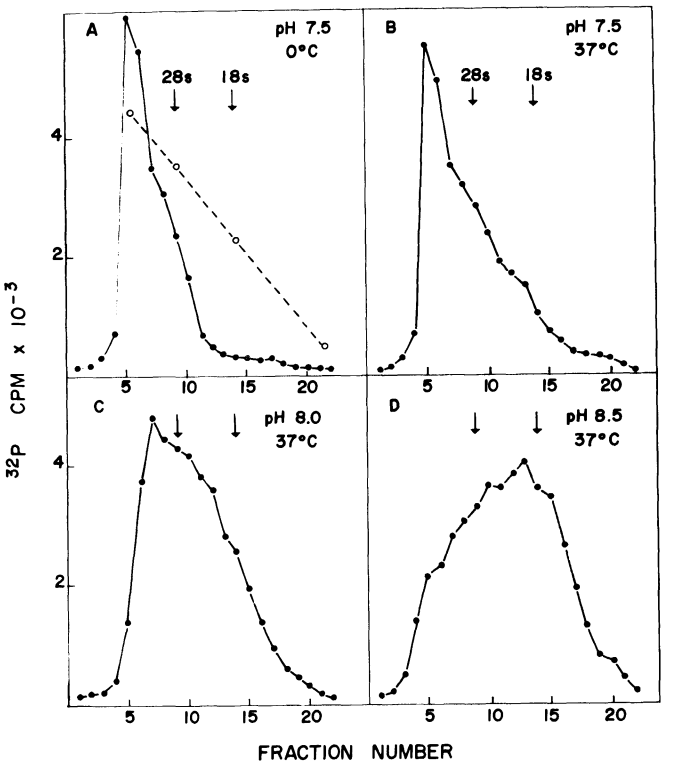

FIGURE 5. Effect of pH on "in situ" viral RNA degradation. ^{32}P-labeled viral suspensions in Eagle's medium were brought to the desired pH by addition of 0.1 M NaOH and incubated for 4 hr at: A) pH 7.5, 0°; B) pH 7.5, 37°; C) pH 8.0, 37°and D) pH 8.5, 37°. After incubation the virions were purified, the RNA extracted twice with phenol-chloroform and analysed in 17 ml 10-30% (wt/vol) sucrose gradient made in 1% NET-SDS buffer, pH 7.4 (Beckman SW 27.1, 17 hr, 22,000 rpm at 20°). Arrows indicate the positions of ribosomal RNAs.

A similar effect was observed at pHs 7.5 and 8.0. By using K^+ or Cs^+ ions instead of NH_4^+ , identical results were obtained; NaCl did not promote

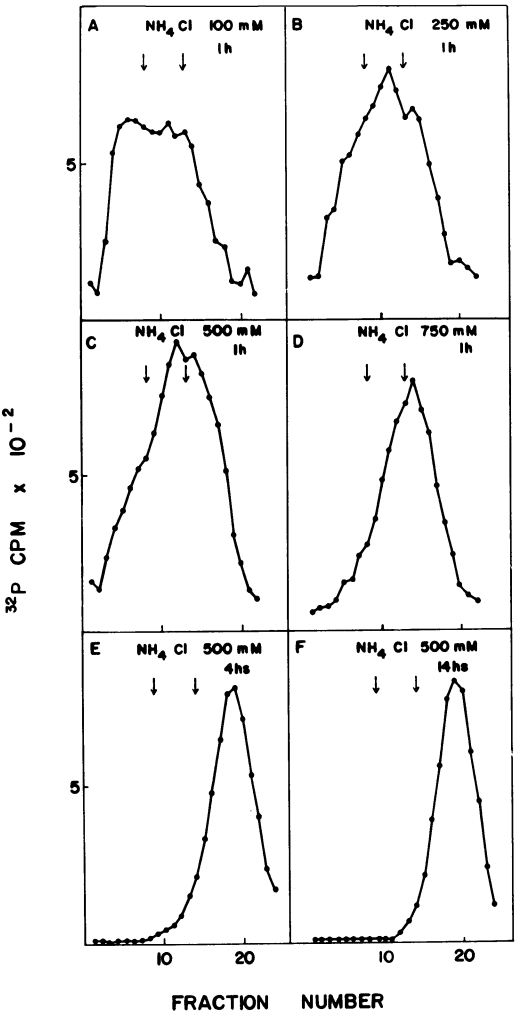

FIGURE 6. Effect of ionic concentration on "in situ" viral RNA degradation kinetics. ^{32}P-labeled viral suspensions in Eagle's medium were incubated for 1 hr at 37°in 0.1 M Tris-HCl, pH 8.5 plus NH_4 Cl: A) 0.10 M; B) 0.25 M; C) 0.50 M; D) 0.75 M or at 37°in 0.10 M Tris-HCl, pH 8.5 for E) 4 hr; F) 14 hr. After incubation the virions were purified and the RNA analysed as in Fig.1 Arrows indicates position of 28 and 18S ribosomal RNAs.

activation of the enzyme. It was argued that selective penetration of unionised ammonia into the virions could increase their internal pH promoting alkaline degradation of the RNA (33). However, this is unlikely since the same effect is caused by monovalent cations such as Cs^+ and K^+ (not shown) which are not capable of increasing the pH inside the virions.

Similar results were obtained incubating highly purified virions (free of any external nuclease contamination, thus ruling out the possibility of RNA degradation by external nucleases). The combined effect of higher pH and higher ionic strength on the nuclease activity produced a marked enhancement of the RNA degradation when compared with that obtained with our previous conditions of incubation (see Fig. 3 B). Under the present conditions the viral RNA is extensively degraded after incubation (4 hr) (Fig. 6 E) and the extent of degradation is greater than that obtained after 50 hr incubation in NET buffer (32). The degradation

profile observed in Fig. 6 E, seems to be maximal, since no further degradation is observed after longer incubation periods (Fig. 6 F).

The sedimentation coefficient and the amount of 140S particles of the incubated samples are the same as in non-incubated controls. Moreover, electrophoretic patterns of their viral capsid proteins are identical to those derived from non-inactivated preparations (not shown).

To see whether the breakdown of the genomic RNA correlates with a loss of the viral infectivity we determined infectivity (as pfu/ml) in viral samples at different times of incubation with NH_4Cl.

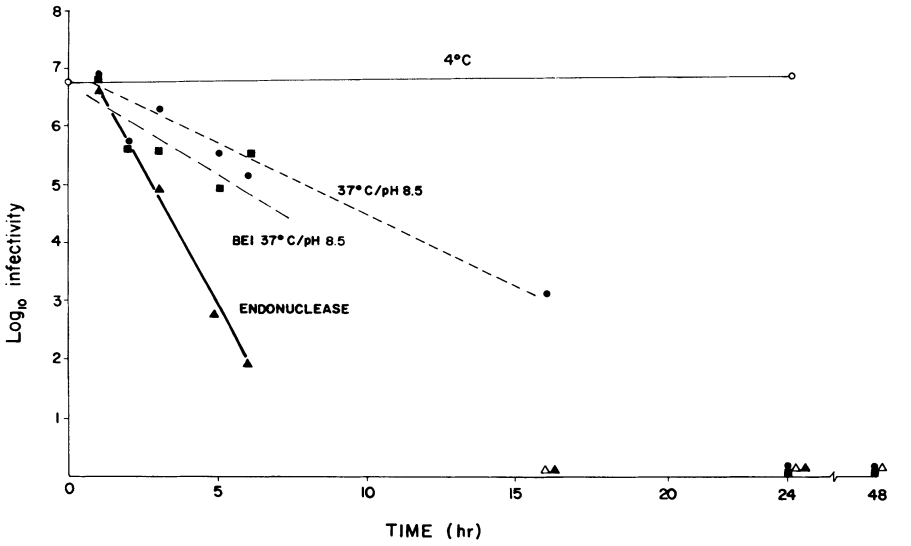

FIGURE 7. Rate of FMDV inactivation under different conditions. FMDV grown in suspensions of BHK cells (1.5 x 10^5 cells/ml) at 37° for 24 hr was harvested after clarification by centrifugation (1500 g for 3 min). The infective virus suspensions were divided into aliquots which were treated as follows: 1) Incubated at 4° for 48 hr. 2) pH adjusted to 8.5 with Na_2CO_3 and incubated for 48 hr at 37°. 3) As 2) above and a solution of bromoethyleneamine added to form BEI. The virus suspension was then incubated at 37°. A second addition of the activating agent was made after 24 hr and the sample kept at 37°for a further 24 hr. 4) pH adjusted to 8.5 by addition of 0.1 vol of 1.1 M Trizma base, 5.5 M NH_4Cl and incubated at 37° for 48 hr. Samples were taken at suitable intervals (0, 1, 2, 3, 4, 5, 6, 16, 22, 24, 48 hr) for infectivity assays.

Fig. 7 shows the kinetics of FMDV inactivation performed by incubation with NH_4Cl or binary ethyleneamine (BEI) which is extensively used as an inactivating agent in FMD vaccine production (34). NH_4Cl inactivation follows first order kinetics and the rate of inactivation is faster than that achieved with BEI. The inactivation rate is slower when the incubation is at 37°, pH 8.5 (Na_2CO_3).

Similar results were obtained with a large number of representative FMDV vaccine strains (not shown). Innocuity test in suckling mice and plaque assays showed that there was not residual infectivity left after 20 hr of inactivation. We are at present studying the feasibility of this inactivation process for industrial use in the production of FMD vaccines.

PART B

Antigenic variation of FMDV is one of the most important causes contributing to the spread of the disease. Failures in the vaccination programs often originate in the fact that new strains are continuously produced while the virus is circulating in the field. These new viruses are the result of two well defined events: 1) the low fidelity of the viral RNA replication (20), and 2) the selective pressure existing by suboptimal levels of neutralizing antibodies in the vaccinated populations. Genetic recombination, which recently has been shown to occur during FMDV replication (21) should also contribute to the virus variability.

There are 7 serotypes of FMDV, the European types O, A and C also present in South America; the South African Territories types SAT_1, SAT_2 and SAT_3, and the Asiatic type $ASIA_1$. Sixty-seven subtypes and hundreds of strains are also recognized. However, subtypification of the virus has been discontinued and at present is only used for diagnostic purposes during outbreaks of the disease.

Identification of viruses is basically performed by complement fixation and recently also by RIA and ELISA tests. The use of monoclonal antibodies for the precise strain identification is also under rapid development. On the other hand, the availability of biochemical procedures for diagnostic purposes opened new and interesting possibilities in this field. Several techniques for the analysis of viral nucleic acids and proteins are being increasingly used for the rapid identification of virus isolates. The combination of these techniques with the serological procedures provides the epidemiologists with an arsenal of new methods to study very precisely

the behaviour, evolution and fate of viruses in the field, after different manipulations in the laboratory and during vaccine production. Among the different procedures, oligonucleotide fingerprinting has been the method of choice for precise identification of a large number of varieties of RNA viruses (35, 36). In the case of FMDV, this method also proved to be very valuable for the identification of virus strains causing outbreaks of the disease in Europe (37) and in South America (38). For that reason, in the second part of this chapter we will present results obtained in our laboratory using RNase T_1 fingerprinting. The data will be specially focused on the practical value of this procedure for the identification of FMDV strains with epidemiological relevance, and in the changes in size of the poly (C) tract of the viral genome by passages in different hosts.

a) Development of a rapid procedure for RNase T_1 fingerprinting. The usual procedure used to make a fingerprint of a viral RNA is to label the genome with ^{32}P orthophosphoric acid during replication, purify the virus and extract its RNA. The purified RNA is then digested with RNase T_1 and the resulting oligonucleotides resolved in two-dimensional polyacrylamide gels (35). This procedure is time consuming and dangerous if routinely used since it involves the isolation of highly labeled viral particles on sucrose or CsCl gradients. In addition, many strains of FMDV are difficult if not impossible to purify due to the intrinsic lability of the virions and their high sensitivity to detergent. On the other hand, another peculiarity of FMDV is that no more than 20% of the 35S induced RNA is incorporated into the mature particles (39) and there are strains of FMDV in which the assembly of virions is also very inneficient resulting in very poor yields of labeled 140S particles. For these reasons it was decided to investigate the possibility of using induced instead of virion RNA for RNase T_1 fingerprinting. This part of the work was performed together with F. Brown and co-workers at the Biochemical Department of The Animal Virus Research Institute (AVRI), Pirbright, England (40). The experiments were performed with the FMDV serotype A subtype 10 strain 61 (A 61). Four monolayers of BHK cells (app. 10^8 cell per bottle) were infected with the virus A 61 at a multiplicity between 10-100. Three of them were harvested at 2-3 and 4 hr postinfection (pi) whereas the 4th one was harvested when a complete cytopathic effect was achieved (app. between 5 and 6 hr pi). The induced 35S RNA was prepared from the cytoplasm of the BHK cells of the first three cultures (40) and analysed on sucrose gradients. It was found that

the amount of 35S was maximum between 3 and 4 hr pi. Radio-labeled virion RNA was extracted from purified virions recovered from the 4th culture and analysed in parallel with the three samples of induced RNA. It was found that at 3 or 4 hr pi., there was 40 times more 35S induced RNA than 35S virions RNA recovered from virus particles. The fingerprints obtained from the induced 35S RNA and the virion RNA were indistinguishable even after performing the electrophoresis of their T_1 hydrolysates (40).

Additional experiments demonstrated that it was not necessary to gradient purify the induced RNA. Total cytoplasmic RNA extracted from infected cells which was labeled in the presence of high doses of actinomycin D gave identical fingerprints to the RNA purified (on gradients or extracted from purified virus) (40). Free phosphorus and small nucleotides which are extracted with the 35S RNA can be easily eliminated by one or two precipitations of the induced RNA with 2.2 M LiCl. Using this procedure, it is possible to obtain a T_1 hydrolysate of RNA of any given FMDV strain within a few hours. It should be noted that the titer of the virus seed was such that a complete cytopathic effect was produced in about 10-12 hr. In addition we used 0.3 mm thick one-dimensional polyacrylamide gels to separate the RNase T_1 oligonucleotides (40). This type of gel was found to be suitable for comparative analysis of different series of viruses. This procedure requires the use of very low levels of counts (between 60,000 to 100,000) to obtain a result within a few hours. An additional advantage of one-dimensional gels is the fact that the poly (C) tract of the viral RNA can be directly analysed whereas this is very difficult on the regular two-dimensional gels. The combined use of the induced RNA with the separation of their RNase T_1 oligonucleotides on one-dimensional gels allows the rapid screening of several different FMDV strains in about 36-40 hr. However, viruses which elicited identical one-dimensional maps, were shown to sometimes have only a few spots of difference when analysed on two-dimensional gels. For this reason, one -dimensional gels can be used for the rapid screening of FMDV strains but the precise identity between two similar isolates can be only established on two-dimensional gels.

b) Analysis of serologically related FMDV strains by RNase T_1 maps. At the end of 1975 and during 1976, in the south of Brasil (State of Rio Grande do Sul) there was an epidemic of FMDV caused by a virus belonging to the serotype A and designated A Bage Brasil/76. This virus was also

active in the field during 1977 and caused outbreaks of the disease in Uruguay and Argentina. In 1979, new viruses appeared in that region and they were designated A Brasil/79 and A Argentina/79. These viruses, as judging by the complement fixation tests, were closely related to the virus A Bage and to the virus A Venceslau which were active in 1976-77 in the states of Santa Catalina and Sao Paulo distant at about 600 and 1000 Km northern from Rio Grande do Sul. In a comparative study with the Pan American Foot-and-Mouth Disease Center at Rio de Janeiro, Brasil, we have analysed these viruses by T_1 maps on one- and two-dimensional gels. Fig. 8 shows the pattern of bands elicited by the RNase T_1 hydrolysates of the viruses A Venceslau, A Bage, A Arg/79 and A Brasil/79 analysed on one-dimensional gels. As can be seen, viruses A/Venceslau and A/Bage (tracts 1 and 3) can be easily distinguished. However, several bands are shared by both viruses indicating that there is an important degree of homology between their genomes. Homologies and differences between them are more easily determined by analysis of the co-electrophoresis of mixtures of their RNase T_1 hydrolysates (tract 2) (40). We have also determined the presence of homologies of the genomes of viruses A/Bage and A/Venceslau with respect to both A/79 viruses (tracts 3-4-5 and 7-8-9) respectively. In this analysis it was also observed that the pattern of bands of viruses A Arg/79 and A Brasil/79 as well as their mixtures were indistinguishable from each other (tracts 5-7 and 6 respectively), suggesting that both viruses are isolates belonging to the same virus strain. The high degree of homology between both viruses was further confirmed by analysis of their RNase T_1 oligonucleotides on two-dimensional gels (fingerprints) where only one spot difference was found between these viruses (see below).

Analysis of the poly (C) tracts of these viruses and their mixtures indicated that viruses A/Bage and A/Venceslau shared poly (C) tracts of identical length (Fig.8, tracts 1-2 and 3). As shown in tracts 5-6 and 7, the length of the poly (C) tract of viruses A Arg/79 and A Brasil/79 are identical but they are clearly shorter than the homopolymers of viruses A/Bage and A/Venceslau.

Analysis of the viruses by two-dimensional gels (fingerprints) also shows that each virus elicited a particular pattern of spots by which they can be clearly identified (Fig. 9). However, the co-electrophoresis (not shown) of their hydrolysates confirmed the presence of an important homology among them.

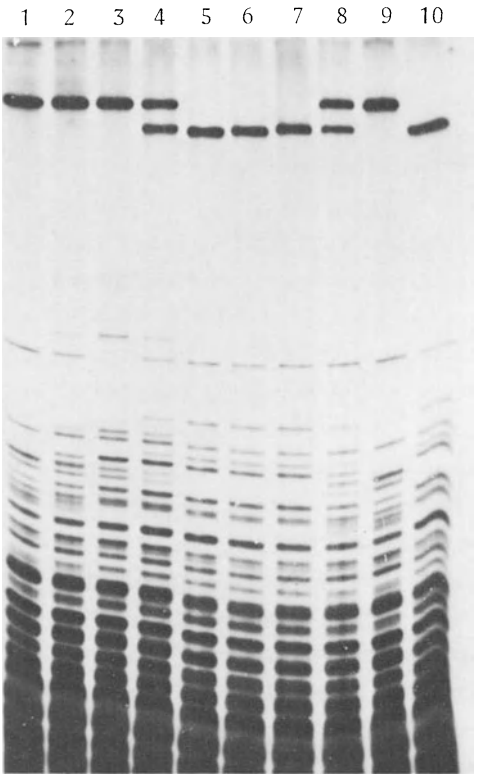

FIGURE 8: RNase T_1 one-dimensional maps of different FMDV strains. 1. A/Venceslau, 3. A/Bage, 5. A Arg/79, 7. A Brasil/79, 9. A/Venceslau, 10. A Arg/79 2. A/Venceslau + A/Bage, 4. A Arg/79 + A/Bage, 6. A Arg/79 + A Brasil/79 8. A/Venceslau + A Brasil/79.

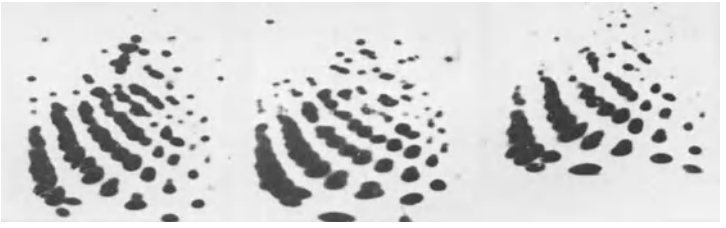

FIGURE 9. RNase T_1 two-dimensional fingerprints of a) A/Bage b) A/Venceslau; c) A Arg/79.

In this regard it should be pointed out that virus A Arg/79 and A Brasil/79 shared all spots except one (23 nucleotides compared); viruses A/Bage and A/Venceslau shared 11 spots (18 nucleotides compared); viruses A/Bage and A Arg/79 had 12 common spots (21 nucleotides compared) and

viruses A/Venceslau and A Arg/79 shared 9 spots (20 nucleotides compared).

These results, taken together with the epidemiological and serological data available, suggest that viruses A/79 were originated from virus A/Bage (Brasil/76). Analysis of some intermediate strains epidemiologically related to this group of viruses (E. Meo Guzman et al, in prep.) are in agreement with this hypothesis. Such evolution of the viruses in the field provides an excellent model to study the molecular basis of antigenic drift and shift of FMDV and which are the protein domains more frequently involved in these antigenic changes.

c) Host-dependent modification of the poly (C) tract of FMDV RNA. It is very well known that FMDV RNA has a polyribocytidilic acid (poly (C)) tract internally located at about 400 nucleotides from its 5' end (14). The length of this homopolymer is very variable (between 80 to 200 residues long) among different virus isolates. The origin of this length variation is at present unknown. However, Harris and Brown (22) reported that the poly (C) tract of an attenuated SAT-1 FMDV strain was considerably shorter than the original virus from which it was derived. We have also found that the poly (C) tracts of 4 different attenuated strains of FMDV were shortened with respect to their original parents (Costa Giomi et al, in prep.). However, in two of those strains, the poly (C) tract was only slightly shortened. Since the attenuation process involves a large number of passages in different nonsusceptible hosts, we thought that the shortening of the poly (C) could be dependent on the host in which the virus was attenuated rather than on its intrinsic degree of attenuation.

In order to study the behaviour of the poly (C) tract of FMDV-RNA in different samples of the FMDV type C_3 strain Resende, one virus was attenuated by 42 passages in 14 days-old chick embryos. The other sample was the original virus from which the attenuated strain was derived (the virus was obtained from the Pan American Foot-and-Mouth Disease Center). Inoculums of both viruses were divided into two equal aliquots and used to infect two different cell lines: BHK 21 clone 13 cells and a swine derived cell line: IB-RS 2 cells. The induced FMDV-RNA was extracted from the four cultures, diggested with RNase T_1 and the resulting oligonucleotides analysed on one-dimensional gels (40). As can be seen in Fig. 10, the poly (C) tract of the attenuated strain was considerably shorter than that of the original virus, both in BHK cells (tracts 2 and 4) and in IB-RS cells (tracts 1 and 3).

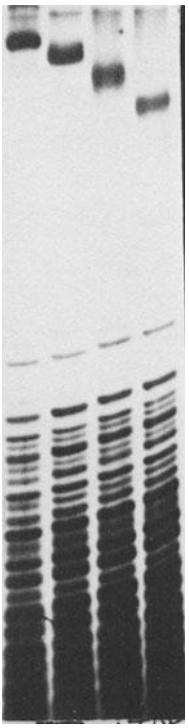

FIGURE 10. one-dimensional RNase T_1 maps of virus replicating in different host cells.
1. C_3 wild type in IB-RS cells
2. C_3 wild type in BHK cells
3. C_3 attenuated (chick embryos) in IB-RS cells. 4. C_3 attenuated (chick embryos) in BHK cells.

However, the poly (C) tract of the original virus and that of the attenuated strains, were enlarged when the viruses were replicated in IB-RS-2 cells. Compare tracts 1 and 2 and 3 and 4 respectively. Preliminary results indicate that the changes in size of the poly (C) tract are due to an increase in the number of the C residues and not to changes affecting the flanking sequences of the homopolymer (Costa Giomi et al, in prep.). It could be also pointed out that when the viruses recovered from IB-RS-2 cells were again passaged through BHK cells, the poly (C) tract recovers its original size that is to say that replication in BHK promotes the shortening of the homopolymer. Similar results were obtained in in primary bovine kidney cells (Costa Giomi et al, in prep.).

These observations are really surprising since these changes in the size of the poly (C) tract are produced in only one passage of the virus in the new host and it needs apparently just one single replication cycle to occur. We do not know at present which is the mechanism by which the poly (C) tract changes its size in different hosts. Nevertheless, these changes are likely to be produced sometime during transcription of the RNA in the different hosts. Hence it is possible that host-dependent factor(s) could be involved in the replication of FMDV RNA as it was postulated for poliovirus-infected Hela cells (41). This idea is further supported by the observation that FMDV strains which are attenuated for cattle or cattle cells growing "in vitro", replicate with similar efficiency as the corresponding wild

type strains in other hosts, such as: swine, swine-derived cells, or BHK cells (P. Auge de Mello, personal communication).

On the other hand, we do not know at present what effect(s) the changes in size of the poly (C) tract of FMDV RNA could have on the biological properties of the virus.

It is very well known that FMDV has a wide host range; however, the physicochemical basis for this behavior is at present unknown. For this reason, the reported changes in size of the poly (C) tract of the FMDV RNA in different hosts, together with the differential suceptibility of cell lines of different hosts to attenuated FMDV strains, could provide an excellent model to ascertain the molecular basis of the virus-cell interactions.

REFERENCES

1. Bachrach, H.L. Beltsville Symposia in Agricultural Research. I Virology in Agriculture. (Ed. J.A. Romberger). Allanheld, Osmun & Co., Montclair, 1977, pp. 3-32.
2. De las Carreras, A. The XI Pan American Meeting on the Control of Foot-and-Mouth Disease Virus and Other Epizooties. Washington, 1978.
3. Matthews, R.E.F. Intervirology 12: 129-296, 1979.
4. Strohmaier, K. and Adam, K.H. Zbl. Vet. Med. B. 23: 483-506, 1976.
5. Vasquez, C., Denoya, C.D., La Torre, J.L. and Palma, E.L. Virology 97: 195-200, 1979.
6. Sangar, D.V. J. Gen. Virol. 45: 1-13, 1979.
7. Sangar, D.V., Rowlands, D.J., Cavanagh, H.D. and Brown, F. J. Gen. Virol. 31: 35-46, 1976.
8. Laporte, J., Grosclaude, J., Wantyghem, J., Bernard, S. and Rouze, P. Compte Rendu Hebdomaire des Seances de L'Academie des Sci. Paris, Serie D 276: 3399-3401, 1973.
9. Bachrach, H.L., Moore, D.M., McKercher, P.D. and Polatnick, J. Immunol. 115: 1636-1641, 1975.
10. Cavanagh, D., Sangar, D.V., Rowlands, D.J. and Brown, F. J. Gen. Virol. 35: 149-158, 1977.
11. Brown, F. and Smale, C.J. J. Gen. Virol. 7: 115-127, 1970.
12. Talbot, P., Rowlands, D.J., Burroughs, N.J., Sangar, D.V. and Brown, F. J. Gen. Virol. 19: 369-380, 1973.
13. Baxt, B., Grubman, M. and Bachrach, H.L. Virology 98: 480-483, 1979.
14. Brown, F., Newman, J., Scott, J., Porter, A., Frisky, D., Newton, C., Carey, N. and Fellner, P. Nature 251: 342-344, 1974.
15. Nomoto, A., Lee, Y.F., Wimmer, E. Proc. Natl. Acad. Sci. USA 73: 375-380, 1976.
16. Hewlett, M.J., Rose, J.K. and Baltimore, D. Proc. Natl. Acad. Sci. USA 73: 327-330, 1976.
17. Sangar, D.V., Rowlands, D.J. and Harris, T.J.R. and Brown, F. Nature 268: 648-650, 1977.

18. Dubra, M.S., La Torre, J.L., Scodeller, E.A. and Vasquez, C. Virology 116: 349-353, 1982.
19. Brown, F. and Wild, T.F. Biochem. Biophys. Acta 119: 301-303, 1966.
20. Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, D. and VandePol, S. Science 215: 1577-1585, 1982.
21. King, A.M.Q., McCahon, D., Slade, W.R. and Newman, J.W.I. Cell 29: 921-928, 1982.
22. Harris, T.J.R. and Brown, F. J. Gen. Virol. 34: 87-105, 1977.
23. Chaterjee, N.K., Bachrach, H.L. and Polatnick, J. Virology 69: 369-377, 1976.
24. Arlinghaus, R.B., Polatnick, J. and Vande Woude, G.F. Virology 30: 541-550, 1966.
25. Breese, S.S.Jr. J. Gen. Virol. 31: 1-8, 1976.
26. Brown, F., Cartwright, B. and Stewart, D.L. J. Gen. Microbiol. 31: 179-186, 1963.
27. Matheka, H.D., Trautman, R. and Bachrach, H.L. Biochem. Biophys. 121: 325-330, 1967.
28. Denoya, C.D., Scodeller, E.A., Gimenez, B.H., Vasquez, C. and La Torre, J.L. Virology 84: 230-235, 1978.
29. Gauntt, C.J. and Griffith, M.M. J. Virol. 13: 762-764, 1974.
30. Dimmock, N.J. Virology 31: 338-353, 1967.
31. Denoya, C.D., Scodeller, E.A., Vasquez, C. and La Torre, J.L. Arch. Virol. 57: 153-159, 1978.
32. Denoya, C.D., Scodeller, E.A., La Torre, J.L. and Vasquez, C. Virology 89: 67-74, 1978.
33. Ward, R.L. J. Virol. 26: 299-305, 1978.
34. Bahnemann, H.G. Arch. Virol. 47: 47-56, 1975.
35. Clewley, J.P. and Bishop, D.H.L. In New Developments in Practical Virology. Liss, A.R. Inc. New York, 1982.
36. Bishop, H.L. Current Topics in Microbiology and Immunology 104: 259-271, 1983.
37. King, A.M.Q., Underwood, B.O., McCahon, D., Newman, J.W.I. and Brown, F. Nature 293: 479-483, 1981.
38. Wild, T.F. and Brown, F. Biochem. J. 107: 395-401, 1968.
39. Costa Giomi, M.P., Durini, L., Bergmann, I.E., Mazzuca, O., Rau, M.E., Fernández, G., Parisi, J.M. and La Torre, J.L. Bull. of. the 16th Conference of the FMDV Commission (1): 263-271, 1982.
40. La Torre, J.L., Underwood, B.O., Lebendiker, M.A., Gorman, B.M. and Brown, F. Infection and Immunity 36:142-147, 1982.
41. Dasgupta, A., Zabel, P. and Baltimore, D. Cell 19: 423-429, 1980.

ACKNOWLEDGMENTS

The authors wish to thank O. Basarab, I.E. Bergmann and M.P. Costa Giomi for allowing the use of unpublished results and to Graciela Teglia and Silvia Rojana for clerical work.

15

THE EXPRESSION AND ORGANIZATION OF THE ALPHAVIRUS GENOME

H. Söderlund, K. Takkinen, A. Jalanko and N. Kalkkinen

1. SUMMARY

The alphaviruses have a positive strand RNA genome about 12,000 nucleotides in length. About 7,500 nucleotides from the 5' end codes for the synthesis of the four viral nonstructural proteins. These are synthesized principally as a polyprotein with an apparent molecular weight of 290,000 from which the final products are formed by proteolytic processing. The cleavage sites can be identified by direct N-terminal sequence analysis of the radiolabeled polypeptides. The 3' third of the genome is transcribed into a subgenomic message, the 26S RNA. This codes for another polyprotein which is proteolytically processed to yield the four structural proteins of the virion. The structure of the RNA from defective interfering particles reveal the regions of the genome which are necessary for RNA-replication and encapsidation.

2. INTRODUCTION

The two best studied members of the alphavirus group are Semliki Forest virus (SFV) and Sindbis virus, but additional information is available also on the Eastern, Western and Venezuelan equine encephalitis, Chikungunya and Middelburg viruses. In some aspects these viruses are clearly different but they are still similar enough to allow generalizations concerning major biological properties (1, 2). Some other togaviruses, like rubella virus, seem to have a general replication scheme resembling that of alphaviruses (3), but the flaviviruses may differ in major aspects of their replication strategy (4).

Y. Becker (ed.), VIRAL MESSENGER RNA.

The particle of an alphavirus consists of an icosahedral nucleocapsid, which contains the capsid protein (MW 30,000) and the RNA genome. The genome has the features of a eukaryotic message with a 5' cap and a 3' poly(A) region. It is also translatable in vitro. The nucleocapsid is surrounded by an envelope, composed of a lipid bilayer with two major glycoproteins E1 and E2 (MW 52,000 and 53,000, respectively). A third glycoprotein E3 (MW 10,000) is associated with the envelope of SFV. The corresponding protein is not found in Sindbis virions but is secreted from the infected cells (1, 2).

The genome of the alphavirus is infectious. Consequently, the four non-structural proteins must be the first virus-specific products early in infection. Two to three hours after infection viral RNA and protein synthesis can be detected. Five hours post-infection a massive synthesis of both viral genomes and the subgenomic 26S RNA message takes place. The protein synthesis machinery is primarily devoted to the synthesis of viral structural proteins while host cell protein synthesis becomes inhibited (1, 2). The newly formed capsid protein binds to 60S ribosomes being possibly associated with the 28S RNA and is then transferred to the genome RNA in the process of nucleocapsid assembly (5, 6). The envelope proteins become glycosylated and are transported to the plasma membrane by the cellular machinery (2, 7). The virus matures by budding of nucleocapsids through the plasma membrane, a process in which the virions acquire their envelope (1, 2).

During high multiplicity passaging of the virus defective interfering (DI) particles may arise. Structurally the particles appear similar to standard virions, but the full-sized genome is replaced by truncated RNA molecules of varying size. The presence of DI-RNA in the cell causes a severe inhibition of RNA synthesis of standard virus and affects also to some extent the encapsidation (8, 9). The DI RNA has lost the coding capacity of the viral genome, but has by necessity retained signals for replication and encapsidation. DI particles can consequently help us to gain insight into the functions of different genomic regions.

3. RESULTS

3.1. Protein synthesis in infected cells

Alpha-virus infection leads to a strong inhibition of host cell protein synthesis (10). Thus the viral structural proteins can easily be detected from infected cells by gel electrophoresis after labeling with radioactive amino acids.

A salt concentration of 300-350 mM in the cell growth medium inhibits protein synthesis initiation, but allows elongation to proceed (11). Release from this hypertonic block leads to synchronization of the protein synthesis. By this simple method the polyprotein nature and the gene order of the SFV structural proteins could be shown (Fig. 1) (12, 13). Interestingly, the same experiments performed with a flavivirus, suggest that the protein synthesis of these viruses is strikingly different. In this case the structural proteins seemed to initiate separately (4).

The nonstructural (ns) proteins of alphaviruses have been more difficult to analyze. These proteins are mainly produced early in infection, when host-cell inhibition is only partial and the total amount of ns proteins synthesized is much smaller than that of the structural proteins (10). Fortunately, a temperature-sensitive mutant of SFV, ts-1, overproduces the nonstructural proteins and facilitates their analysis, but still radioactive labeling is a necessity for their detection (14, 15). Hypertonic media can also be used to suppress host-cell protein synthesis. The additional use of salt-synchronized proteins synthesis allowed Keränen and Ruohonen (16) to detect the four nonstructural proteins of SFV and to determine their gene order. We call them here ns P1, ns P2, ns P3 and ns P4, the numbering referring to the translation order (c.f. 17).

The synthesis scheme of the nonstructural proteins shows that the ns proteins are synthesized principally as a poly-protein (MW 290,000). This is proteolytically processed to yield two intermediate precursors, ns p155 and ns p135, which in turn are cleaved to the final products ns P1 - ns P2 and ns P3 - ns P4, respectively (Fig. 1). The large precursor ns p290 is apparently processed as nascent to yield the

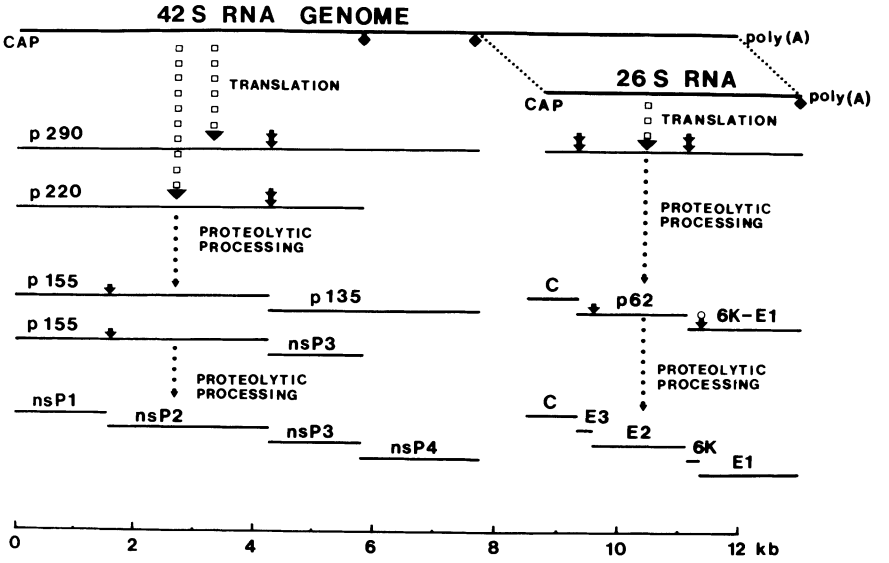

FIGURE 1. Processing scheme of the alphavirus nonstructural and structural precursor polyproteins. The nonstructural precursor proteins ns p290 and ns p220, of which ns p220 may be the predominant precursor (see 3.3) are usually not found in infected cells due to their nascent cleavage to the intermediate proteins. The intermediates ns p155 and ns p135 can in contrast easily be detected. The intact structural precursor protein (the translation product of the subgenomic 26S RNA) is not either found in cells due to cleavage of the nascent polypeptide to C, p62 and 6K-E1. p62 is then cleaved late during virus maturation to form E3 and E2. The 6K protein, which contains the signal for membrane penetration of E1 is cleaved from E1 possibly due to the host cell signal peptidase activity.

◆ translation termination site
↡ cleavage of the nascent polypeptide chain
↓ proteolytic processing site
⚲↓ suggested cleavage by host cell signal peptidase.

intermediates ns p155 and ns p135.

A second primary precursor, ns p220, has also been observed (18, Fig. 1). It may be that this is the predominant precursor which gives rise to ns P1, ns P2 and ns P3 (see 3.3).

Ns P4, the last nonstructural protein, can be generated only from ns p290. Thus, the amount of this protein in infected cells must be lower than the amount of the other ns-proteins. The pulse-chase experiments by Keränen and Ruohonen (16) actually show that ns P4 really is present in lower amounts than the other ones and also have a shorter half life.

3.2. Cloning of the SFV-genome

The cDNA corresponding to the virion RNA of Sindbis and SFV has been synthesized by reverse transcription and subsequently cloned (17, 19, 20). We have used cDNA clones from the SFV 26S RNA region as primers to clone the cDNA corresponding to the nonstructural part of the SFV genome. A first set of cDNA was cloned into the E. coli plasmid pBR322 by G:C tailing. With restriction maps known for these first clones we could then cleave a second set of cDNA with appropriate restriction enzymes to obtain clones which specifically spanned known gaps. No individual clone was longer than 3 kb at this stage, but together they cover the whole region.

To ensure that the clones were true copies of the RNA, we performed a colinearity experiment by hybridizing ^{32}P-labeled viral RNA to the individual clones in conditions which favour DNA:RNA hybridization, whereafter unprotected RNA and DNA was digested with nuclease S1. A clone was considered colinear when the size of the protected RNA was identical to the size of the insert in the clone (21). This test turned out to be crucial, since in several clones a specific deletion in the cDNA was observed. The 5' end of the RNA also gave aberrant cDNA. An interesting fact was that the 3' end of the SFV-RNA is so U-rich that the poly(A) tail apparently loops back and functions as an efficient "self-primer". Thus a significant proportion of the clones mached the 3' end of the genomic-RNA even when primers complementary to its central parts were used.

3.3.Nucleotide sequence of alphavirus genomes

The RNA of Sindbis, SFV and parts of other alpha-viruses have been sequenced either directly from cDNA (22, 23) or after cloning (17, 19, 20). The Sindbis genome is sequenced completely (17), while about 5% of the nucleotide sequence from the SFV genome is unknown at present (19, 20, our data). The overall picture corresponds well to that postulated on the basis of protein data. Flanked by short noncoding regions the genome is divided into two blocks with long open reading frames (Fig. 1). Strikingly it was observed that in Sindbis and Middelburg virus RNA there is a UGA stop codon in frame between the genes for ns P3 and ns P4 (23). This opal stop codon is apparently read through to form the full-sized precursor yielding ns P4 upon cleavage.

Sequence comparison between SFV and Sindbis (and Middelburg) coding sequences reveal regions with significant homology and others devoid of strong similarities (Fig. 2).

3.4.Amino acid sequence of alphavirus proteins

From the nucleotide sequence the primary structure of the proteins can be deduced. However, the proteolytic cleavage sites can be found only by direct protein sequencing. Technically this offered no problems in the case of the structural proteins, which could be isolated in pure form from the virion. Their N-terminal amino acid sequences could then be determined after Edman-degradation and their C-terminal structure after digestion with carboxypeptidases (24, 25, 26). The determination of the amino acid sequences of the ns proteins turned out to be more difficult, since there is little hope to isolate these proteins in a chemically pure form.

To solve this problem we have adopted another strategy. The ns proteins are labeled with one or several radioactive amino acids using high salt treatment to direct most of the label towards the protein in question (see 3.1). The proteins are then isolated by preparative SDS-gel electrophoresis and subjected to automated Edman degradation. The amino acid derivatives from each cycle are then analyzed for radioactivity.

If a single label is used, radioactivity as such gives the positions of that amino acid. With multilabel additional separation of the amino acid phenylthiohydantoin by HPLC is necessary before radioactivity measurement (27). In this way we obtained a partial amino acid sequence, which even with a few positive assignments was enough to unambiquously align the nucleotide sequence with the amino acid sequence (Fig. 3).

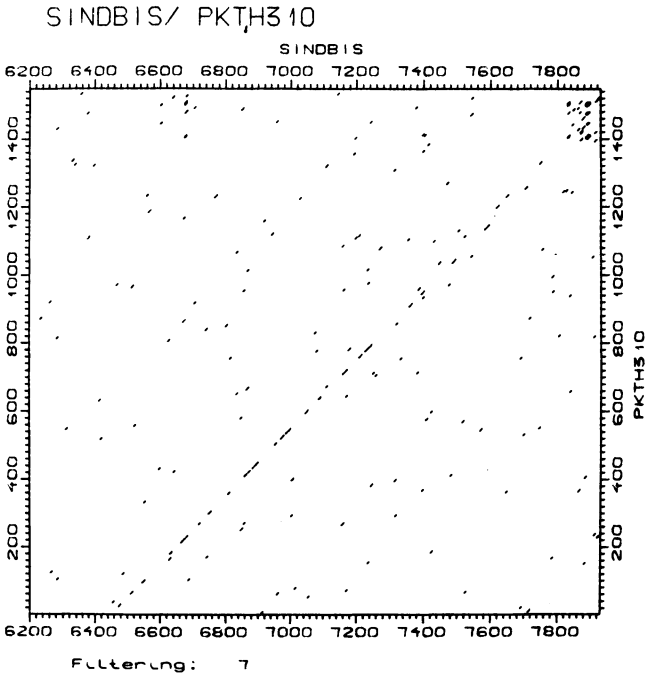

FIGURE 2. Comparison of a part of the nucleotide sequence of SFV and Sindbis. The data shown is the comparison of a SFV cDNA clone pKTH310 with the corresponding region of the Sindbis genome (17). The region extends from the 26S RNA junction (pos. 7600 in Sindbis) towards the 5' end of the genomes. The comparison is a simple matrix where a dot is marked for every position in which a 7-nucleotide-stretch of SFV is found on the Sindbis genome within the relevant region. Homology is indicated by a diagonal against the scattered pattern of random matches.

Unfortunately, we failed to obtain the N-terminal sequence of ns P1 as well as of its precursor ns p155. This may indicate that their N-termini are blocked in one way or another. Thus, we cannot localize the nucleotide sequences coding for their N-termini. However, the AUG codon at nucleotides 86-88 of the SFV genome, which is followed by an alanine codon, is the probable initiator, since it is followed by a long open reading frame, and the initiation dipeptide has been shown to be met-ala (28, 29).

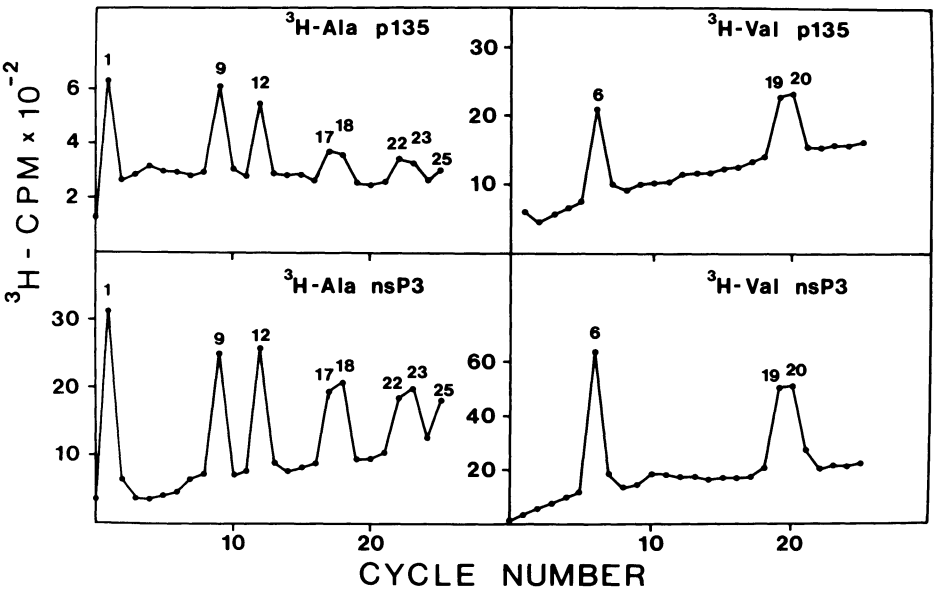

FIGURE 3. Amino-terminal radio-sequence analysis of ns P3 and ns p135. The proteins were labeled in vivo by incorporation of either ^{3}H-alanine or ^{3}H-valine and then purified by preparative SDS-polyacrylamide gel electrophoresis. The results show recovery of ^{3}H-alanine (left panel) and ^{3}H-valine (right panel) from radiochemically pure ns P3 and ns p135 in Edman degradation cycles in a liquid phase sequencer. The obtained results show that ns P3 and ns p135 have identical amino termini. The nucleotide sequence corresponding to these amino termini could unambiquously be localized on the genome at a position 4.2 kb from its 5' terminus.

The N-termini of ns p135 and ns P3 are open for Edman degradations. Using radiochemically pure proteins labeled with ^{3}H-alanine, ^{3}H-valine and ^{3}H-leucine we could show that these proteins have identical N-termini (Fig. 3). Thus there is no additional N-terminal processing of ns P3 after the primary cleavage of ns p290 to ns p155 and ns p135.

The C-terminus of ns P3 is formed in two ways, either by termination at the UGA-codon or by proteolytic cleavage of the precursor ns p135 (Fig. 1). Our preliminary amino acid sequence data suggest that the proteolytic cleavage site does not correspond to the stop codon. Ns P3 formed by cleavage is some amino acids longer than that formed upon termination.

3.5. Specific signals within the SFV genome

In addition to the coding sequences the alphavirus genome should by necessity contain also three other signals. The viral RNA replicases use exclusively viral genome and antigenome RNA as template in vivo. Thus the genome contains signals for replicase binding. Secondly, under in vivo conditions only genome RNA is assembled into nucleocapsids. Consequently, an encapsidation signal is present. Thirdly, distinct subgenomic 26S RNA is synthesized from genome-sized minus-strand templates indicating the presence of an internal initiation signal for RNA synthesis (1).

One way to identify the replicase binding and encapsidation signals is to study the structure of DI RNAs. These RNAs can be considered as deletion mutants, which have lost essential coding functions, but have retained the capacity for replication and encapsidation in the presence of standard virus as a helper. We have cloned and sequenced two DI RNAs. Nucleotide sequence regions, which are present in both RNA species, should contain the essential signals.

Both the 5' and 3' regions of the genome are conserved in the two DI-RNAs (DI301 and DI309; 29, 30). The 5' end may in some cases have an additional extension (31, 32). A striking feature is that the DI-RNA has a central part which is composed of sequence "blocks", which are amplified.

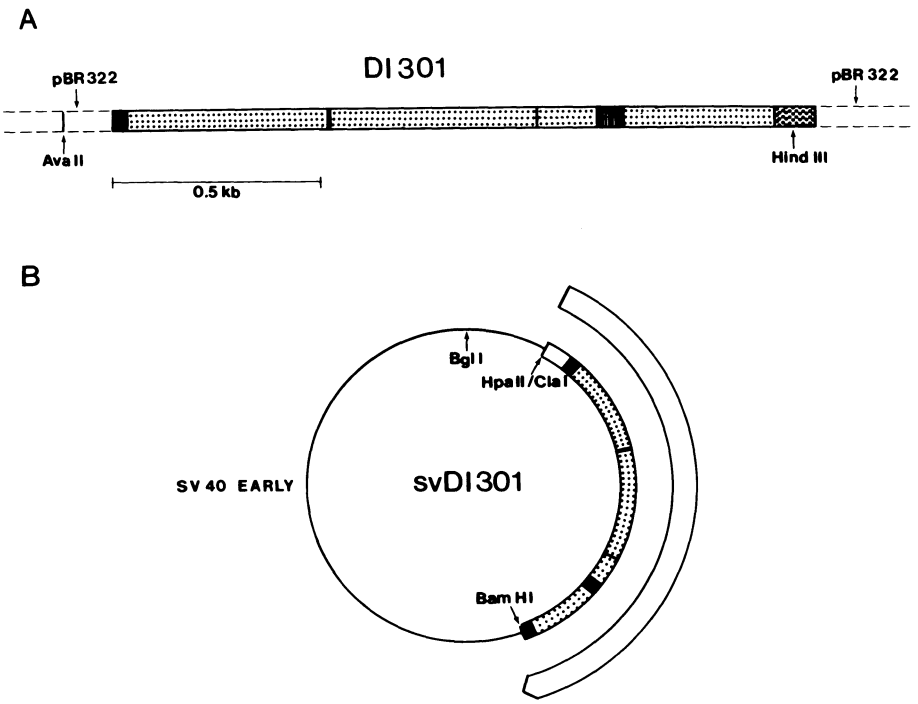

FIGURE 4. Construction of SV40 recombinant (svDI301) expressing the repeated region of a SFV DI-RNA. The cDNA clone (in pBR322) contained the entire DI-molecule except for a few nucleotides at its 5' end (A). For insertion into the late region of SV40 (between the HpaII and BamHI sites) the DI-insert was removed by cleavage within the unique 3' end with HindIII and slightly upstream from the 5' end in the pBR322 region with AvaII. The sites were then altered to BamHI and ClaI, respectively, and introduced into the SV40. A SV40 recombinant virus stock was propagated using a tsA mutant as a helper. Upon infection of monkey kidney cells this mixture expressed the repeated part of the DI-RNA together with flanking parts of the vector (arrow) (B). The origin of the DNA regions are indicated by the shadowing. The stipled area indicates the amplified unit, repeated linearly three times (separated by vertical lines). The third unit contains a small insert (vertical stripes).

The amplification level seems to be increased during passaging of the DI particles (9). Comparison of the two sequenced DI-RNAs of SFV with each other shows that only 518 bases in addition to the termini are in common. Within this region there is a stretch of 45 bases found also in Sindbis virus DI-RNA (33).

The 3' and 5' termini of the genome RNA are the initiation points for the replication of minus and plus-strand RNAs, respectively. It is thus not surprising that these regions are conserved in DI RNA. The amplified internal regions can, however, be either polymerase or capsid protein binding loci. We made an effort to discriminate between these possibilities by studying RNA containing the central part only. To this aim we constructed a SV40 virus DNA with the late region replaced by the central part of DI301 (Fig. 4). By cotransfection of this DNA into CV-1 cells together with the DNA of a temperature-sensitive mutant of SV40, defective in its early region (tsA58) we then obtained by complementation a SV40 recombinant and tsA mixture, which is able to replicate (34). This virus expressed the DI RNA in B-Vero cells as shown by Northern blot analysis (not shown). After superinfection with SFV we then followed the RNA synthesis and virus assembly of SFV. Despite the fact that the number of DI RNA recombinant molecules per cell is rather low, it should be comparable to early passages of natural DI RNA (9) and thus give clear effects. Especially RNA synthesis is strongly inhibited by DI RNA. Our data indicated that recombinant DI RNA did not affect SFV RNA synthesis to any significant extent, but packaging of 42S RNA into virions was inhibited (Fig. 5). We could not find recombinant DI molecules in mature virions, but at present it is not clear whether this RNA cannot be encapsidated or whether our detection system was not sensitive enough. In any case the results show that the amplified nucleotide sequence of the DI RNA interferes with the encapsidation process, suggesting that these sequences are involved in nucleocapsid assembly.

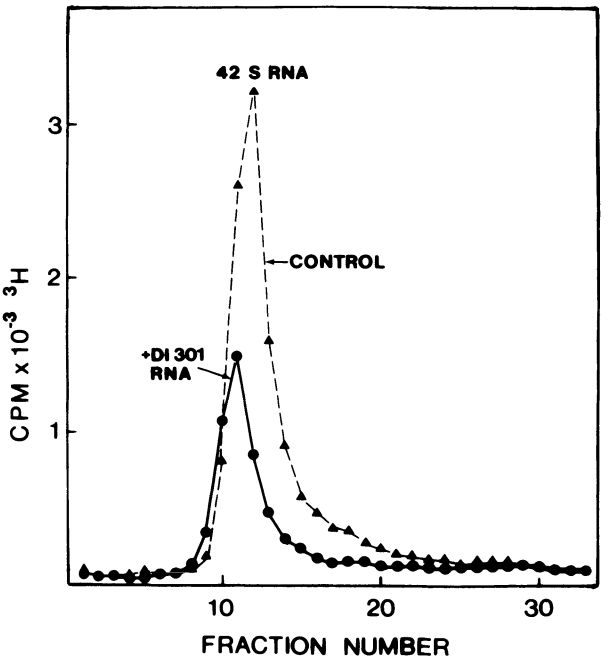

FIGURE 5. Quantitation of virus released from SFV infected B-Vero cells. The cells (8 x 10^5) were infected with svDI301 (●—●) or with wild-type SV40 as a control (▲--▲). 24 h later the cells were superinfected with SFV (moi. = 5), and ^{3}H-uridine label given 2-6 h after the SFV infection. The culture medium was harvested and released virus pelleted through a 30% sucrose cushion. Viral RNA was released by SDS treatment and analyzed on 15-30% sucrose gradients.

4. DISCUSSION

The overall replication strategy of the alphaviruses is well established. The message-sense genome codes for the nonstructural proteins while a subgenomic message codes for the structural proteins. Both sets of proteins are synthesized as polyproteins and the final products are released by proteolytic processing. This processing event is not yet clarified. It is obvious that the proteases involved must have an extraordinarily high substrate specificity. The processing

requires at least 7 proteolytic steps of which possibly one is performed by host cell signalase (2). Whether a single or several, host or virus, protease(s) is/are required for the other cleavages is an open question. There is, however, some evidence suggesting that the capsid protein has autoproteolytic activity, releasing itself from its precursor (35, A. Jalanko, unpublished).

The capsid protein may also play a role in the control of protein synthesis in the infected cells. Van Steeg et al. (36, 37) have shown that ribosomal washes containing capsid protein to some extent discriminates between 26S RNA and other messages in an in vitro translation assay.

Another interesting feature of the viral translation process is the stop codon between ns P3 and ns P4 (23). The read-through necessary for synthesis of ns P4 may well be a control element in the replication process. The major effect would be to create less ns P4 than the 3 other nonstructural proteins. The two suggested forms of ns P3 may of course also have different properties. The mechanism of read-through is at present unknown, but it should be kept in mind that the alphaviruses have an extremely wide host range (1, 2). A specific non-universal, host function appears therefore improbable, assuming that the function of ns P4 is necessary for virus replication in all cell types.

The binding sequences for the RNA polymerase and the encapsidation signal for packaging are unidentified even if the possibilities are narrowed down to a few hundred nucleotides. It will be interesting to see whether foreign RNA can be replicated and packaged if correct elements from the alphavirus genome are joined to them.

REFERENCES

1. Kääriäinen, L. and Söderlund, H. Curr. Top. Microbiol. Immunol. 82: 15-69, 1978.
2. Garoff, H., Kondor-Koch, C. and Riedel, H. Curr. Top. Microbiol. Immunol. 99: 1-50, 1982.
3. Oker-Blom, C., Ulmanen, I., Kääriäinen, L., and Pettersson, R.F. J. Virol. 49: 403-408, 1984.

4. Westaway, E.G. Virology 80: 320-335, 1977.
5. Ulmanen, I., Söderlund, H., and Kääriäinen, L. Virology 99: 265-276, 1979.
6. Ranki, M., Ulmanen, I. and Kääriäinen, L. FEBS Lett. 108: 299-302, 1979.
7. Saraste, J. and Kuismanen, E. Cell. in press.
8. Stollar, V. In: "The togaviruses. Biology, structure and replication." (R.W. Schlesinger, ed.). Academic Press, NY. 1980. pp. 427-457.
9. Kääriäinen, L., Pettersson, R.F., Keränen, S., Lehtovaara, P., Söderlund, H., and Ukkonen, P. Virology 113: 686-697, 1981.
10. Lachmi, B. and Kääriäinen, L. J. Virol. 22: 142-149, 1977.
11. Saborio, J.L., Pong, S.S. and Koch, G. J. Mol. Biol. 85: 195-211, 1974.
12. Clegg, J.C.S. Nature 254: 454-455, 1975.
13. Garoff, H. and Söderlund, H. J. Mol. Biol. 124: 535-549, 1978.
14. Keränen, S. and Kääriäinen, L. J. Virol. 16: 388-396, 1975.
15. Lachmi, B. and Kääriäinen, L. Proc. Natl. Acad. Sci. USA, 73: 1936-1940, 1976.
16. Keränen, S. and Ruohonen, L. J. Virol. 47: 505-515, 1983.
17. Strauss, E.G., Rice, C.M. and Strauss, J.H. Virology 133: 92-110, 1984.
18. Lehtovaara, P., Ulmanen, I., Kääriäinen, L., Keränen, S. and Philipson, L. Eur. J. Biochem. 112: 461-468, 1980.
19. Garoff, H. Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. Proc. Natl. Acad. Sci. USA 77: 6376-6380, 1980.
20. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. and Delius, H. Nature 288: 236-241, 1980.
21. Söderlund, H., Keränen, S., Lehtovaara, P., Palva, I., Pettersson, R.F. and Kääriäinen, L. Nucl. Acids Res. 9: 3403-3417, 1981.
22. Rice, C.M. and Strauss, J.H. J. Mol. Biol. 150: 315-340, 1981.
23. Strauss, E.G., Rice, C.M. and Strauss, J.H. Proc. Natl. Acad. Sci USA 80: 5271-5275, 1983.
24. Kalkkinen, N. FEBS Lett. 115: 163-166, 1980.
25. Kalkkinen, N., Jörnvall, H., Söderlund, H. and Kääriäinen, L. Eur. J. Biochem. 108: 31-37, 1980.
26. Kalkkinen, N., Jörnvall, H. and Kääriäinen, L. FEBS Lett. 126: 33-37, 1981.
27. Kalkkinen, N., Laaksonen, M., Söderlund, H., and Jörnvall, H. Virology 113: 188-195, 1981.
28. Glanville, N., Ranki, M., Morser, J., Kääriäinen, L. and Smith, A.E. Proc. Natl. Acad. Sci USA 73: 3059-3063, 1976.
29. Lehtovaara, P., Söderlund, H., Keränen, S., Pettersson, R.F. and Kääriäinen, L. J. Mol. Biol. 156: 731-748, 1982.
30. Lehtovaara, P., Söderlund, H., Keränen, S., Pettersson, R.F. and Kääriäinen, L. Proc. Natl. Acad. Sci USA 78: 5353-5357, 1981.
31. Pettersson, R.F. Proc. Natl. Acad. Sci. USA 78: 115-119, 1981.
32. Monroe, S.S. and Schlesinger, S. Proc. Natl. Acad. Sci, USA 80: 3279-3283, 1983.

33. Monroe, S.S., Ou, J.-H., Rice, C.M., Schlesinger, S., Strauss, E.G. and Strauss, J.H. J. Virol. 41: 153-162, 1982.
34. Elder, J.T., Spritz, R.A. and Weissmann, S.H. Ann. Rev. Genet. 15: 295-340, 1981.
35. Aliperti, G. and Schlesinger, M.J. Virology 90: 366-369, 1978.
36. Van Steeg, H., Thomas, A., Verbeek, S., Kasperaitis, M., Voorma, H.O. and Benne, R. J. Virol. 38: 728-736, 1981.
37. Van Steeg, H., Kasperaitis, M., Voorma, H.O. and Benne, R. Eur. J. Biochem. 138: 473-478, 1984.

16

A SHORT SEGMENT OF VIRAL DNA IS REQUIRED FOR THE STIMULATION OF TRANSCRIPTION OF MOUSE MAMMARY TUMOR VIRUS GENES BY GLUCOCORTICOID HORMONES

H. Diggelmann, E. Buetti, D. Owen, N. Fasel, J. Firzlaff, A.L. Vessaz

SUMMARY

Transcription of mouse mammary tumor virus (MMTV) DNA is stimulated by glucocorticoid hormones in vivo and in mammary tumor cells in culture. We have shown that cloned exogenous and endogenous proviral DNA is expressed in transfected mouse and mink cells. This expression is hormonally controlled and the regulation takes place on the level of initiation of transcription. The element responsible for the glucocorticoid stimulation was first localized in the long terminal repeat (LTR) region of the viral DNA. To more closely locate the important sequences we constructed a plasmid containing the LTR in front of the coding region of the herpes simplex thymidine kinase gene from which the promoter had been eliminated. Specific portions of the LTR were removed and the deleted molecules were tested for their ability to respond to glucocorticoid hormones after transfection into Ltk$^-$ aprt$^-$ cells. These experiments demonstrated that the DNA sequences between -105 and -204 base pairs upstream from the initiation site of viral transcription are required for glucocorticoid stimulation. The postulated protein coded for by the viral LTR is not involved in the hormone regulation of transcription.

INTRODUCTION

The observation that the expression of mouse mammary tumor virus genes is regulated in cells in culture by glucocorticoids (for review see 1) has made this system particularly attractive to study the mechanism of hormonal control of gene expression. According to the generally proposed model of steroid hormone action the hormone diffuses passively into the cell, binds to the cytoplasmic hormone receptor and the hormone receptor complex is then translocated to the nucleus, where it is

Y. Becker (ed.), VIRAL MESSENGER RNA.

supposed to interact with specific regions of the chromatin to elicit the specific regulation of a set of genes. Pulse-labeling experiments suggested that the increase in the amount of viral products was due to a rapid increase in the rate of viral RNA synthesis (2, 3,4,5) and this effect was shown to be independent of ongoing protein synthesis (6, 7). More recently the number of nascent viral RNA chains in nuclei of L-cells transfected with cloned endogenous MMTV DNA was directly measured in presence and absence of glucocorticoids (8). These data demonstrate that the hormone increases the number of RNA polymerase molecules that are engaged in transcription of the viral genome.

The experiments performed using MMTV infected cells did not allow to determine conclusively if an element within the viral genome was required for an interaction with the glucocorticoid-receptor complex, or if the hormone response could be due solely to cellular sequences in the vicinity of the integrated viral genome. We therefore decided to clone a complete unintegrated circular MMTV provirus of the GR strain (9) and to study its expression after transfection into mouse Ltk$^-$ aprt$^-$ cells (9) and into mink lung epithelial cells (10) in presence or absence of glucocorticoids. We were able to show that the proviral DNA itself contained the sequences required for the hormonal stimulation of transcription and we proceeded to identify these sequences more precisely. At the same time we also sequenced regions of the MMTV DNA which contain important signals for initiation and processing of viral RNA (11, 12).

RESULTS

The cloning of circular unintegrated MMTV DNA of the GR strain in bacteriophage λ has been described (9). The biological activity of the cloned viral DNA was tested in cotransfection experiments after transfer into mouse cells (9) and into mink lung epithelial cells which are devoid of endogenous MMTV related sequences (10). In both cell types the foreign DNA was stably integrated. Viral RNA synthesis was monitored first by dot-blot hybridization; the result for a representative number of individual Ltk$^+$ transfected cell clones is shown in Figure 1. Ltk$^-$ cells do not synthesize MMTV specific RNA. This shows that the L-cell endogenous MMTV proviruses are transcriptionally inactive. Individual

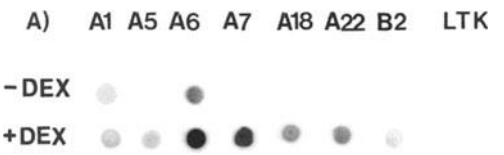

Fig. 1 : Dot-blot analysis of total transfected cell RNA. Ten micrograms of total RNA extracted from individual cell clones grown in the presence or absence of 10^{-6} M dexamethasone was applied to a nitrocellulose filter and hybridized to cloned ^{32}P-MMTV DNA. Ltk : RNA from untransfected Ltk$^-$ cells.

Ltk$^+$ cell clones transfected with the purified 9 kb MMTV insert (clones A) or with the complete λ phage containing the MMTV insert (clone B2) make variable but small amounts of viral RNA in absence of dexamethasone, a synthetic glucocorticoid (Fig. 1 A). In most cell clones the amount of viral RNA increases after overnight treatment of the cells with 10^{-6} M dexamethasone. Only one cell clone (A1), out of at least 20 that were tested, did not respond to glucocorticoids, for reasons that were not further investigated; the lack of response might be due to a defect in the provirus or in the cells. The levels of viral RNA synthesis in individual cell clones were generally correlated to the copy number of transfected DNA (see e.g. 13). The thymidine kinase gene (tk) which was also introduced into the Ltk$^-$ cells as a selective marker was not transcribed in a hormone-dependent fashion (13), suggesting that it was expressed independently. Dot-blot analysis of RNA is useful as a rapid screening method of a large number of cell clones. However aberrant transcripts from abnormal initiation sites or from the opposite strand will also contribute to the hybridization signal. To avoid this problem we examined the same RNA samples for the presence of transcripts starting at the MMTV promoter by the S1 nuclease mapping procedure (14) using an end-labeled DNA probe (15, 16). The result of such an experiment is shown in Figure 2. A unique start site for viral

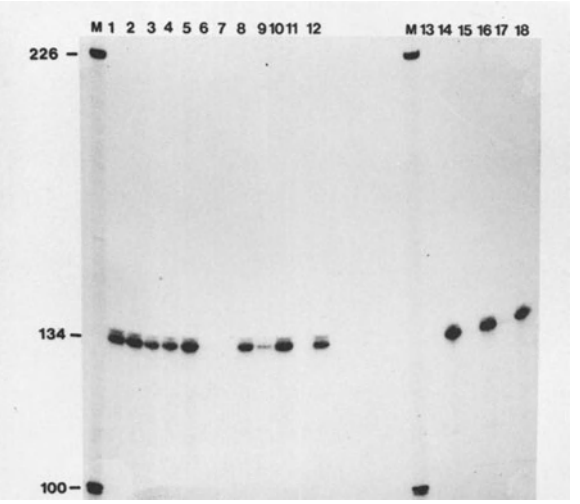

Fig. 2 : S-1 nuclease mapping of the MMTV-initiated transcripts in transfected cells. Total cellular RNA (50 µg) was hybridized to a 5'-end labeled Pvu II fragment (for details see ref. 16) covering the area of initiation of viral transcription. After treatment of hybrids with 1,000 units/ml of S1 nuclease the protected DNA fragments were separated on a 6% polyacrylamide-urea sequencing gel. M : pBR-322 DNA fragments of 226 and 100 basepairs as marker. RNA was isolated from individual Ltk^+ cell clones transfected with cloned exogenous MMTV DNA (lanes 1-10, 13-18), grown in the presence of 10^{-6} M dexamethasone for $\sim$16 hrs (lanes 1-10, 14, 16, 18) or in the absence of the hormone (13, 15, 17). Lane 12 : RNA extracted from GR tumor cells in presence of dexamethasone. Lane 11 : RNA extracted from Ltk^- cells.

transcription in GR tumor cells (lane 12) is located 134 nucleotides upstream from the end of the viral LTR (see Fig. 5). In individual transfected cell clones the identical start site for transcription was observed (lanes 1-10, 13-18). The amount of properly initiated RNA varied from cell clone to cell clone as observed in the dot-blot analysis (Fig. 1), and in some cell clones the level of MMTV transcription was below detection. In three cell clones the transcripts synthesized in presence or absence of dexamethasone were compared

(lanes 13-18). Transcription was below the level of detection in absence of the hormone in two (lanes 13 + 15) and barely detectable in one of the clones (lane 17), while the band was very strong in the RNA samples from glucocorticoid treated cells. Based on this experiment we estimate the stimulation of transcription by the hormone to be at least 20-fold (16, 17).

To accurately map the unique start site of MMTV transcripts the S1 nuclease-protected fragment was electrophoresed on a sequencing gel together with the products obtained by subjecting the end-labeled probe to the chemical degradation procedure of Maxam and Gilbert (18). As is shown in Figure 3 the S1-protected fragment runs about 1.5 nucleotides

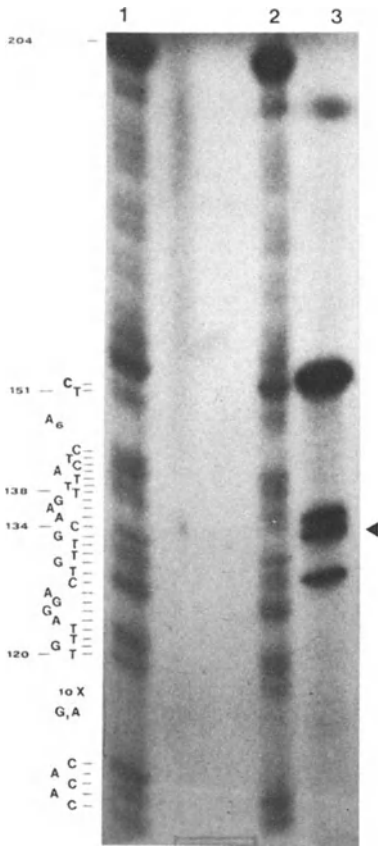

Fig. 3 :
Determination of the 5' end of MMTV-GR tumor cell RNA by S-1 nuclease mapping. The ^{32}P probe is a mixture of 5' end labeled Alu I fragments of the LTR region (16). The protected cap fragment is indicated with an arrow (lane 3). Lanes 1 and 2 show sequence ladders (18) of the T + C reaction performed on the purified Alu I fragment containing the cap site. The sequence of the non-coding strand of the DNA is given (11).

slower than the fragment of 134 nucleotides, in agreement with the theoretical prediction (19). We therefore conclude that the initiation site for viral transcription is located ∿134 nucleotides upstream from the end of the LTR (see Fig. 4 and 5). Viral RNA's synthesized in transfected cells were also analyzed for their size by gel electrophoresis and subsequent blot analysis. These data have been published (9, 10, 13, 17). A summary of these data is given in Figure 4. Two major RNA species are readily found in transfected cells (20, 9, 10). Full length 9 kb viral RNA, which is indistinguishable from the 9 kb viral mRNA, codes for the viral core proteins gag and pol (21, 22). The 4.4 kb mRNA codes for the envelope precursor polypeptide (22a). These RNA species are identical with the RNA species previously found in MMTV infected cells (23, 24, 25).

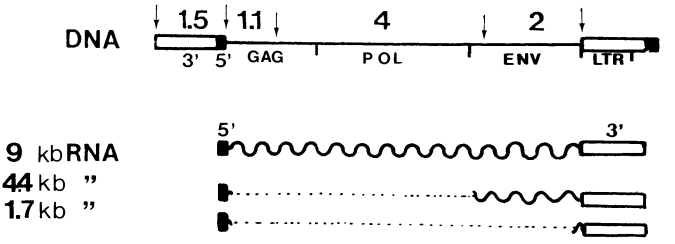

Fig. 4 : Schematic representation of MMTV mRNA's. The arrows indicate the Pst I cleavage sites in GR MMTV DNA. The individual fragments were subcloned in pBR 322.

The existence of a third RNA species has been postulated on the basis of the DNA sequence determined for the LTR's of different strains of MMTV. Starting with the A of an initiation codon outside the LTR, an open reading frame of 960 nucleotides was detected, potentially coding for a protein of 36 K (11, 26, 27). Such a protein had previously been

found in in vitro translation studies using 3'-end fragments of viral RNA (28). When over-exposed films of RNA blots of transfected cells are carefully examined for the presence of a low molecular weight RNA, a discrete band of 1.7 kb can be seen in cells which contain a large copy number of transfected DNA (9, 17). Very recently a corresponding spliced mRNA of 1.7 kb has been found in preneoplastic and neoplastic mouse mammary tissue of Balb/c mice (29) and in mammary glands of several mouse strains (30). Using nick-translated probes corresponding to the different regions of the MMTV provirus (e.g. the Pst I restriction enzyme fragments of 1.45, 1.1, 4 and 2 kb shown in Fig. 4, subcloned in the plasmid pBR 322) a rough map of the subgenomic RNA's was established (17). Using the S1 nuclease mapping technique, a unique 5'-splicing site for these mRNA's, located $\sim$ 288 nucleotides downstream from the 5' end of the viral RNA, was determined recently (12). The same 5'-splicing site was observed for the 1.7 kb mRNA by Van Ooyen et al. (30). The 3' splicing sites for the 4.4. kb env mRNA (31, 32) and the 1.7 kb mRNA (29, 30) have been recently mapped.

After having shown that MMTV DNA contained the sequences necessary for the stimulation of viral RNA synthesis we proceeded to localize these sequences more precisely. Using the 1.45 kb Pst I restriction enzyme fragment containing most of the LTR and a short segment of unique 5' DNA in transfection experiments we were able to attribute hormone responsiveness to this fragment (12). Using recombinant DNA's between MMTV fragments and a marker gene, other laboratories came to the same conclusion (33, 34, 35). However in many of these constructions additional viral or cellular sequences were also present (33, 35) and their contribution to the observed effect could not be excluded. We further wanted to determine what part of the 1328 bp long LTR is involved in the hormonal control. We therefore removed increasing portions upstream of the transcriptional start site in a chimaeric plasmid where the MMTV LTR had been linked to the coding region of the thymidine kinase (tk) gene of herpes simplex virus (HSV). Plasmids with deletions were then introduced into Ltk$^-$ cells and stably transfected cell clones were isolated (16). The initiation of transcription at the MMTV promoter and its regulation by glucocorticoids were then studied.

A summary of these recently published results (16) is shown in Figure 5. RNA was extracted from cell clones tranfected with the deletion

mutants and transcripts from the MMTV promoter were mapped by the S1 nuclease technique. Amounts of virus-specific RNA's made in absence or presence of glucocorticoids were estimated by scanning the signals of the 134-nucleotide protected cap fragment. Due to the difference in DNA copy number in individual cell lines the absolute amounts of viral RNA vary from cell clone to cell clone. However the specific stimulation observed by the addition of the hormone to an individual cell clone was independent of the absolute level of transcription. As is indicated in Figure 5 removal of sequences up to ~200 nucleotides upstream from the

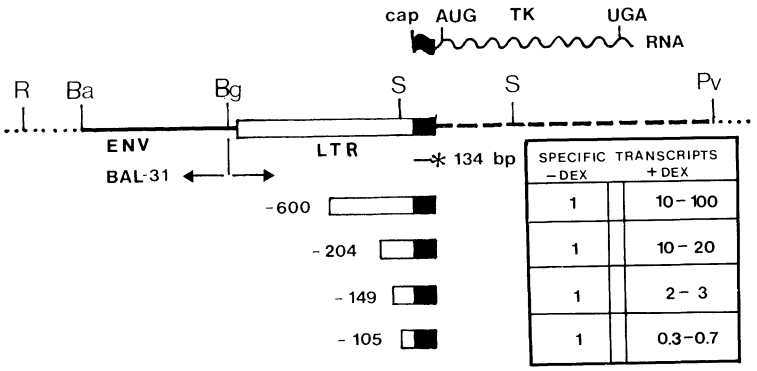

Fig. 5 : Hormone dependence of transcription in cells transfected with LTR deletion mutants (for details see text).

cap site does not lead to a significant reduction of the hormone response. When the deletion extended to -149 bp, the results were less clear, but mainly pointed to a greatly reduced, if not completely abolished, response to the glucocorticoid. Removal of all LTR sequences upstream of -105 bp from the cap resulted in a complete loss of the hormone response, while a base level of transcription starting at the correct site was maintained. From these data we conclude that glucocorticoid regulation of the MMTV promoter requires an upstream DNA region between -105 and -204 base pairs.

DISCUSSION

Our cloning and transfection experiments demonstrate that MMTV proviral DNA contains the necessary element for glucocorticoid regulation of viral transcription. A contribution of endogenous MMTV sequences to the phenomenon was excluded by transfecting the same cloned DNA into mink lung epithelial cells which do not contain such sequences (10). Similar experiments performed with cloned integrated endogenous proviruses and their flanking cellular sequences gave basically the same result (36, 10, 13). These proviruses are not expressed and do not respond to glucocorticoids in vivo. This suggests that the cloned endogenous and exogenous MMTV proviruses contain the necessary sequences for hormonal response; however this response can only be functional when viral transcription is already turned on. In this respect the hormone responsive element can be considered equivalent to an enhancer sequence (37) which allows the modulation of the promoter efficiency, but which is not sufficient to turn on a silent promoter. Several laboratories localized the sequences necessary for the hormone response in the viral LTR (33, 34, 11, 35). Our experiments summarized above (16) demonstrate that a fragment of MMTV DNA from $\sim$ -100 to $\sim$ -200 bp upstream from the initiation site of transcription is required for hormone stimulation. The sequence of the relevant area is shown in Figure 6.

Similar deletion experiments have been reported (35, 38) with a chimaeric DNA construction containing MMTV env and LTR sequences in addition to some cellular sequences linked to the complete tk gene of HSV including its own promoter of transcription. Using this construction the authors did not detect any transcripts initiated at the normal MMTV cap site in absence of glucocorticoids. This result differs from our observation of a low level of constitutive specific transcription in absence of the hormone. The reason for this difference might be that in our experiments the selection of tk^+ cells in absence of the hormone depended on sufficiently high basal level of MMTV promoter activity in order for the cells to become resistant to the selective HAT medium. We might have therefore selected for cells containing a high copy number of chimaeric molecules (probably obtained by amplification (16)) and therefore high enough levels of stable mRNA to give a positive signal in the S1 nuclease assay. It is also possible that the close vicinity of the tk

promoter in the construction of Hynes et al. (38) somehow competes or interferes with the MMTV promoter for the binding of polymerase molecules in absence of the hormone.

In presence of the hormone the authors observed good transcriptional activity when 202 bp of upstream sequences were present. This 5' boundary for the sequences required for the hormonal response is compatible with our findings. However when the deletion of Hynes et al. extends to -50 nucleotides from the cap site transcription is still observed in presence of the hormone, suggesting the presence of DNA sequences in this plasmid which confer hormone sensitivity to the MMTV promoter. This is in contrast to our data which show that sequences up to -105 bp upstream from the transcriptional start site are not sufficient to establish hormone regulation.

It has been suggested by Groner et al. (35) that the viral promoter region overlaps with the sequences required for hormonal regulation of transcription. We think that this is not the case. The basal level of correctly initiated transcripts found with the -105 bp deletion mutants is comparable with the level observed with the entire proviral DNA in absence of glucocorticoids, suggesting that sequences required for hormone response and sequences necessary for correct initiation of transcription do not overlap significantly. Our finding was recently confirmed by Majors and Varmus (39) who observed the persistance of a constitutive level of RNA synthesis in cells transfected with a deletion mutant to position -80.

Using different techniques, recently several groups have reported that partially purified hormone-receptor complexes preferentially bind in vitro to certain cloned MMTV DNA fragments (40, 41, 42, 43, 44). Although binding sites have been observed in different regions of MMTV DNA and of flanking mouse DNA, all authors agree on the presence of at least one binding site in the LTR. Finer mapping place it in the 400 bp at the 3' end of the LTR. Nuclease protection experiments (45) identified two partially homologous receptor binding sequences upstream from the MMTV promoter. The protected areas on the coding DNA strand are marked with bars in Figure 6. The first one is found between nucleotides -189 and -164. A second one is not contiguous, but seems to be composed of three smaller DNA stretches extending from nucleotides -123 to -70. When these binding sites are compared to our functional studies a good

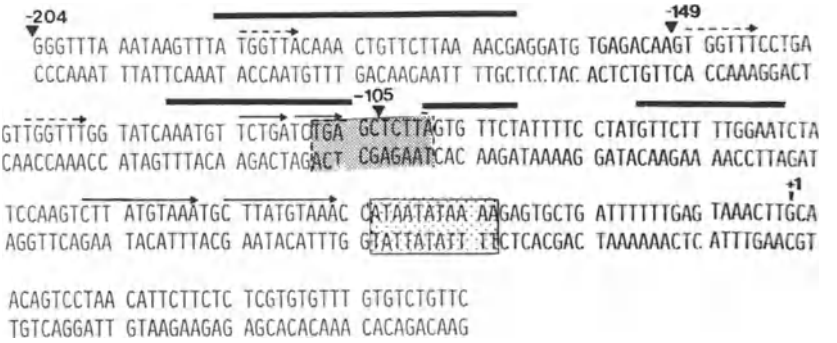

Fig. 6 : Sequence of the region of the GR LTR with deletion breakpoints. Both DNA strands are shown. +1 indicates the cap nucleotide. Arrows indicate repeats. The lightly shaded box indicates the TATA box, the heavily shaded box the consensus sequence among glucocorticoid regulated genes (see Discussion). The dark bars correspond to the binding regions of the hormone receptor complex in the coding DNA strand (45).

coincidence of the relevant areas is evident. The deletions which are fully functional (-204 deletions) contain the two binding sites identified by Scheidereit et al.. The deletion which extends to -149, that has lost most of the glucocorticoid response, does not contain the contiguous first hormone-receptor binding site. The deletion extending to -105 is biologically completely inactive and contains only about half of the second, interrupted binding site. It is possible that both of these binding sites are needed to render the gene fully responsive to glucocorticoids. However our transcription data do not exclude the possibility that the region between -204 and -149 and therefore one single binding site, is sufficient for the increased rate of transcription. Nuclear transcription data performed in this laboratory with the cells transfected with the deleted molecules also demonstrate (8) that the deletion extending to -204 has full polymerase loading in presence of the hormone while the number of nascent RNA chains in presence and absence of the hormone is indistinguishable in the deletions -149 and -105.It has been observed recently that sequences located upstream from

-105 bp in the MMTV LTR can confer hormone sensitivity to the thymidine kinase gene of HSV (46). The hormonal stimulation of tk activity measured in cell clones transfected with different MMTV fragments varied considerably (from 1.9 to 16 fold). No stimulatory activity was observed when the left half of the MMTV LTR was used, while a fragment containing the upstream sequences between -450 and -105 bp of MMTV was still capable of conferring a moderate hormone sensitivity to the tk gene (3-fold stimulation of enzyme activity). A similar conclusion is also drawn by Hynes et al. (38) from their results with the chimaeric molecules containing the MMTV and the tk promoter. A slight hormonal stimulation of the expression was observed when LTR sequences upstream from -137 bp were present. We are currently testing whether the fragment of -100 bp to -200 bp identified in our experiments is sufficient by itself to confer hormone responsiveness to another gene.

Inspection of the DNA sequence between -1 and -204 (11, Fig 6) reveals a number of specific features. A 10 bp direct repeat separated by 2 bp is found just before the TATA box. Direct repeats of 5 bp (-108 to -117) and 6 bp (-143 -148 and -130 to -135) as well as an incomplete form of the latter (-182 to -187) are present in the important area for the glucocorticoid response. It has been noted (47) that one of the direct repeats found at position -148 to -143 and -130 to -135 resembles a "core" sequence 5' GTGG $\frac{AAA}{TTT}$ G3' which is present in several viral enhancers (48).

In addition we compared upstream sequences of different glucocorticoid regulated genes. Schmid et al. (49) found by computer analysis a 7/10 consensus sequence (TXAGXTCTXA) at position -101 to -111 between the rat tryptophane oxygenase (TO) gene and MMTV LTR's of three different strains. We found (50) 1 bp variants of this sequence at 2 additional locations in MMTV DNA (-252 in the LTR and -1230 at the end of the env gene, just upstream from the right LTR) and in two other glucocorticoid-regulated genes, the rat growth hormone gene (49) and the human proopiomelanocortin gene (52). Although direct and inverted repeats have been shown to be present in essential regions of other regulated genes, as e.g. the heat shock gene (53), their significance is as yet unknown. The most simple hypothesis, based on analogies to

prokaryotic systems (54) would predict that the region between -105 and -204 (Fig. 6) contains DNA sequences or DNA structures which are recognized by the hormone-receptor complex, which through binding to this region of the DNA would increase the efficiency of RNA polymerase binding to the MMTV promoter.

All of the data presented and reviewed here clearly exclude the possibility that the postulated protein coded for by the left 2/3 of the MMTV LTR is involved in the glucocorticoid stimulation of viral transcription. The fully active deletion mutants extending to position -204 contain no DNA sequences of the open reading frame, which terminates 30 nucleotides upstream from the deletion breakpoint.

REFERENCES

1. Varmus, H.E., Ringold, G.M. and Yamamoto, K.R. In : Glucocorticoid Hormone Action (Eds. J.D. Baxter and G.G. Rousseau), Springer, New York, 1979, pp. 253-278.
2. Ringold, G.M., Yamamoto, K.R., Bishop, J.M. and Varmus, H.E. Proc. Natl. Acad. Sci. USA 74 : 2879-2883, 1977.
3. Young, H.A., Shih, T.Y., Scolnick, E.M. and Parks, W.P. J. Virol. 21 : 139-146, 1977.
4. Ucker, D.S., Ross, S.R. and Yamamoto, K.R. Cell 27 : 257-266, 1981.
5. Groner, B., Hynes, N.E., Rahmsdorf, U. and Ponta, H. Nucleic Acids Res. 11 : 4713-4725, 1983.
6. Ringold, G.M., Yamamoto, K.R., Tomkins, G.M., Bishop, J.M. and Varmus, H.E. Cell 6 : 299-305, 1975.
7. Scolnick, E.M., Young, H.A. and Parks, W.P. Virology 69 : 148-156, 1976.
8. Firzlaff, J. and Diggelmann, H., 1984 (manuscript in preparation).
9. Buetti, E. and Diggelmann, H. Cell 23 : 335-345, 1981.
10. Owen, D. and Diggelmann, H. J. Virol. 45 : 148-154, 1983.
11. Fasel, N., Pearson, K., Buetti, E. and Diggelmann, H. EMBO J. 1 : 3-7, 1982.
12. Fasel, N., Buetti, E., Firzlaff, J., Pearson, K. and Diggelmann, H. Nucleic Acids Res. 11, 6943-6955, 1983.
13. Diggelmann, H., Vessaz, A.L. and Buetti, E. Virology 122 : 332-341, 1982.
14. Berk, A.J. and Sharp, P.A. Cell 12, 721-732, 1977.
15. Weaver, R.F. and Weissmann, C. Nucleic Acids Res. 7 : 1175-1193, 1979.
16. Buetti, E. and Diggelmann, H. EMBO J. 2 : 1423-1429, 1983.
17. Diggelmann, H., Buetti, E., Fasel, N., Owen, D. and Vessaz, A. In : Understanding Breast Cancer : Chemical and laboratory Concepts (Eds. M. Rich and P. Furmanski), Marcel Dekker, New York, in press.
18. Maxam, A. and Gilbert, W. In : Methods in Enzymology 65 (Eds. L. Grossman and K. Moldave), Academic Press, New York, 1980, pp. 499-560.
19. Sollner-Webb, B. and Reeder, R.H. Cell 18 : 485-499, 1979.
20. Hynes, N.E., Kennedy, N., Rahmsdorf, U. and Groner, B. Proc. Natl. Acad. Sci. USA 78 : 2038-2042, 1981.

21. Nusse, R., Asselbergs, F.A.M., Salden, M.H.L., Michalides, R.J.A.M. and Bloemendal, H. Virology 91 : 106-115, 1978.
22. Dahl, H.H.M. and Dickson, C. J. Virol. 29 : 1131-1141, 1979.
22a. Dickson, C. and Atterwill, M. J. Virol. 35, 349-361, 1980.
23. Groner, B., Hynes, N.E. and Diggelmann, H. J. Virol. 30 : 417-420, 1979.
24. Robertson, D.L. and Varmus, H.E. J. Virol. 30 : 576-589, 1979.
25. Sen, G.C., Smith, S.W., Marcus, S.L. and Sarkar, N.H. Proc. Natl. Acad. Sci. USA 76 : 1736-1740, 1979.
26. Donehower, L.A., Fleurdelys, B. and Hager, G.L. J. Virol. 45 : 941-949, 1983.
27. Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N.E., Herrlich, P., Michalides, R. and Van Ooyen, A.J.J. Nature 295 : 622-624, 1982.
28. Dickson, C. and Peters, G. J. Virol. 37 : 36-47, 1981.
29. Wheeler, D.A., Butel, J.S., Medina, D., Cardiff, R.D. and Hager, G.L. J. Virol. 46 : 42-49, 1983.
30. Van Ooyen, A.J.J., Michalides, R.J.A.M. and Nusse, R. J. Virol. 46 : 362-370, 1983.
31. Redmond, S.M.S. and Dickson, C. EMBO J. 2 : 125-131, 1983.
32. Majors, J.E. and Varmus, H.E. J. Virol. 47 : 495-504, 1983.
33. Huang, A.L., Ostrowski, M.C., Berard, D. and Hager, G.L. Cell 27 : 245-255, 1981.
34. Lee, F., Mulligan, R., Berg, P. and Ringold, G. Nature 294 : 228-232, 1981.
35. Groner, B., Herrlich, P., Kennedy, N., Ponta, H., Rahmsdorf, U. and Hynes, N.E. J. Cell Biochem. 20 : 349-357, 1982.
36. Hynes, N.E., Kennedy, N., Rahmsdorf, U. and Groner, B. Proc. Natl. Acad. Sci. USA 78 : 2038-2042, 1981.
37. Banerji, J., Rusconi, S. and Schaffner, W. Cell 27 : 299-308, 1981.
38. Hynes, N.E., Van Ooyen, A.J.J., Kennedy, N., Herrlich, P., Ponta, H. and Groner, B. Proc. Nat. Acad. Sci. USA 80 : 3637-3641, 1983.
39. Majors, J. and Varmus, H.E. Proc. Natl. Acad. Sci USA 80 : 5866-5870, 1983.
40. Payvar, F., Wrange, O., Carlstedt-Duke, J., Okret, S., Gustafsson, J.A. and Yamamoto, K.R. Proc. Natl. Acad. Sci. USA 78 : 6628-6632, 1981.
41. Payvar, F., Firestone, G.L., Ross, S.R., Chandler, V.L., Wrange, O., Carlstedt-Duke, J., Gustafsson, J.A. and Yamamoto, K.R. J. Cell Biochem. 19 : 241-247, 1982.
42. Pfahl, M. Cell 31 : 475-482, 1982.
43. Govindan, M.V., Spiess, E. and Majors, J. Proc. Nat. Acad. Sci. USA 79 : 5157-5161, 1982.
44. Geisse, S., Scheidereit, C., Westphal, H.M., Hynes, N.E., Groner, B. and Beato, M. EMBO J. 1 : 1613-1619, 1982.
45. Scheidereit, C., Geisse, S., Westphal, H.M. and Beato, M. Nature 304 : 749-752, 1983.
46. Chandler, V.L., Maler, B.A. and Yamamoto, K.R. Cell 33 : 489-499, 1983.
47. Parker, M. Nature 304 : 687-688, 1983.
48. Khoury, G. and Gruss, P. Cell 33 : 313-314, 1983.
49. Schmid, W., Scherer, G., Danesch, U., Zentgraf, H., Matthias, P.,Strange, C.M., Röwekamp, W. and Schütz, G. EMBO J. 1 : 1287-1293, 1982.

50. Fasel, N. Ph.D. Thesis, University of Lausanne, 1983.
51. Barta, A., Richards, R.I., Baxter, J.D. and Shine, J. Proc. Natl. Acad. Sci. USA 78 : 4867-4871, 1981.
52. Cochet, M., Chang, A.C.Y. and Cohen, S.N. Nature 297 : 335-339, 1982.
53. Mirault, M.E., Southgate, R. and Delwart, E. EMBO J. 1 : 1279-1285, 1982.
54. Ptashne, M., Jeffrey, A., Johnson, A.D., Maurer, R., Meyer, B.J., Pabo, C.O., Roberts, T.M. and Sauer, R.T. Cell 19 : 1-11, 1980.

17

THE INTERACTIONS OF VIRAL PROTEINS WITH ROUS SARCOMA VIRUS RNA AND POSSIBLE CONTROL OF REVERSE TRANSCRIPTION, TRANSLATION AND VIRION ASSEMBLY

JEAN-LUC DARLIX, CLAUDE MERIC and PIERRE-FRANCOIS SPAHR

SUMMARY

The interactions between Rous Sarcoma virus RNA and the viral proteins have been analysed in the virus and in vitro. We show that it is protein P12 and not P19 that binds tightly to RSV RNA both in the virus and in vitro. Specific RNA sequences are recognized by protein P12 in the virus and such sequences are close to or involve the splice sites, and the dimer linkage site likely to be required for viral packaging. In vitro P12 does not impair proviral DNA synthesis whereas it can completely inhibit RNA translation. Such data suggest that P12 may control negatively viral RNA splicing and translation and positively the packaging process.

1. INTRODUCTION

There is little doubt that genetic expression of nucleic acids is mediated, either directly or indirectly, by protein-nucleic acid interactions. The interplay of RNA with proteins has been and still is extensively studied and the RNA-protein interactions involving the RNA genome of Rous Sarcoma virus (RSV)* has emerged as a model system for such analysis due to the following reasons:

(i) RSV RNA is multifunctional in that it can direct reverse transcription, translation, splicing and virion assembly. (ii) The system is well defined since the sequence of RSV RNA, that of the

*Abbreviations used: RSV, Rous sarcoma virus; PAGE, polyacrylamide gel electrophoresis; UV ultraviolet light; nt, nucleotides; RNP, ribonucleoprotein.

structured domains, its encoded proteins and much of its mode of expression are all known. (iii) Several RNA clones as well as antibodies to the viral proteins have been isolated and (iv) a large number of mutants are now available (reviewed in refs. 1,2; 3-5).

The structural organization of the viral RNA is well known and the genome can be briefly described as consisting of : the leader sequence (initiation sites for reverse transcription and translation), the Gag gene encoding the internal structural proteins of the virion, the Pol gene encoding the viral DNA polymerase, the Env gene encoding the proteins found on the surface of the virion envelope, the Src gene encoding the transforming protein, and the U3-R region.

When cells are infected by RSV, the RNA is reverse transcribed by the viral DNA polymerase and the synthesized proviral DNA is integrated into the cellular DNA. The integrated proviral DNA is then continuously transcribed without disturbing cellular proliferation. The primary viral transcript is full genomic and this may be transported into the cytoplasm, without change, where it can serve both as a message for the translation of the Gag and Pol genes as well as a genome for the progeny. Alternatively the viral transcript may be spliced to yield the viral Env and Src mRNAs which are formed by the removal of "intron coding sequences" that are the Gag and Pol genes (Env mRNA) or the Env gene as well (Src mRNA) (2).

Thus, RSV RNA can direct four essential biological functions : reverse transcription, RNA splicing, translation and virion assembly. What features of this RNA molecule could account for these multiple functions including their control and modulation ? What are the roles, if any, of the viral proteins in controlling the functions of RSV RNA in the viral life cycle ? A partial answer to these questions might come from the fact that the purified viral RNA has an unusually high ordered folding (6). Based on experimental data we have recently proposed a cruciform structure for the leader, present in all three viral RNAs, and which contains the initiation sites for both reverse transcription and translation (7). This structure would readily explain the

apparent paradox of transcription and translation initiated on the same sequence of 100 nt (2,7) : although the initiation sites are very close to each other they are structurally independent (7). In addition the viral RNA was shown to interact with its own encoded proteins within the virus, with strong RNA-protein interactions occurring at 12-14 sites, all located in structured regions of the RNA (8-11).

Here we present a summary of our results on the interactions of the viral proteins P12, P19 and DNA polymerase with RSV RNA both within the virus and in vitro. The effects of these interactions on the functions of the viral RNA have been analysed and will be discussed.

2.1. Identification of the viral proteins interacting with the RNA within the virus

RSV (Pr-B LA23), a temperature-sensitive mutant for cellular transformation, was grown in cultures of chicken embryo fibroblasts and purified by sucrose gradient centrifugation. The virus was then irradiated with UV light which produces covalent linkages between the tightly interacting protein molecules and the RNA. This photo-induced cross-linking is so specific that it only occurs when the interacting molecules are both very close and their reactive groups are separated by not more than one bond length. The virus was subsequently lysed with SDS and the viral RNA purified by SDS-sucrose gradient centrifugation (see Fig. 1a). The 60-80S and > 80S control and UV-treated RNAs were recovered, heat denatured and further purified by Cs_2SO_4 density gradient centrifugation (Fig. 1b); control viral RNA banded at a density of 1,63-1,65 g/cm^3 in a Cs_2SO_4 density gradient whereas both UV-RNAs were at 1.57-1.60 g/cm^3. After extensive dialysis and RNA digestion with pancreatic and T_1 RNases, the proteins present in the control RNA and in both UV RNAs were analysed by SDS-polyacrylamide gel electrophoresis (Figure 2). Minute amounts of proteins remain associated with RSV RNA even in the absence of UV irradiation of the virus. Two major proteins of apparent molecular weights of 14 and 25 Kd are present in the 70S UV RNA, whereas most, if not all, the viral proteins are present

in the > 80S UV RNA with, however, a majority of proteins which have been identified as P12 and viral DNA polymerase.

We have shown that the rabbit antisera anti P12, anti P19, anti P27 and anti DNA polymerase were specific (10) as they recognize only the respective viral proteins. Immunoprecipitation of the in vivo S^{35}-labelled or in vitro P^{32}-labelled proteins UV linked to the viral RNA in the virus were carried out using the

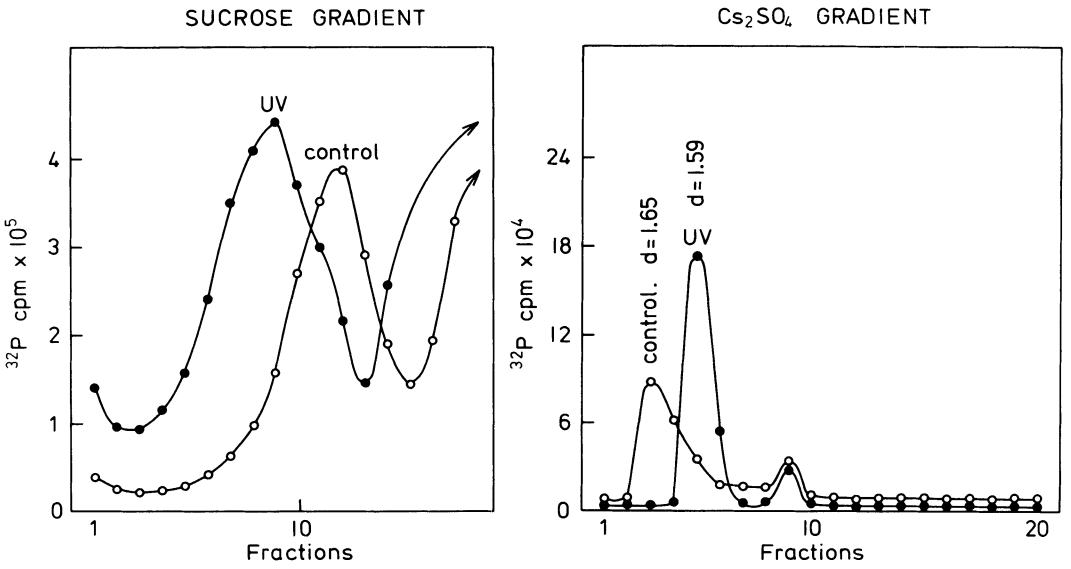

FIGURE 1. Purification of the viral UV RNAs
RSV (LA23-B) was grown (250 ml medium), labelled with P^{32}-orthophosphate (10 mCi) and S^{35}-methionine (1 mCi), and purified as previously described (12). Virus was irradiated with UV light (252 nm) at 4°C for 5 min (9) and viral RNA from control and UV irradiated viruses was purified by 5-23% sucrose gradient centrifugation in conditions previously described (9). 60-80S control and UV RNAs corresponded to fractions 8 to 14 and 5 to 12, respectively; > 80S control and UV RNAs to fractions 1 to 6 and 1 to 4, respectively. The fractions were pooled, diluted with 3 volumes of sterile water, heated 3 min at 95°C and adjusted to a density of 1,49 with Cs_2SO_4 at 20°C. Conditions for centrifugation were as described by Sen and Todaro (8). Control 70S and > 80S RNAs had the same density and fractions 2 to 5 were pooled; UV 70S and > 80S had a density of 1,57-1,60 g/cm^3 and fractions 4 to 7 were pooled. After extensive dialysis the RNAs were ethanol precipitated (2 vol) and redissolved in 7 M urea, 20 mM citrate pH 5.0 if proteins present were to be analysed by SDS-PAGE.

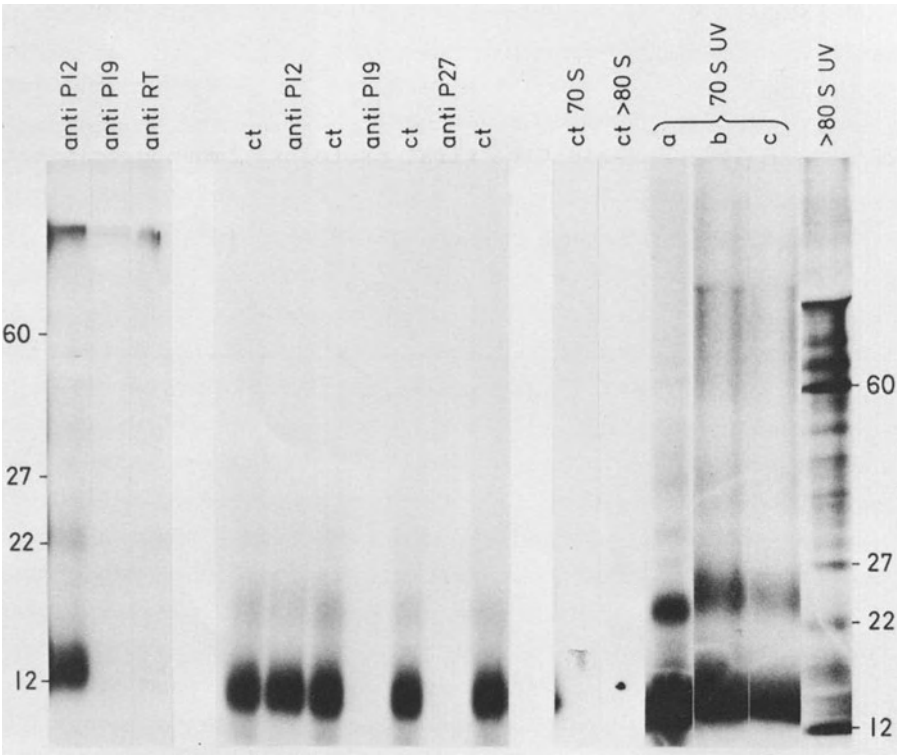

FIGURE 2. Analysis and characterization of the proteins UV linked to the viral RNA in the virus

The RNAs were prepared as reported in the legend of Figure 1 and digested with 10 µg of panc. RNase and 1 µg of T_1 RNase per assay (56°C for 1 h). Proteins were analysed on a 7.5%-15% SDS-polyacrylamide gel. ^{35}S-proteins were detected by fluorography (1-2 days of exposure).

Right panel : ct 70S and ct>80S correspond to the proteins present in the control 70S and > 80S RNA, 70S UV to the proteins UV cross-linked to the 70S RNA purified by sucrose gradient centrifugation (a) and further by banding in a Cs_2SO_4 density gradient (b and c). > 80S UV corresponds to the proteins present in the > 80S UV RNA purified by sucrose gradient and Cs_2SO_4 density gradient centrifugation.

Middle panel : 70S UV RNA was purified as above and digested with T_1 and panc RNases as described (9). The proteins were P^{32}-labelled in vitro using P^{32}-ATP and T_4 polynucleotide kinase (9). Reaction was stopped by adding EDTA to 10 mM. Antisera against each of the viral proteins were first adsorbed to protein A sepharose and subsequently washed to remove traces of nucleases, phosphatases and proteases. Immunoprecipitations were carried out for 1 h at 4°C with antisera against P27, P19 and P12. Controls correspond to P^{32}-labelled proteins incubated with the

antisera and protein A-sepharose. The results (Fig. 2) demonstrate that P12 is the major protein constituent of the 70S UV RNA and its apparent m.w. of 14 Kd instead of 12 Kd is very probably due to short oligoribonucleotides covalently linked to the protein. The dimeric (25 Kd protein) and trimeric forms of P12 probably correspond to protein molecules linked together by RNA sequences that are protected by P12 against RNase degradation. Small amounts of P19 and P22 have been detected by Western analysis (10) and radioactive quantitations indicate that the sum of P19 plus P22 (a precursor of P19; 13) is 10-15 times less abundant than P12. No viral DNA polymerase is detected in the 70S UV RNA whereas the enzyme does appear in the 20-30 times less abundant > 80S UV RNA (14). In conclusion, protein P12 and not P19 as previously reported by Sen and Todaro (8) is, by far, the major protein cross-linked to the viral RNA upon UV irradiation of the virus. Small amounts of P19, P22 and viral DNA polymerase are also present. The number of P12, P19 plus P22 and viral DNA polymerase molecules UV linked to the viral RNA was estimated using S^{35}-labelled virus. Between 0.7 to 1% of all P12 molecules were found to be crosslinked to the RNA and since there are 1200 P12 molecules present per 35S RNA within the virion (16), it can be concluded that 8 to 12 molecules of P12 are UV linked to one 35S RNA. Similarly the number of molecules of P19 plus P22 and viral DNA polymerase is estimated to be 0,5-1 each per 35S RNA.

2.2. The binding sites of viral protein P12 onto RSV RNA

90% of the protein UV linked to the viral RNA within the virus is P12 and its binding sites along RSV RNA have been located by T_1 fingerprint analysis (9). Control RNA and the UV RNA were purified as described above, further digested with T_1 RNase and material above 30 Kd was purified by Sephadex G75 filtration. The

---

respective antisera, the last control is without incubation. Numbers on the right of the figure refer to protein markers in Kd for the gels of the right and middle panels.

Left panel: S^{35}-proteins cross-linked to the 70S RNA were immunoprecipitated as above using the antisera against viral DNA polymerase, P19 and 12. Numbers on the left of the panel refer to the protein markers in Kd. Viral proteins used to prepare the viral antisera were purified as described in ref. 15.

RSV RNA content of this material was analysed by two-dimensional gel electrophoresis and by the T_1 fingerprint technique (9). The RSV RNA fragments were also 5'-P^{32} labelled and sequenced using the rapid enzymatic sequencing method (17,18). The RNA sequences, reported in Table 1, are in the order of decreasing yield of recovery, and their relative locations on the genome were assigned by comparison with PrC RNA sequences (3,9). The 12 to 13 P12 binding sites appear to contain direct or inverted repeats and as 8-12 molecules of P12 are UV linked to a 35S RNA, it can be assumed that one P12 is present per binding site.

The P12 binding sites, and not the P19 sites as previously reported (8,9), are located within the P19 coding sequence, the 3' end of the Gag and Pol genes and within the Env-Src gene junction. The P12 site located in the P19 coding sequence at position 510-523 is of special interest for several reasons : (i) this site is recovered with the highest efficiency (see Table 1 and ref. 9), (ii) it is within the dimer linkage structure (DLS) thought to hold together the two 35S subunits of the genomic RNA (19) and (iii) the DLS appears to be a sequence required for avian retroviral packaging (20).

5' Position*	Sequence
510	GUCCCCUUCAGCCUUAUAUUCCCCG
891	CUCCUCCUCCUCCUUAUG
2393	GCUCCUCUUCCCCG
1949 / 7015	AAAAAAUCG
5723	ACCUCCGCCUCUAACUAUACG
6197	UAGCACCCCACCAUACAG
6425	-14 AACAG GACAACAUCACG
872	GAAACGGCCACACCUAAAACCG
4963	AAAAACUG
4750	-22 AGGGGGAC AAUCCCCACCAGCAAACAGGGGGA
2955	UUCUCCCACUAUCUG
430	GUAAAACCUAUUGCGGGAAAACCUCUCCUUCUAAG

Table 1. Sequences at and around P12 binding sites on RSV RNA

Purification of P12 protected viral RNA fragments, and subsequent sequencing of the 5'-P^{32} RNA have been reported. Underlined sequences are either direct or inverted repeats. Numbers -22 and -14 refer to positions of residues relative to positions 4750 and 6425, respectively.

The binding site of viral DNA polymerase on RSV RNA is very probably the $tRNA^{trp}$ primer since the enzyme recognizes specifically the $tRNA^{trp}$ amongst cellular tRNAs (11) and binds very tightly to 70S RNA and not to viral RNA lacking $tRNA^{trp}$ (14). The enzyme initiates reverse transcription by elongation of the $tRNA^{trp}$ primer either in a reconstructed reaction or within the disrupted viral particle (2). These observations strongly suggest that viral DNA polymerase is tightly bound to the $tRNA^{trp}$ primer when the latter is hybridized to the viral RNA. We have, as yet, not been able to determine where the 0,5-1 molecule of P19 binds to the RSV RNA but this is under progress.

2.3. Characteristics of the interactions between the viral proteins and nucleic acids in vitro

Results of the interactions between the viral proteins and RNA in the virus have been extended by in vitro studies. Pure RSV P12, P19 and AMV DNA polymerase were tested for their binding affinities to viral and cellular RNAs as well as to DNAs. Increasing amounts of these proteins were adsorbed onto nitrocellulose filters and binding was conducted by incubating these filters in varying concentrations of P^{32}-labelled RNA or DNA. The results are reported in Figure 3, and the quantitation of the binding of P12 to the nucleic acids is shown in Figure 4. The relative binding affinities of P12 for the various tested nucleic acids follows the order : 70S RNA $>$ 35S RNA, rRNA $>$ ssDNA $>$ tRNA $\gg$ dsDNA. Similarly, with viral DNA polymerase the order of affinities is : 70S RNA $\gg$ 35S RNA, rRNA, ssDNA $>$ tRNA, dsDNA. Using this procedure binding of P19 to various nucleic acids appears to be very weak in agreement with the report of Bowen et al. (22) which is in marked contrast with reports of Sen and Todaro (8) and Leis et al. (24,25). In conclusion, the tight binding of P12 to genomic RNA in vitro agrees with the UV crosslinking of P12 to RSV RNA. Similarly the strong binding of viral DNA polymerase to 70S RNA, but weak binding to other nucleic acids lacking the $tRNA^{trp}$ primer, agrees with the very probable binding of the enzyme to the $tRNA^{trp}$ hybridized to the genomic RNA.

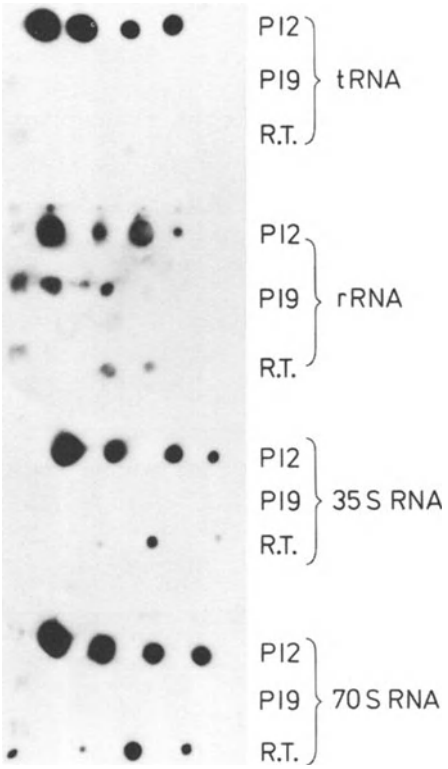

FIGURE 3. Binding of nucleic acids to viral proteins in vitro
RSV 70S RNA and 35S RNA P^{32} labelled in vitro were prepared as described previously (12). Ribosomal 28S RNA and tRNAs were prepared from cultures of chicken embryo fibroblasts and labelled with P^{32} orthophosphate in vivo. Specific radioactivity of the RNAs was 10^6 cpm/µg. The DNA probe was Pr-C cloned in PBR322 (a gift of Dr. Guntaka) and it was labelled in vitro using αP^{32} dCTP and DNA polymerase to a specific radioactivity of 10^7 cpm/µg. RSV P12 and P19 were purified from viruses according to the procedure of Johnson et al. (21). AMV DNA polymerase was of commercial source. Proteins were diluted to 50 µg/ml in 10 mM Na phosphate pH 7,5, 1 mM $MgCl_2$, 1 mM DTT and 10% glycerol and 0.1, 0.2, 0.4 and 0.8 µg were spotted on nitrocellulose. Nucleic acid binding was at 0,5 to 1 x 10^5 cpm P^{32}/ml and carried out as described by Bowen et al. (22) in the presence of 75 mM NaCl and at 37°C for 10 min. The filters were washed in the same buffer except that 0.1 M NaCl was added. Autoradiography was for 3-10 hrs. Increasing the concentration of NaCl to 0.25 M did not abolish binding of the nucleic acids to P12 and DNA polymerase, but it is decreased by a factor 2-3 and specificity for RSV 70S RNA is more pronounced.

2.4. Effects of P12 on the functions of the viral RNA.

We have shown previously that the P12 UV linked to the viral RNA caused the complete or partial inhibition of *in vitro* RNA translation and reverse transcription (9), respectively. P12 binds tightly to the genomic RNA *in vitro* (10), and at a ratio of one P12 molecule per 20-25 nt a fast sedimenting P12-RNA complex is generated (to be reported in detail elsewhere). This RNP complex is reminiscent of the viral RNP which can be isolated from detergent-treated virions and shown to consist of P12 and RNA at a ratio of one protein molecule per 10 nt (16). The effects of P12 on reverse transcription and translation were estimated by measuring the amounts of cDNA and polyprotein pr76 Kd (the *Gag* gene product, precursor to the viral proteins P19, P10, P27, P12 and P15), synthesized in conditions where more and more P12 were bound to the viral RNA until the fast sedimenting RNP complex was

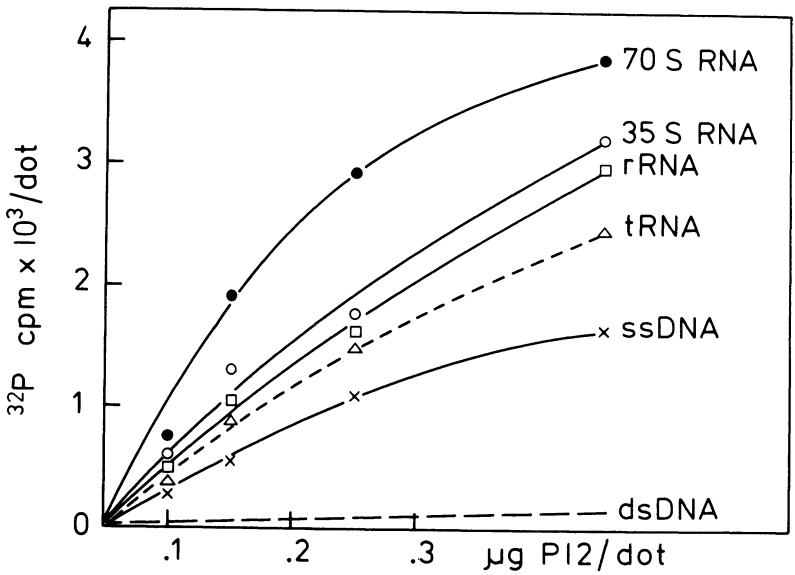

FIGURE 4. Relative binding affinities of nucleic acids for protein P12 in vitro

Binding experiments were conducted as described in the legend of Figure 3. Radioactive spots were then cut out and their radioactivity determined in a liquid scintillation counter.

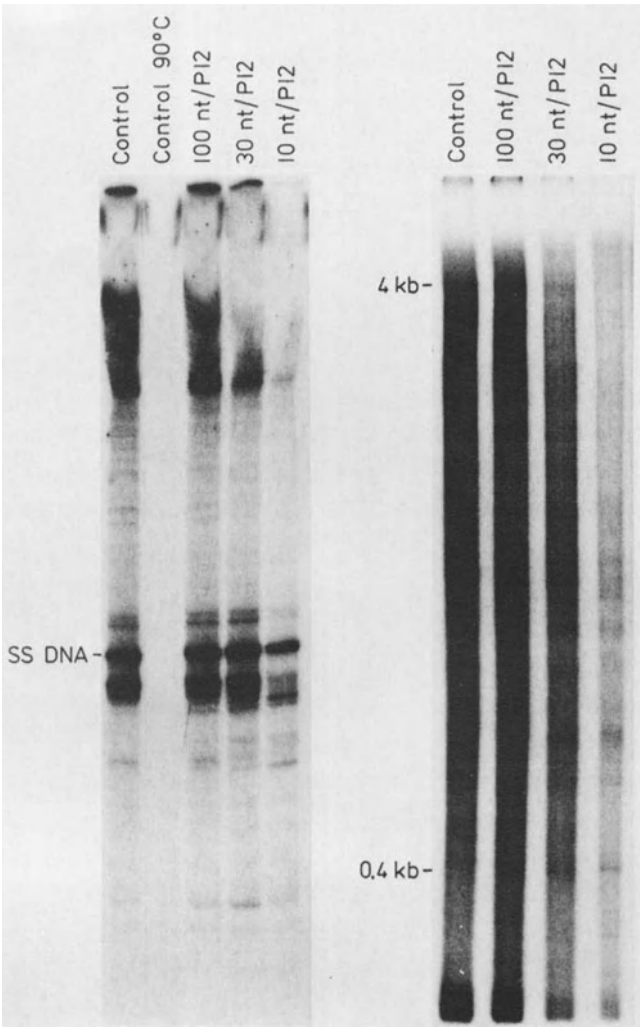

FIGURE 5. Effect of protein P12 on in vitro reverse transcription of RSV RNA

RSV RNA, P12 and AMV DNA polymerase were obtained as described in the legends of the previous figures. Reverse transcription was performed exactly as described before (23) except that αP^{32}-dCTP (Amersham) was at 50 µM and 30 µCi/ml. The P12 to RSV RNA ratio is indicated in the figure. The P12 was added to the incubation mixture prior to AMV DNA polymerase : after 10 min at 37°C, cDNA synthesis was carried out at 44°C for 6 min in order to detect solely strong stop DNA (12) and for 60 min to look at very large cDNA transcripts (23).

formed. Reverse transcription is not disturbed at one P12 per 30-100 nt, is slightly inhibited at one P12 per 10 nt and the length of the cDNA transcripts is not modified (Figure 5). In vitro translation of RSV RNA in the rabbit reticulocyte system leads to the synthesis of the polyprotein Pr76 Kd (9,26), and addition of P12 to the system progressively inhibits translation which is totally abolished at one P12 per 10 nt (Figure 6).

2.5. Unfolding of viral 70S RNA by protein P12

The viral RNA is extensively structured as indicated by its strong resistance to T_1 and pancreatic RNases (6), its susceptibility to RNases III and IV (27), and by electron microscopic observations (19,28). In addition we have been able to sequence the structured domains of RSV RNA (5).

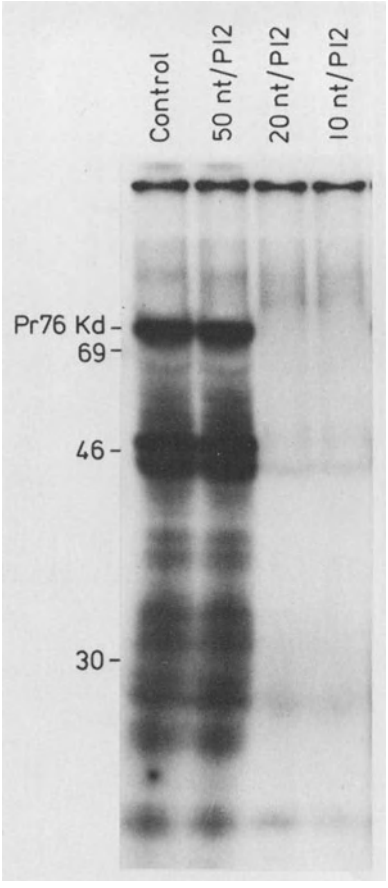

FIGURE 6. Effect of protein P12 on in vitro translation of RSV RNA

Translation of RSV RNA in the rabbit reticulocyte lysate was performed as described before (12) with a RNA concentration of 20 µg/ml. The P12-RNA complexes were formed at 33°C for 10 min prior to translation. After 60 min translation at 33°C reactions were stopped and Pr76Kd made was analysed on a 10% SDS-PAGE. The amount of Pr76Kd labelled with S^{35}-methionine was as follows : 5500 cpm (control), 6300 cpm (50 nt/P12), 700 cpm (20 nt/P12), 700 cpm (10 nt/P12).

The extent of RNA structure was estimated by analysing the T_1 RNase-resistant fragments on a polyacrylamide gel run under denaturing conditions (6). The P12 protein was bound to RSV RNA,

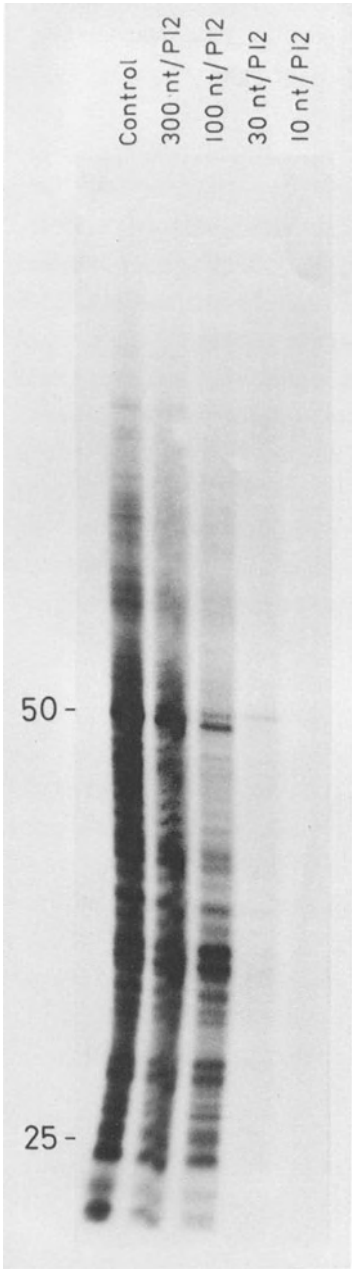

FIGURE 7. Unfolding of viral 70S RNA by protein P12

P12-RNA complexes were formed as described in legend of figure 3 except that 0.1 M NaCl was replaced by 0.1 M Tris-HCl pH7,5. After 10 min at 37°C, the reactions were put on ice and T_1 RNase was added at an enzyme to RNA weight ratio of 1/100 (6). The nuclease digestion was carried out for 30 min at 4°C and was stopped by the addition of 1% SDS. After two phenol-SDS-EDTA extractions and three ethanol precipitations, the RSV RNA fragments recovered were heat denatured and 5'-P^{32} labelled with P^{32}-ATP and T_4 kinase (Amersham). The 5'-P^{32} RNA fragments were analysed on a 12% PAGE in Tris-borate-EDTA pH 8,3 and 7 M urea.

the complex subsequently treated with T_1 RNase and following the removal of P12 and T_1 RNase, the structured RNA's were 5'-P^{32}-labelled and analysed. At ratios of one P12 per 50-300 nt the protein has little effect on the structure of RSV RNA, but as soon as the complex is formed (one P12 per 10-30 nt), the overall folding of the viral RNA is drastically modified and the RNA appears to be poorly structured at one P12 per 10 nt (Figure 7, and to be reported in more detail elsewhere).

3. DISCUSSION

We have studied the interactions of RSV RNA with the viral proteins both in the virus and *in vitro*. Upon UV light irradiation of virions we have shown that it is P12 and not P19 (8-10) which is the major protein (90%) to be UV-linked to the RNA. Between 8 to 12 P12 molecules per 35S RNA are thus associated sufficiently closely in the virus to be UV-linked to the RNA. The P12 proteins are located near (1) P12 binding sites at position 430, at the 3' end of the *Pol* gene and in the *Env-Src* junction which are in the close vicinity of the 5' and 3' splice sites, and (2) the P12 site at position 510 which is within a sequence called the dimer linkage structure very probably necessary for the efficient packaging of the viral RNA (19,20).

Viral DNA polymerase at a ratio of 0,5 to 1 molecule per 35S RNA is also UV-linked to the RNA and several lines of evidence indicate that it is bound to the $tRNA^{trp}$ primer (11,14). The location of the 0,5-1 molecule of P19 remains to be determined. The P12 molecules UV-linked to RSV RNA are also known to be very close to each other and in addition, the binding sites are located in structured RNA sequences probably brought together by RNA interactions (6,9). These data have led us to propose a model for the structure of RSV 35S RNA in which the nucleoprotein core takes into account three types of interactions : 1) RNA-RNA, 2) RNA-protein and 3) protein-protein interactions (Fig. 8). Other features of the model are that most of the RNA is on the outside as large loops and that both 5' and 3' ends of RSV RNA are free allowing binding of the viral DNA polymerase at the 5' end which will then synthesize strong-stop DNA and perform the 5' to

3' jump to resume reverse transcription (12,29). In the RNP most of the P12 molecules should interact with the large RNA loops (see Fig. 8) and, in fact, preliminary data indicate that the UV-linked P12 molecules are in close contact with non UV-linked P12 and also P27, but not with either P15 or P19.

What could be the functions of the viral protein P12 in the life cycle of RSV ? In the virus the RNP consists of 70S RNA complexed to approximately 2000-2500 molecules of P12 (16). We have tried to reconstruct the RNP *in vitro* using 70S RNA and pure P12 and the results obtained (to be reported in detail elsewhere) indicate that at a ratio of 2000 molecules of P12 per 70S RNA, most of the P12 are associated with the viral RNA and form a rapidly sedimenting complex. RSV RNA is efficiently reverse transcribed either within the P12-complex or in virions treated with a small concentration of detergent and thus it can be concluded that P12 does not impair the synthesis of the proviral DNA in the cell after viral infection. Possible functions of P12 in the synthesis of the double-stranded proviral DNA and in its integration into the host genome are currently under investigation.

The primary transcript of the integrated proviral DNA is the genomic 35S RNA which can be transported to the cytoplasm or spliced to give the viral *Env* and *Src* mRNAs. The binding of protein P12 to sites very close to the donor and acceptor splice sites (3) favors the idea that P12 can control the splicing of the 35S RNA and thus can regulate the relative amounts of RSV full-genomic, *Env* and *Src* RNAs. Large amounts of P12 inhibit completely the translation of the 35S RNA *in vitro* and this is probably due to the unfolding of the leader resulting in the disruption of the secondary structure of the ribosome binding site which would prevent initiation of translation according to our model (7). If P12 molecules accumulate locally on RSV 35S RNA in the cell, translation would cease and the genomic RNA would then be packaged into particles. Consequently P12 (should) exert a negative control on RSV RNA translation and regulates positively virion formation in the cell. Formation of the RNP should be initiated by

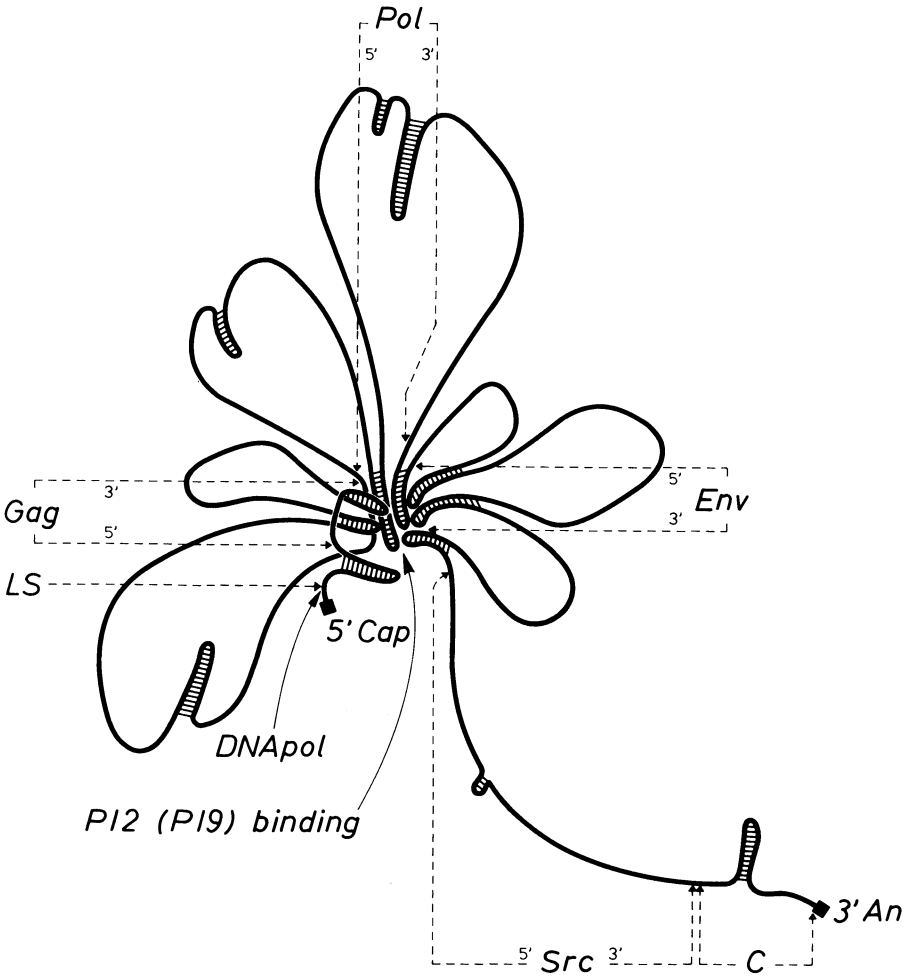

FIGURE 8. A model structure for RSV 35S RNA and its interactions with viral proteins

A model structure for RSV 35S RNA is presented showing the interactions between the UV linked protein molecules of P12 and DNA polymerase, and the structure core of 35S RNA as suggested by data presented here and elsewhere (9). According to this model the non-UV linked P12 should interact with the large RNA loops in the ribonucleoprotein particle.

the binding of P12 to the dimer linkage structure and then to other sites on RSV RNA as strongly suggested by the work of Pugatsch and Stacey (20). With one P12 every 10 nt the viral RNA is extensively unfolded and in the budding virions the RNA is complexed with P12 molecules in a helical conformation as seen by electron microscopy (30).

In conclusion, viral protein P12 very probably plays a critical role in the formation of competent viral particles by first inhibiting translation and then by packaging the viral RNA. Obviously further work is needed in order to know how the interactions P12-RNA take place in the cell, and to evaluate the role of P12, if any, on the synthesis of the proviral DNA and on the formation of the viral subgenomic RNAs.

Finally the viral DNA polymerase, proteins P19 and P12 can be phosphorylated *in vivo* (31,32). The β subunit of DNA polymerase is phosphorylated at its carboxyl terminus; a large fraction of P19 is phosphorylated whereas only a small portion of P12 is phosphorylated, and in both cases the phosphate is predominantly on serine residues (31). Although it has been reported that phosphorylation of reverse transcriptase can enhance its enzymatic activity (32), and that the phosphorylation of proteins may regulate their binding to nucleic acids, the roles of the differential phosphorylations of the viral proteins are unclear and are being studied in our laboratory.

ACKNOWLEDGEMENTS

Thanks are due to M. Schwager and C. Gabus for their technical assistance, to F. Veuthey for the computer facilities, to O. Jenni for the drawings and B. Brun for typing the manuscript. We acknowledge R. Southgate and P. Arrigo for their helpful comments. This work was supported by the grant No 3.321.82 to Professor P.-F. Spahr from the Swiss National Science Foundation.

REFERENCES

1. Darnell, J.E. Nature 297:365-371, 1982.
2. Weiss, R., Teich, N., Varmus, H. and Coffin, J. In: RNA tumor viruses (Eds. R. Weiss, N. Teich, H. Varmus and J. Coffin), Cold Spring Harbor Laboratory, 1982, chapters 4-7.
3. Schwartz, D.E., Tizard, R. and Gilbert, W. Cell 32:853-869, 1983.
4. Kitamura, N., Kitamura, A., Toyoshima, T., Hirayama, Y. and Yoshida, M. Nature 297:205-208, 1982.

5. Darlix, J.L. and Spahr, P.F. Nucleic Acids Res. 11:59-53, 1983.
6. Darlix, J.L., Schwager, M., Spahr, P.F. and Bromley, P.A. Nucleic Acids Res. 8:3335-3354, 1980.
7. Darlix, J.L., Zuker, M. and Spahr, P.F. Nucleic Acids Res. 10:5183-5196, 1982.
8. Sen, A.C. and Todaro, G.J. Cell 10:91-99, 1977.
9. Darlix, J.L. and Spahr, P.F. J. Mol. Biol. 160:147-161, 1982.
10. Méric, C., Darlix, J.L. and Spahr, P.F., submitted for publication.
11. Panet, A., Haseltine, W.A., Baltimore, D., Peters, G., Harada, F. and Dahlberg, J.E. Proc. Natl. Acad. Sci. (USA) 72:2535-2539, 1975.
12. Darlix, J.L., Spahr, P.F., Bromley, P.A. and Jaton, J.C. J. Virol. 29:597-611, 1979.
13. Rohrschneider, J.M., Diggelmann, H., Ogura, H., Friis, R.R. and Bauer, H. Virology 75:177-187, 1976.
14. Golomb, M. and Grandgenett, D.P. J. Biol. Chem. 254:1606-1613, 1979.
15. Henderson, L.E., Sowder, R. and Oroszlan, S. In: Chemical Synthesis and Sequencing of Peptides and Proteins (Eds. Liu, Schechter, Heinrikson and Condliffe), North Holland, Inc., 1981, pp. 251-260.
16. Davis, N.L. and Rueckert, R.R. J. Virol. 10:1010-1020, 1972.
17. Donis-Keller, M., Maxam, M. and Gilbert, W. Nucleic Acids Res. 4:2527-2538, 1977.
18. Simonscits, A., Brownlee, G.G., Brown, R.S., Rubin, J.R. and Guilley, M. Nature (London) 269:833-836, 1977.
19. Murti, K.G., Bondurant, M. and Tereba, A. J. Virol. 37:411-419, 1981.
20. Pugatsch, T. and Stacey, D.W. Virology 128:505-511, 1983.
21. Johnson, S.P., Veigl, M., Vanaman, T. and Leis, J.P. J. Virol. 45:876-881, 1983.
22. Bowen, B., Steinberg, J., Laemmli, U.K. and Weintraub, H. Nucleic Acids Res. 8:1-20, 1980.
23. Darlix, J.L., Bromley, P.A. and Spahr, P.F. J. Virol. 23: 659-668, 1977.
24. Leis, J.P., McGinnis, J. and Green, R.W. Virology 84:87-98, 1978.
25. Leis, J.P., Scheible, P. and Smith, R.E. J. Virol. 35:722-731, 1981.
26. Vogt, V.M., Eisenman, R. and Diggelmann, H. J. Mol. Biol. 96:471-493, 1975.
27. Darlix, J.L., Spahr, P.F. and Bromley, P.A. Virology 90:317-329, 1978.
28. Kung, H.J., Bailey, J.M., Davidson, N., Vogt, P., Nicolson, M.O. and McAllister, R.M. Cold Spring Harbor Symp. Quant. Biol. 39:827-834, 1974.
29. Haseltine, W.A., Maxam, A.M. and Gilbert, W. Proc. Natl. Acad. Sci. (USA) 74:989-993, 1977.
30. Sarkar, N.H., Nowinski, R.C. and Moore, D.H. J. Virol. 8: 564-572, 1971.
31. Lai, M.C. Virology 74:287-301, 1976.
32. Lee, S.G., Miceli, M.V., Jungmann, R.A. and Hung, P.P. Proc. Natl. Acad. Sci. (USA) 72:2945-2949, 1975.

INDEX