Expression of Animal Virus Genomes

DAVID BALTIMORE

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

INTRODUCTION

The world of animal viruses appears to offer an unfathomable diversity of specimens but, as the molecular biology of the replication of many viruses has been studied, a pattern of behavior has emerged. The viruses can be divided into classes, each of which has its own method of transmitting its genetic information from one generation to the next and its own style of expressing its genetic information. Although in some cases the data are still fragmentary, it is possible to outline the behavior of these systems and to place them in a formal scheme. In this paper I will present such a scheme and I will discuss in some detail the behavior of three viral systems which have been investigated in my laboratory. Furthermore, I will discuss some of the implications of the existence of these viral systems in the context of the behavior of normal cells.

CENTRAL ROLE OF MESSENGER RIBONUCLEIC ACID

Viruses have evolved to the point where they are concerned mainly with two processes: duplication of their genetic material and controlled expression of the information inherent in the genetic material. The specific mechanisms which a given virus uses to carry out the processes of replication and transcription can be considered the viral genetic system (6). There are two end products of the viral genetic system: new genetic material and messenger ribonucleic acid (mRNA).

The one process which all viruses must perform is mRNA synthesis. The central role of mRNA derives from the fact that viruses use cellular ribosomes and soluble factors to translate their mRNA. Therefore, the job of the virus is almost finished once it has made mRNA, except for certain minor alterations in the functioning of the translation machinery.

THE VIRAL GENETIC SYSTEM

The specific mechanism for mRNA synthesis used by a given virus depends on the structure of the viral genetic material. Figure 1 is a composite diagram of the pathways of mRNA synthesis which are used by various classes of viruses. To maintain unity in the diagram, mRNA has been defined as "+" ribonucleic acid (RNA) for all the systems. This does violation to tradition in certain of the viral systems, but the gain in order seems worth the change of previous definitions.

The data on which this diagram are based are fairly conclusive for all cases except class VI. A general discussion of each class follows.
FIG. 1. Examples: I = T4 phage, vaccinia virus; II = +X/74; III = reovirus; IV = RNA phages, poliovirus; V = vesicular stomatitis virus, Newcastle disease virus; VI = RNA tumor viruses.

Class I

Class I consists of all viruses which have a double-stranded deoxyribonucleic acid (DNA) genome. These viruses, like cells, carry out an asymmetric transcription of their DNA to give rise to mRNA. Because in many viral systems different mRNA species come from different DNA strands, the designation of "+" and "−" strands in the genome is not meaningful, although locally the designation is possible.

Class II

Class II consists of viruses which have a single-stranded DNA genome of the same polarity as the mRNA (23, 38, 40). To my knowledge, no virus with single-stranded DNA complementary to mRNA has yet been found, although viruses with strands of both polarity in separate particles have been characterized (19, 33, 37).

Class III

Class III consists of viruses which have a double-stranded RNA genome. The mRNA for these viruses represents an asymmetric transcript of the genome (31). The known viruses of this type have multiple pieces of double-stranded RNA, each of which apparently contains the information for synthesis of a single protein (41).

Class IV

Class IV consists of single-stranded RNA viruses whose mRNA is identical in base sequence to virion RNA. In the case of poliovirus, the mRNA appears to be identical to the virion RNA in both size and base sequence (2), and it would appear that all of the picornaviruses share this property. The arboviruses (or togavirus) probably fall into this class also, although their mRNA has not been precisely defined. As noted below, the ability to isolate infectious arbovirus nucleic acid (48) is one reason for including them in this category. In order for the class IV viruses to synthesize mRNA (or more virion RNA), they must first produce a template "−" strand of RNA (2).

Class V

Class V consists of viruses which have a single-stranded RNA genome which is complementary in base sequence to the messenger RNA. Synthesis of their mRNA therefore involves transfer of information from one single-stranded RNA to its complementary RNA (17, 25). Vesicular stomatitis virus (VSV) is representative of this group (25), but Newcastle disease virus (NDV; reference 17) and possibly influenza (18) are also members. This group may then include all of the rabdo-, paramyxo-, and myxoviruses. If the myxoviruses are, in fact, part of this class, then some members have a single RNA molecule as their genome which is transcribed into partial replicas (VSV and NDV; references 17, 25), whereas others have multiple pieces of genome RNA (influenza; reference 20).

Class VI

Class VI consists of the RNA tumor viruses and other viruses which have a DNA intermediate in their growth and a single-stranded RNA genome. To fit these viruses into the scheme, it has been necessary to stretch the available evidence a long way, because the intracellular DNA is still more of a postulate than a reality and the mRNA has not yet been characterized. However, the evidence from metabolic inhibition studies (1, 44, 45) and the existence of the RNA-dependent DNA polymerase in the virions of these viruses (3, 46) strongly implies a DNA intermediate. Furthermore, the easiest (but not necessarily the correct) scheme for the handling of this DNA would be an asymmetric transcription to give rise to both virion RNA and mRNA. For this reason, I suggest that the two types of RNA probably have the same rather than complementary base sequences.

There are many viruses about which so little is known that they cannot be placed in the scheme in Fig. 1. However, as their transcription becomes understood, either they should fall into place in a recognized pattern or new classes will have to be added.

The overall viral genetic systems as defined above consist of both the transcription and replication systems (6). Figure 1 omits replication because, except for the systems where replication and transcription are synonomous (class IV and maybe VI), little is known about replication of the virion RNA. Viruses with similar
transcriptional systems could have different replicational systems leading to the necessity to extend the class designations.

**VIRION NUCLEIC ACID POLYMERASES**

One important aspect of the transcriptional systems of animal viruses has emerged during the last few years: the role of virion-associated nucleic acid polymerases. The initial discovery of the virion-associated DNA-dependent RNA polymerase of vaccinia virus (30, 35), followed a year later by the discovery of the double-stranded RNA-dependent RNA polymerase of reovirus (16, 39), opened up a new concept of how viruses carry out their initial transcriptional processes. More recently, the finding that VSV (8) and NDV (24) both have virion-associated RNA-dependent transcriptases has led us to see such enzymes as the rule rather than the exception. Finally, the existence of an RNA-dependent DNA polymerase in the virions of the RNA tumor viruses (3, 46) has led us to the concept that, whenever the first function performed by virion nucleic acid after its introduction into the cell is the transfer of its information to another nucleic acid, the enzyme responsible for this transfer is likely to be found in the virion.

The discovery of virion polymerases has provided a rationale for an old puzzle in virology. Why is it that infectious nucleic acid can be extracted from the virions of some viruses but not others? Virion polymerases provide the clue to this puzzle; where they exist the nucleic acid is noninfectious, where the nucleic acid is infectious they could not exist or at least they could not serve an obligate role. The infectivity of arbovirus RNA was the major reason for including these viruses in class V; the infectivity of their RNA suggests that virion RNA and mRNA must be of the same strand of RNA for these viruses.

**POLIOVIRUS**

The foregoing generalities provide a framework for considering the molecular biology of different types of viruses. In the case of poliovirus, which is the prototype of the group of viruses which is understood in most detail, the picornaviruses, the fundamental problems which are posed by the scheme in Fig. 1 are the mechanism of synthesis of the complementary RNA (the "-" strand) and the mechanism of synthesis of virion RNA. These two problems define the mechanism of the picornavirus genetic system because they encompass both transcription and replication. The existence of "+" and "-" strand synthesis in cells has been documented, and the nature of some of the intermediate molecules has been investigated (see reference 2), but the lack of an in vitro system in which the molecular details can be analyzed has frustrated progress.

One aspect of the poliovirus genetic system which has led to some surprising results is the existence of a single mRNA molecule which is the template for synthesis of a number of viral polypeptides (2, 5). The initial discovery of this situation was unremarkable because it was assumed that the mRNA represented a polycistronic mRNA in the sense in which such molecules are understood in bacteria. However, analysis of protein synthesis in infected cells has shown that a heretofore unknown mechanism is operative. The mRNA encodes a single polypeptide chain of about 250,000 daltons which is cleaved by proteolytic enzymes to form the functional viral proteins (5, 7, 26, 28). The evidence for this model has been presented in detail elsewhere (5, 7, 26, 28) and need not be reiterated here, except to note that two different types of experiments have revealed the single, large precursor protein: exposure of infected cells to amino acid analogues and incubation at high temperature.

The cleavage of this large polypeptide proceeds in at least three steps. First, smaller polypeptides are cleaved from the single, large polypeptide while it is in the process of synthesis; these can be called nascent cleavages (26). Their existence explains why the large polypeptide is never seen in untreated, infected cells: it never really exists in such cells because it is cut up before it is finished. The second class of cleavages involves processing of the initial products and occurs more slowly. Scission of the single capsid precursor molecule into three pieces typifies this series of events (26, 28). The capsid precursor has a half-life of about 15 minutes. A third type of cleavage is known for which only a single example has been seen. One of the capsid proteins is cleaved into two pieces in an event which occurs as RNA is being combined with viral protein to form the virion (27). The immediate precursor of the virion is, we believe, an icosahedral structure (the procapsid) consisting of 60 subunits, each of which is made of three proteins (5). As RNA is added to this structure to make the virion, one of the polypeptides in each of the 60 subunits is cleaved. This cleavage occurs only when RNA is added to the procapsid; in the absence of RNA synthesis no cleavage is observed (27).

**VESICULAR STOMATITIS VIRUS**

Recognition of the existence of the class V viruses came from two types of experiments. The
first was the purification from VSV-specific poly-
ribosomes of mRNA molecules which were much
smaller than the 4.2 × 10^6-dalton single-stranded
virion RNA and which would quantitatively hy-
bridize to virion RNA (25). The second was the
demonstration of the virion RNA polymerase
which would carry out the synthesis of small RNA
molecules from a long single-stranded template
(8). Together these experiments indicated that the
VSV transcription system was different from
those which were previously known. The exist-
ence of large amounts of intracellular NDV RNA
complementary to and smaller than NDV virion
RNA (17) suggested that it was closely related to
VSV, and the recent discovery of an NDV virion
transcriptase (24) has fortified that conclusion.
The very recent identification of an influenza
virion RNA polymerase (18) indicates that the
myxoviruses may belong to the same class, al-
though little is known about their mRNA.

RNA TUMOR VIRUSES

As mentioned above, the genetic system of the
RNA tumor viruses is not known, but a reason-
able guess as to its nature is possible. There are
two classes of evidence which point to a DNA
intermediate. One is the data accumulated by
using metabolic inhibitors and other perturba-
tions of events in cells (1, 44, 45). The measured
variable in these experiments is production of
virions or morphological transformation, or both,
criteria which are distant from the affected
biochemical process and which have therefore
provided circumstantial but not convincing evi-
dence. The second class of evidence is the dem-
ONSTRATION of an RNA-dependent DNA poly-
erase in virions of the RNA tumor viruses (3, 46).
This has been taken as strong evidence for a DNA
intermediate, but, in fact, it shows only how the
DNA intermediate could be formed if it is formed.
A third form of evidence, and potentially the most
powerful, would be the direct demonstration of
the postulated DNA in infected cells; however,
aside from a few suggestive reports (12, 13),
strong evidence has yet to appear.

Recently, my laboratory has been focusing on
two aspects of the behavior of the virion DNA
polymerase in RNA tumor viruses. (i) Does the
enzyme start DNA molecules de novo or does it
require a primer? (ii) Is the enzyme really an
RNA-dependent DNA polymerase, that is, does
it prefer to use ribonucleic acid templates rather
than deoxyribonucleic acid templates? These
questions could be investigated because the RNA
tumor virus DNA polymerase, as opposed to all
other known virion-associated nucleic acid poly-
merases, will utilize added nucleic acids as tem-
plates (34, 36, 42, 43). To investigate the ques-
tions which we posed, it was necessary to use hom-
opolymer templates rather than true nucleic acids,
because their behavior was more easily controlled.

Because most of the data have been recently
published (9, 11), I present the following conclu-
sions of this study.

(i) The enzyme cannot start molecules de novo
but requires at least an oligonucleotide primer to
initiate synthesis. In this property, the enzyme is
like other known DNA polymerases (14, 15, 21,
22) and unlike RNA polymerases, which are able
to start synthesis with merely a nucleoside tri-
phosphate initiator (32).

(ii) The enzyme prefers ribohomopolymers to
deoxyribohomopolymers, but its preference is not
absolute (see Table 1 and reference 9). It will
utilize poly(dC) as a template if an oligo(dG) is
provided to initiate synthesis, but no other deoxy-
ribohomopolymer will act as template even if the
appropriate primer is provided. In contrast, Esch-
erichia coli DNA polymerase I (the “Kornberg
enzyme”) shows a preference for deoxyribohomo-
polymers over ribohomopolymers although,
again, one ambiguity is seen: the enzyme will
utilize both poly(rA) and poly(dA).

IMPLICATIONS FOR CELL BIOLOGY

The foregoing material shows that we are in a
position now to understand viral genetic systems.
It is necessary to stretch evidence in places and
to make some guesses in order to provide unify-
ng concepts at the present time, but the rate of
progress is so great right now that within a few
years the picture should be quite clear. It is im-
portant that investigation of such systems con-
tinue because the results have implications be-
yond the province of virology. I should like to
look at some of these implications if for no other
reason than to illustrate how virology can illumi-
nate cell biology. I will mention three examples
drawn from the three areas of virology described
above.

Protein Synthesis

When we first realized that poliovirus proteins
were all derived from a single precursor, we asked
the obvious question: Why does the virus use this
apparently clumsy mechanism? We suggested, as
a possible answer, that perhaps all mammalian
mRNA is like poliovirus RNA, i.e., it has only a
single initiation site for protein synthesis and only
a single termination site (28). Many results in the
literature were consistent with this idea (see refer-
ence 28 for further discussion), and as yet no con-
tradictory evidence has appeared. A corollary of
this idea is that operons could not exist in mam-
Table 1. Template specificity of DNA polymerases

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Homopolymers&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rA</td>
</tr>
<tr>
<td>AMV DNA polymerase</td>
<td>+++</td>
</tr>
<tr>
<td>MLV DNA polymerase</td>
<td>+++</td>
</tr>
<tr>
<td>Escherichia coli DNA polymerase I</td>
<td>+</td>
</tr>
</tbody>
</table>

<sup>a</sup> Relative activities of different templates for the viral enzymes were estimated from the data in Fig. 3 and 4 in reference 9 and from other experiments in which rates were determined by using high specific activity precursors. Where active polymers are designated +++ or +++, inactive polymers (0) had less than 1% of the activity of their active homologues. The data on E. coli DNA polymerase I were obtained with a highly purified preparation (fraction 7; reference 30) kindly provided by Arthur Kornberg and assayed with 0.05 M potassium phosphate buffer (pH 7.4) and 6 mM magnesium acetate (10). Templates, primers, and substrates were varied as in the assays of the viral DNA polymerases. Poly(dA) was 10-fold more active than poly(rA) as a template for the E. coli enzyme.

Abbreviations: AMV, avian myeloblastosis virus; MLV, murine leukemia virus.

Abbreviations: Homopolymers of riboadenylate (rA), ribocytidylate (rC), riboinosinate (rI), ribouridylate (rU), deoxyadenylate (dA), deoxycytidylate (dC), deoxyinosinate (dI), and deoxythymidylate (dT).

malian cells as they do in bacteria, as units of transcription containing the information for linked but separate units of translation.

RNA-Dependent Transcription in Cells

The mechanism of mRNA synthesis used by VSV and the other class V viruses suggests an intriguing possibility for formation of some mRNA in cells. Possibly the DNA genome is transcribed into anti-mRNA which, in turn, is transcribed into mRNA. Although the postulation of an intermediate RNA may seem gratuitous, it would (i) provide a role for the unstable nuclear heterogeneous RNA; (ii) provide a new level for control of gene expression and a new mechanism for quantitative regulation of mRNA synthesis; and (iii) provide an explanation for the origin of some if not all RNA viruses. Evidence for such a mode of information transfer in cells could come from hybridization of mRNA to nuclear RNA. Experiments in progress indicate that in HeLa cells there is at least limited complementarity between these two species of RNA (M. Stempfer, M. Rosbash, A. S. Huang, and D. Baltimore, unpublished data).

RNA-Dependent DNA Synthesis

Discovery of the RNA-dependent DNA polymerase in RNA tumor viruses has touched off an explosion of interest in the possibility of such an enzyme and such a process in uninfected cells. It is possible to imagine such systems having a role in the processes of differentiation, the workings of the immunological system, and in the mechanisms of memory storage (4). Evidence that ribosomal gene amplification involves RNA-dependent DNA synthesis has already appeared (47).

CONCLUSION

Viruses which have a double-stranded DNA genome can express and duplicate their genetic material by processes which are at least formally identical to those used by cells. Viruses with other types of genomes require special systems for replication and transcription. These processes make up the viral genetic system, and it is possible to group viruses according to general properties of their genetic systems. A central element in the genetic systems of many viruses is a polymerase carried in the virion which carries out a transfer of information from one type of nucleic acid to another as probably the earliest event in viral replication.

A comparison of gene expression in three viral systems indicates the diversity of behavior which exists. The picornaviruses encode multiple protein species in a single polypeptide. The rhabdoviruses and paramyxoviruses carry many genes in a single RNA but synthesize separate, comple-
mentary RNA species. The RNA tumor viruses probably transfer their genetic information from RNA to DNA, where it can be expressed by DNA-dependent RNA synthesis.

Each of these examples suggests the existence of processes in cells which otherwise might not be evident.

ACKNOWLEDGMENTS

The author was a Faculty Research Awardee of the American Cancer Society and was supported by research grants from the Public Health Service and the American Cancer Society.

LITERATURE CITED


