VIRUSES, POLYMERASES AND CANCER

Nobel Lecture, December 12, 1975

by DAVID BALTIMORE
Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.

The study of biology is partly an exercise in natural esthetics. We derive much of our pleasure as biologists from the continuing realization of how economical, elegant and intelligent are the accidents of evolution that have been maintained by selection. A virologist is among the luckiest of biologists because he can see into his chosen pet down to the details of all of its molecules. The virologist sees how an extreme parasite functions using just the most fundamental aspects of biological behavior.

A virus is a form of life with very simple requirements (1). The basic needs of a virus are a nucleic acid to be transmitted from generation to generation (the genome) and a messenger RNA to direct the synthesis of viral proteins. The critical viral proteins that the messenger RNA must encode are those that coat the genome and those that help replicate the genome. One of the great surprises of modern virology has been the discovery of the variety of genetic systems that viruses have evolved to satisfy their needs. Among the animal viruses, at least 6 totally different solutions to the basic requirements of a virus have been found (2).

If we look back to virology books of 15 years ago, we find no appreciation yet for the variety of viral genetic systems used by RNA viruses (3). Since then, the various systems have come into focus, the last to be recognized being that of the retroviruses ("RNA tumor viruses"). As each new genetic system was discovered, it was often the identification of an RNA or a DNA polymerase that could be responsible for the synthesis of virus-specific nucleic acids that gave the most convincing evidence for the existence of the new system.

Now that the life-styles of different types of viruses have been delineated we can ask what relation there is between a virus' multiplication cycle and the disease it causes. In general, this question has no simple answer because disease symptoms do not correlate with the biochemical class of the virus. For instance, both poliovirus and rhinovirus are picornaviruses but one causes an intestinal infection with paralysis while the other causes the common cold. One class of RNA viruses, however, does have a unique symptom associated with it: the biochemically-defined group of viruses called the retroviruses are the only RNA viruses known to cause cancer. For a virologist interested in cancer, the problem is to first understand the molecular biology of retroviruses and then to understand how they cause the disease.

In what follows, I will review my personal involvement in uncovering the different genetic systems of RNA viruses, a story which leads to the recognition of the unique style of retroviruses. I will then consider what is known
about the relationship between the biochemistry of retroviruses and their ability to be oncogenic.

As I tell my story I will mention a few of the many co-workers, teachers and students who have influenced my thinking or contributed their labors and ideas to the products of my laboratory. To mention all of the people to whom I am indebted would make too long a list; I can only say that the honors I receive are in large measure testaments to their accomplishments.

**PICORNAVIRUSES**

My work on the genetic systems of RNA viruses dates back to my graduate school days. As part of my introduction to animal virology during a Cold Spring Harbor course, I heard Richard Franklin describe his then-recent experiments using autoradiography to show that Mengovirus, a picornavirus and a close relative of poliovirus, could shut off the nuclear synthesis of cellular RNA early after infection and could later induce new RNA synthesis in the cytoplasm which appeared to represent synthesis of viral RNA (4). I decided to go to the Rockefeller University as a graduate student with Richard Franklin in order to work on the system he had developed.

Before I began to study how picornavirus RNA was made, it was already known from the work of Simon that picornavirus RNA synthesis was independent of DNA synthesis (5). Furthermore, studies with actinomycin D had shown that neither synthesis nor expression of cellular DNA was involved in viral RNA synthesis (6). These results suggested that Mengovirus might make a cytoplasmic RNA-dependent RNA synthesis system. The concept that viruses induce synthesis of their own enzymes had strong precedents in bacteriophage systems - Seymour Cohen's work had shown decisively that new virus-specified enzymes were found in infected bacteria (7).

We approached the problem of the virus' effect on intracellular RNA synthesis as a question in enzymology. We first showed that the nuclei from Mengovirus-infected cells were greatly reduced in their ability to carry out cell-free DNA-dependent RNA synthesis compared to nuclei of uninfected cells (8). Later, we showed that cytoplasmic extracts of Mengovirus- or poliovirus-infected cells contained an RNA synthesis activity not evident in uninfected cells and not inhibited by actinomycin D (9). When we learned that the system made RNA of the size and structure of virion RNA (10), it became clear that it represented the postulated viral RNA-dependent RNA synthesis system.

While there has been extensive further analysis of crude cytoplasmic systems (11, 12) and impressive enrichment of the RNA synthesis system has been achieved (13), no pure enzyme able to make picornavirus RNA has ever been isolated so the detailed mechanism of viral RNA synthesis still remains obscure. Most of our knowledge of how picornavirus RNA is made has come from studies on the virus-specific RNA molecules in infected cells and their kinetics of labeling by radioactive precursors. Such research has been carried out in many laboratories (11, 12); my work in this area was done in association with James Darnell and Marc Girard. Together we found and studied
the relations between the poliovirus double-stranded RNA, the poliovirus replicative intermediate and the poliovirus replication complex (14). Marc Girard’s precise in vitro analysis of RNA synthesis capped this whole series of experiments (15).

A crucial part of the viral genetic system is the manner of translation of the viral messenger RNA. While working on viral maturation, my first graduate student, Michael Jacobson, and I began to realize that proteolytic cleavage was an important part of the process (16). Our work led us to suggest that the whole 7500-nucleotide viral genome might be translated into a single continuous polypeptide that we have called a polyprotein (17, 18). Recently, this work culminated in the demonstration that poliovirus RNA can be translated into this continuous polyprotein in a cell-free system and that some of the cleavages that make the polyprotein into the functional proteins appear to occur in extracts of uninfected cells (19).

The demonstration that the poliovirus genome RNA is the messenger RNA for the synthesis of viral proteins, coupled with the demonstration of the infectivity of viral RNA (20), implies that the poliovirus genetic system is very simple. The existing evidence confirms this simplicity - as seen diagrammatically in Figure 1: it appears that the incoming “plus” strand of

The Poliovirus Genetic System

![Diagram of the Poliovirus Genetic System](image)

Figure 1. Schematic representation of poliovirus-specific RNA synthesis in the cytoplasm of infected cells.
RNA synthesizes a "minus" strand which in turn synthesizes a series of plus strands. This diagrammatic simplicity of poliovirus replication hides a fair amount of as yet undeciphered complexity as shown by the work of Eckard Wimmer and his colleagues as well as by work in my laboratory. For instance, the 3'-ends of the virion RNA and messenger RNA are both poly(A), the 5'-end of the minus strand is poly(U), so we assume that they are templates for each other (21). But these homopolymer stretches have very variable lengths even in the progeny of a cloned virus; what then determines their length? The poly (A) serves some obscure but necessary function in the life-cycle of the virus (22); what is this function? There is no triphosphate 5'-terminus, either free or capped, on the virion RNA or messenger RNA (23, 24); how then is the synthesis of these molecules initiated? The 5'-end of the virion RNA and messenger RNA are different (24); what does this mean?

VESICULAR STOMATITIS VIRUS
Most of the work in my laboratory until 1969 centered on poliovirus. We had assumed that all RNA viruses would be similar in their basic molecular biology but during the 1960's results emerging from various laboratories implied that poliovirus, rather than being a model for all RNA viruses, used one out of a collection of different viral genetic systems. Probably the first dramatic demonstration of the variety in RNA viruses came from next door to Richard Franklin's laboratory at the Rockefeller Institute where Peter Gomatos and Igor Tamm found that reovirus has double-stranded RNA as its genome (25). The peculiarity of reovirus was underscored by the demonstration later that an RNA polymerase in the virion of reovirus is able to assymmetrically transcribe the double-stranded RNA (26). This was the first virion-bound RNA-dependent RNA polymerase ever found and followed after the finding of the first nucleic acid polymerase in a virion - the DNA-dependent RNA polymerase found by Rates and McAuslan and Munyon et al in virions of vaccinia virus (27).

Another observation that suggested there were profound differences among the RNA viruses was the finding that in cells infected by the paramyxovi-ruses, Newcastle disease virus or Sendai virus, much of the virus-specific RNA was complementary to the virion RNA (28). This result was in sharp contrast to what was found in poliovirus-infected cells where most of the virus-specific RNA was of the same polarity as virion RNA (11).

We branched away from concentration only on poliovirus to include the study of vesicular stomatitis virus (VSV) because of the lucky circumstances that brought Alice Huang to my laboratory. She joined me first at the Salk Institute and then we both came to M.I.T. in 1968. Alice had studied VSV as her graduate work with Robert R. Wagner at Johns Hopkins. We decided that the techniques developed for studying poliovirus should be applied to VSV, hoping that the peculiar ability of VSV to spawn and then be inhibited by short, defective particles could be understood at the molecular level. A graduate student, Martha Stampfer, joined in this work and together we
found that we had bitten off an enormous problem because VSV induced synthesis of so many species of RNA. In poliovirus-infected cells, only three species of RNA are seen but we found at least 9 RNA’s in VSV-infected cells and one of these RNA’s was clearly heterogeneous (29) - later work showed it had 4 components (30, 31). In our description of this work we said that 9 RNA species “seems exorbitant” (29) but we soon realized that each RNA had its place in the cycle of growth and growth inhibition of VSV.

As we were beginning to unravel the multiple VSV RNA’s, Schaeffer et al (32) published a paper showing that the major VSV-induced RNA’s in infected cells, like those induced bySendai and Newcastle disease viruses, were complementary in base sequence to the virion RNA. We confirmed and extended that observation, showing that the virus-specific RNA recovered from the polyribosomes of infected cells (the viral messenger RNA) was all complementary to virion RNA (33). This finding presented a pregnant paradox: if all viruses were like poliovirus and induced a new polymerase in the infected cell how could a virus that carried as its genome the strand of RNA complementary to messenger RNA ever start an infection? There seemed two possible solutions: the RNA came into the cells and was copied by a cellular enzyme to make the messenger RNA to initiate the infection cycle or the RNA came into the cell carrying an RNA polymerase with it.

Because no convincing evidence for RNA to RNA transcription existed (or exists) for any uninfected cell, the possibility of a polymerase with the incoming RNA seemed attractive. This possibility implied that there might be polymerase activity demonstrable in disrupted virions of VSV. Almost as soon as the power of this reasoning was clear to us, we had shown the existence of the virion RNA polymerase (34). The demonstration of this enzyme was the piece of evidence that led to the realization that there is a huge class of viruses, now called negative strand viruses (35), that all carry the strand of RNA complementary to the messenger RNA as their genome and that carry an RNA polymerase able to copy the genome RNA to form multiple messenger RNA’s.

RETVIRUSES
The discovery of a virion polymerase in VSV led us to search for such enzymes in other viruses. Because Newcastle disease virus made a lot of complementary RNA after infection it seemed a logical candidate and after an initial failure (34), we found activity in virions of that virus (36). But a more exciting possibility occurred to me; maybe by looking for a virion polymerase, light could be shed on the puzzle of how RNA tumor viruses multiply.

In his Nobel lecture, Howard Temin has outlined how he was led to postulate a DNA intermediate in the growth of RNA tumor viruses (37). Although his logic was persuasive, and seems in retrospect to have been flawless, in 1970 there were few advocates and many skeptics. Luckily, I had no experience in the field and so no axe to grind - I also had enormous respect for Howard dating back to my high school days when he had been the guru of the Summer School I attended at the Jackson Laboratory in Maine. So I decided
to hedge my bets - I would look for either an RNA or a DNA polymerase in virions of RNA tumor viruses.

To make the foray into tumor virology, I needed some virus. Peter Vogt first sent me some Rous sarcoma virus and, although I later used it to good advantage, I initially assayed for an RNA polymerase in this viral preparation and failed to find anything. Then George Todaro helped me utilize the resources of the Special Virus Cancer Program of the National Cancer Institute to get some Rauscher mouse leukemia virus. Using that virus preparation I set out to look for a DNA polymerase activity. With little difficulty, I was able to demonstrate that virions of Rauscher virus contained a ribonuclease-sensitive DNA polymerase activity and, after confirming the results with Rous sarcoma virus, I knew that the machinery for making a DNA copy of the RNA genome was wrapped up inside the virions of RNA tumor viruses (38). Simultaneously with my work, Temin and Mizutani discovered the DNA synthesis activity in Rous sarcoma virus (39).

BIOCHEMISTRY OF REVERSE TRANSCRIPTASE

Once the DNA polymerase activity had been demonstrated in the virions of what we now call retroviruses, many laboratories began to study the enzyme. A correspondent of *Nature* dubbed the enzyme "reverse transcriptase" (40) and this romantic name has become common parlance. About 2 years after its discovery, Howard Temin and I reviewed the literature on the enzyme (41). That review and later compendia (42) make a detailed rehash of the biochemistry of retroviruses superfluous. So, I will only present a brief sketch of the picture we have today of how a retrovirus multiplies and how the reverse transcriptase functions. I will not attempt to credit all of the people who have helped to clarify this picture.

There are two separate time-periods that can be distinguished after infection of a cell by a retrovirus (Fig. 2). The first period, which lasts a few hours, involves reverse transcription and establishment of the DNA provirus as an integrated part of the cellular DNA; the second period involves the expression of the integrated genome and synthesis of progeny virions.

Analysis of the first period of retrovirus growth has focused on the types of virus-specific DNA molecules that are produced. One important type of DNA that has been found is a closed circular duplex DNA of about $5.7 \times 10^6$ daltons (43). This DNA can transfect cells with one-hit kinetics (44) and therefore contains the total viral genetic information. Other DNA's that may be on the way to becoming the closed circular form are also evident (45). It has been hard to get definitive evidence as to what DNA form integrates but presumably it is the circular duplex DNA. Whatever the form that integrates, the evidence is quite good for acquisition of proviral DNA by the chromosomes of infected cells (46, 47).

The second period in a productive retrovirus growth cycle starts when the integrated genome begins making viral RNA (48). Synthesis of viral proteins and progeny virions ensues and the cell ever-after continues to make viral products except for variations imposed by the cell cycle (49). Among the
Based on present knowledge (42), the life cycle of an RNA tumor virus can be separated into two parts. In the first part the virion attaches to the cell and somehow allows its RNA along with reverse transcriptase to get into the cell’s cytoplasm. There the reverse transcriptase causes the synthesis of a DNA copy of the viral RNA. A fraction of the DNA can be recovered as closed, circular DNA (43) and it is presumably that form which integrates into the cellular DNA. Once the proviral DNA is integrated into cellular DNA it can then be expressed by the normal process of transcription. The two types of product which have been characterized are new virion RNA and messenger RNA. Much of the messenger RNA which specifies the sequence of viral protein is of the same length as the virion RNA but there may also be shorter messenger RNA’s (48). The virus-specific proteins have 2 known functions: one is the transformation of cells which occurs when, for instance, a sarcoma virus infects a fibroblast, the second is to provide the protein for new virion production. The transforming protein is shown here as acting at the cell surface but that is only a hypothesis.

viral proteins made in the infected cell may be a product that changes the growth properties of the cell (50); in such a case the retrovirus becomes a tumor virus.

The second period of the infection cycle can be dissociated from the first in a number of experimental systems. For instance, mammalian cells infected by avian viruses can gain viral DNA but not express it (46). Also, cells can have viral genomes that they inherited from their ancestors and such genomes are generally not being transcribed. Nonexpressed genomes can be activated: bromodeoxyuridine and iododeoxyuridine, for instance, stimulate the expression of inherited, silent viral genomes (51). The exact mechanism of activation of the genome for transcription, initiation of the transcript and termination of transcription are obscure, as are any processing events of the initial transcript which may occur.
It is evident that the key piece of machinery provided by the virus for this unique life cycle is the reverse transcriptase. Purified reverse transcriptase has the properties of most DNA polymerases: it is a primer-dependent enzyme that makes DNA in a 5' → 3' direction using deoxyribonucleoside triphosphates as substrates and taking the direction of a template for determining the base sequence of the product. The enzyme differs from normal cellular DNA polymerases by having a unique polypeptide structure, having an associated ribonuclease H activity and being able to make copies of RNA templates as readily as DNA templates (41). Genetics has shown us that the avian leukemia viral enzyme, at least, is encoded by viral RNA and needed only in the first period of the infection cycle (52).

The primer-dependence of the reverse transcriptase means that the enzyme can only elongate nucleic acid molecules, it cannot initiate DNA synthesis de novo. How then does the enzyme initiate the copying of viral RNA? The answer is that the genome RNA has attached to it a primer RNA molecule which is, in the case of avian leukosis viruses, cellular tryptophan transfer RNA (53). The avian leukosis virus reverse transcriptase has a high-affinity binding site for that transfer RNA which the enzyme presumably uses for precise initiation of reverse transcription (54).

RETROVIRUSES AND CANCER

The last 15 years of research in animal virology has allowed us to see the diversity of genetic systems used by the various types of RNA viruses and has most recently shown us how distinct the retroviruses are from the others. Rather than using an entirely RNA-dependent replication and transcription machinery, the retroviruses have included the DNA provirus in their lifecycle. Having a DNA intermediate does not make their mode of growth especially complicated - the DNA formally takes the place of the "minus" strand in the picornavirus genetic system - but the DNA is probably the due to why retroviruses are the only ones able to cause cancer. The DNA provides the necessary stability to the virus-cell interaction so that a viral gene product can permanently change the growth properties of an infected cell. Equally significant, the DNA stage is probably important to the ease with which retroviruses carry out genetic recombination (55); it is quite possible that the recombination system can bring together host cell genetic information with viral information and that in this way non-oncogenic retroviruses become oncogenic (56).

So the inclusion of a proviral stage in the retrovirus life-cycle may provide critical capabilities towards the development of an oncogenic potential. But the actual transformation of cells by retroviruses is a highly selective process; each type of oncogenic virus transforms a very limited range of cell types (57). If we assume that all transforming genes of viruses are like those of Rous sarcoma virus, genes that are not necessary to the growth of the virus (50, 58), then we can postulate that each type of transforming virus makes a specific type of transforming protein. Such a protein, by this model, would not be
critically involved in the multiplication cycle of the virus. Isolation of such transforming proteins and elucidation of their mechanism of action is the present challenge of cancer virology. Not only will such work help us to understand carcinogenesis, it may also be important to the study of developmental biology because of the intimate relationship between the differentiated state of cells and the type of virus able to transform them.

Another implication of the occurrence of a proviral stage in the life-cycle of retroviruses is that cells can harbor such viruses as genetically silent DNA molecules. In fact, in most, if not all, animal species, the normal cells of the body have DNA related to one or more types of retroviruses (59). They receive that DNA by inheritance, not infection, and in favorable cases it can be as precisely located in the chromosomes as any gene (60). What is the significance of these genes that look like viruses?

There have been three types of explanations for virus-related genes that are inherited in the germ line of so many animal species:

1) They are an aspect of the normal genetic complement of the animal and they are virus-related because they are the progenitors of retroviruses. These genes play some important role in the life of the animal and so are not dispensable. This explanation is basically the protovirus hypothesis put forward by Howard Temin (37).

2) They are genes inserted into the chromosome of some ancestral animal by a retrovirus infection of the germ cells of that animal. Because once the provirus is integrated it remains stably associated with the chromosome, the viral genes are inherited by progeny of the original infected animal. There is one force that can eliminate such genes from a species, the slow but inexorable process of mutation. As part of this explanation of inherited viral genomes, therefore, it has been suggested that the viral genes have some positive influence on the life of the animal and so are maintained by positive selection. This explanation is closely related to the virogene-oncogene hypothesis (61).

3) The previous explanation can be modified by the exclusion of any positive role for viral genes in the life of the infected animal. There are a number of reasons why positive selection may be an unnecessary attribute to postulate. For one thing, mutations are rare events and totally inactivating mutations are much rarer. Also, the virus can be genetically invigorated by becoming a replicating virus in the body of the animal and then reinfecting the germ line. When the virus starts multiplying as an independent entity, the burden of debilitating mutations it might have accumulated could be purged if a sufficient number of generations intervened between the activation of the latent provirus and its reintegration into the germ line. The reintegration might even replace the original provirus (62). If the viral genes were not transcribed in most cells that have the viral genome, as appears to be the case, the negative effect of having one or a few integrated genomes would be slight and probably insufficient to cause a serious negative selection against animals with inherited proviruses.

I would argue that the third explanation above is the one most likely to be correct. It is an explanation that maintains the separation of viral activities
and cellular activities and does not require the ad hoc postulation of beneficial properties of viral products. It treats retroviruses like any other virus, as an entity with its own lifestyle and its own accommodation with evolution.

In summary, I have tried here to develop the view of retroviruses as one of a number of solutions to the problem of creating a virus. Each virus directs synthesis of two critical classes of proteins: proteins for replication and proteins for constructing the virus particle. By encoding the reverse transcriptase, retroviruses have evolved the ability to integrate themselves into the cell chromosome as a provirus. This is a very sheltered environment in which to live, only mutation interferes with the continual transmission of the virus to the progeny of an animal that is infected in its germ cells. In this context, the ability of some retroviruses to cause cancer is a gratuitous one. But it is today the most challenging and important attribute of these retroviruses and the one that will dominate future research efforts in this area.

REFERENCES:
21. Alexander, H. E., Koch, G., Mountain, I. M., Sprunt, K. and Van Damme, O., Vi-
24. Colter, J. S., Birel, H. H., Mayer, A. W. and Brown, R. A., Vi-
26. Alexander, H. E., Koch, G., Mountain, I. M., Sprunt, K. and Van Damme, O., Vi-