

# RETROVIRUSES AND ONCOGENES I

Nobel Lecture, December 8, 1989

by

HAROLD E. VARMUS

Departments of Microbiology and Immunology, and Biochemistry and Biophysics, University of California at San Francisco, San Francisco, U.S.A.

## INTRODUCTION

The story that Mike Bishop and I will tell in these two lectures is one in which retroviruses, oncogenes, our personal histories, and the history of tumor virology are closely interwoven. It begins with some simple questions about the origin and behavior of viral genes and takes us to a vantage point from which we can survey many aspects of retroviruses and animal cells, including some of the aberrations that lead to cancer. We now know that retroviruses capture normal cellular genes and convert them to cancer-causing genes, called oncogenes. Such transductions are rare, but depend upon the normal events of an intricate virus life cycle. Retroviruses have introduced us to more than forty cellular genes with the potential to become oncogenes, some discovered as components of viral genomes, others as genetic targets for viral insertion mutations.

It has been our privilege to participate in a generous share of the experiments that established these principles. But we have required the help of many talented people in our laboratories at UCSF, as well as the collaboration and friendly competition of others elsewhere. (I mention as many names as the narrative can bear, but inevitably I must apologize to valued colleagues who remain anonymous here.) Several viruses also figure in our tale, but Rous sarcoma virus again has the leading role, yet another of many tributes to the pioneering work of Peyton Rous and to the principle of delayed gratification in science. The product of his diligence in pursuing a single chicken tumor nearly eighty years ago (1), Rous' virus remains the only retrovirus that could have satisfied the genetic and biochemical criteria for the work we accomplished in the era that preceded molecular cloning.

## FIRST TASTE OF MOLECULAR BIOLOGY: HYBRIDIZATIONS WITH THE LAC OPERON

My commitment to experimental science occurred, by today's standards, dangerously late in a prolonged adolescence. As an undergraduate at Amherst College, I was devoted to Dickensian novels and anti-Establishment journalism, while marginally fulfilling premedical requirements. I then indulged myself with a year of Anglo-Saxon and metaphysical poetry at Harvard graduate school, before beginning medical studies at Columbia

University, with a primary interest in psychiatry. But my ambitions soon turned towards an academic career in internal medicine. So just after graduation in 1966, like many of my contemporaries, I applied for research training at the National Institutes of Health. Perhaps because his wife was a poet, Ira Pastan agreed to take me into his laboratory, despite my lack of scientific credentials.

At the time, Ira was studying the biochemical effects of thyroid stimulating hormone on tissue slices, a subject close enough to clinical endocrinology not to be intimidating. But one day, while still an intern at Columbia-presbyterian Hospital, I received a telephone call from Ira, telling me that a lecture by Earl Sutherland had inspired him to begin work on the effects of cyclic AMP on regulation of the *lac* operon in *E.coli*. Late that night, alone in the house staff library, I peered for the first time into the *Journal of Molecular Biology* - it is no small tribute to Columbia that this journal was there - and attempted to read the seminal papers on the *lac* operon by Jacob and Monod (2). I knew then that, one way or another, my life was about to change.

Science is largely the making of measurements, and I soon learned from Ira how much more important a new measurement could be than an old theory. He and Bob Perlman had just discovered that cyclic AMP reversed catabolite repression of the *lac* operon (3). They suggested that I use the relatively new technique of molecular hybridization to ask whether regulation by cyclic AMP occurs at the transcriptional level. Apart from the pleasure of just getting results (as Gunther Stent has said, results are wonderful because they give us something to talk about (4) these measurements had enormous intellectual appeal, because they very simply resolved the ambiguity of hypothesis, demonstrating unequivocal changes in synthesis of *lac* messenger RNA (5). Furthermore, they were carried out with technical subtleties that ultimately shaped the way I later thought about the problems of detecting single genes in more complex, eukaryotic cells. We annealed radiolabeled *E.coli* RNA to filter-bound DNA from a pair of bacteriophages that differed only by the presence or absence of the *lac* operon; and we minimized irrelevant hybridization by including, as competitor, unlabeled RNA from an *E.coli* mutant from which the *lac* operon was deleted. An aesthetic merger of genetics with molecular biology, itself as pleasing as the results!

#### INTRODUCTION TO THE PROVIRUS AND THE VIROGENE-ONCOGENE HYPOTHESES

A major feature of life at the NIH in late 1960's was the extraordinary offering of evening courses for physicians attempting to become scientists as they neared thirty. Two classes had direct and specific effects on my subsequent work because they introduced me to important problems I believed approachable with the methods I had acquired in my brief apprenticeship.

Like many of my peers, I was excited by the prospect of applying reduc-

tionist methods to eukaryotic organisms, particularly in a way that might be informative about human disease. From some dilatory reading in the early 1960's, I knew enough about viruses and their association with tumors in animals to understand that they might provide a relatively simple entry into a problem as complex as cancer. In fact, for anyone interested in the genetic basis of cancer, viruses seemed to be the only game in town. What surprised and beckoned me were two rather simple but heretical hypotheses that described curious ways the genes of RNA tumor viruses might mingle with the chromosomes of host cells.

The more daring of these two hypotheses was the provirus hypothesis, first enunciated by Howard Temin (6). (John Bader, one of our NIH lecturers, was among the few others to espouse it in public (7).) The provirus hypothesis stated that the genes of RNA tumor viruses were copied into DNA, which became stably associated with the host cell; the proviral DNA then provided the information for production of new virus particles. With its existence supported principally — some said feebly — by studies with inhibitors of DNA and RNA synthesis, and its plausibility doubted in the absence of any precedent for information transfer from RNA to DNA, the provirus seemed to be a provocative target for a definitive decision with molecular hybridization.

The other hypothesis, the virogene-oncogene hypothesis, was more complex (8). George Todaro and Robert Huebner proposed that normal cells must contain genes related to those found in RNA tumor viruses, since viral proteins could often be found in cells of apparently uninfected animals, especially chickens and mice. Such genes, known as virogenes, were believed to be transmitted vertically as components of chromosomes, expressed in response to a variety of agents, and acquired by infection of germ cells at some time in the past. Since some RNA tumor viruses were known to be highly oncogenic, it was also proposed that tumor-inducing genes of such viruses (viral oncogenes) might also be transmitted through the germ line as a consequence of ancient infection. Activation of these endogenous viral oncogenes by substances we recognize as carcinogens — chemicals, radiation, other viruses — could serve to initiate a neoplastic process.

#### TRANSITION TO RNA TUMOR VIROLOGY IN REVOLUTIONARY TIMES

During the summer of 1969, I combined a backpacking vacation in California with a search for a suitable place to study tumor viruses. Acting on a tip from Harry Rubin in Berkeley, I sought out a small group, composed of Mike Bishop, Leon Levintow, and Warren Levinson, that was beginning to work with Rous sarcoma virus at UC San Francisco. (Rubin, it should be said, was more eager for me to meet Peter Duesberg, but Duesberg was out of town.) A brief conversation with Mike was sufficient to convince me of our intellectual compatibility (happily, one of the few convictions to have survived twenty years in this field), and I made plans to join the UCSF group as a post-doctoral fellow the following summer.

Before the intervening year had passed, however, two major discoveries changed the landscape of tumor virology. Satoshi Mizutani and Temin (9) and David Baltimore (10) found the predicted enzyme, reverse transcriptase, in virus particles, thereby erasing most of the skepticism about the provirus hypothesis by providing a means to synthesize the heretical DNA copy of an RNA genome. And Steve Martin isolated a crucial mutant of Rous sarcoma virus (11), one that lost its ability to transform cells at elevated temperature and regained it when the temperature was reduced. Martin's mutant offered the first clear definition of the gene we later called *src*, and it showed that the gene — and, by implication, a protein the gene encoded — was required to instigate and sustain the transformed state. Since the mutant virus grew normally at the temperature that blocked transformation, oncogenic and replicative functions could be dissociated, a facet of the story that will soon resurface.

#### FIRST FORAYS WITH RSV: SEEKING PROVIRAL DNA

Reverse transcriptase was properly greeted as strong evidence for the provirus hypothesis, and defused the urgency of challenging it. Yet it still seemed important to detect the provirus directly, most obviously by molecular hybridization, and to follow the pathway of its synthesis, especially in infected cells, not just *in vitro*. Reverse transcriptase obligingly offered a means to simplify the job, through the synthesis of potentially powerful probes, radioactive virus-specific DNA copied from a template of viral RNA.

Some of my initial efforts to proceed with these problems look, in retrospect, frustrating, if not quixotic — though the results were published in prominent journals. We chose at first to use double-stranded products of the RSV reverse transcriptase as hybridization probes, in order to measure gene copies through the accelerating effects of cellular DNA on reassociation kinetics (12). However, the RNA template was unevenly copied by reverse transcriptase *in vitro* (13), so that the products had complicated reannealing kinetics and did not uniformly represent the viral genome (whose genetic composition was in any case still unknown.) When I attempted to measure RSV-related DNA in the most obvious settings — uninfected and RSV-infected chicken cells — I found multiple copies of virus-related DNA in the normal cells, in apparent confirmation of at least some aspects of the virogene-oncogene hypothesis (14). But I was unable to detect the anticipated increment of RSV DNA in infected chicken cells, until I switched to the use of single-stranded DNA as probes (15). But by then Paul Neiman had already measured the increment by hybridization with radiolabeled RNA from virions (16). And, in the meantime, Hill and Hillova had provided more dramatic support for the provirus hypothesis in an entirely different way, by DNA transfection: addition of DNA from RSV-infected cells to new cells allowed the recovery of the original virus (17). So the entire viral genome must have been present in the DNA of infected cells.

Our approach to the provirus was eased when I abandoned chicken cells,

the traditional hosts for RSV in culture, in favor of cells from other birds — ducks and quail — and from mammals (18). Because we could detect little virus-related DNA in these cells prior to infection, it was relatively simple to measure new copies of RSV DNA following infection, to follow the time course of DNA synthesis, to show that reverse transcription occurred in the cytoplasm, and to define linear, circular, and integrated forms of viral DNA (19).

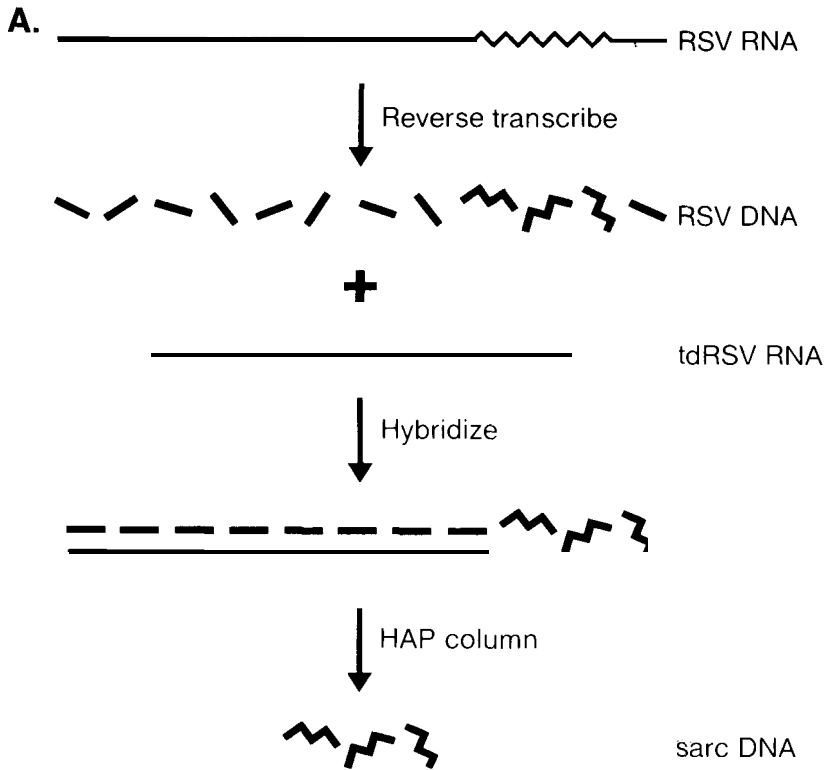
#### MAKING A PROBE TO TEST THE VIROGENE-ONCOGENE HYPOTHESIS

The varied abilities of normal avian DNAs to anneal to RSV-derived probes helped to focus our attention upon the sorts of virus-related sequence we could detect in chicken DNA. Did these sequences constitute genes for viral structural proteins? More importantly, did they include the viral transforming gene, as predicted by the oncogene-virogene hypothesis? To approach these questions it was imperative to have more rigorously defined probes. This was not a trivial challenge in the early 1970's, before restriction mapping and molecular cloning were available to us.

But one potent reagent was available. In 1971, Peter Vogt reported the isolation of transformation-defective, replication-competent mutants of RSV (20). The genomic RNA subunits of these "td" mutants were shown by Duesberg and his colleagues to be about 15 percent shorter than the subunits of wildtype virus (21). The provisional interpretation was that the missing sequence (initially called "x" and later "sarc") included some or all of the viral transforming gene (*v-src*) earlier defined by temperature-sensitive mutants. Like Martin's *ts* mutants, the deletion mutants retained the functions required for replication, despite the extensive loss of sequence, so it was tempting to presume that the deletion was coextensive, or nearly coextensive, with the transforming gene.

Mike and I were intimately acquainted with these conjectures through a collaborative consortium of Californian laboratories, directed by Vogt, Duesberg, and us, which met every six weeks or so, in Los Angeles or the Bay Area. Through these discussions, we recognized that if we could prepare radioactive DNA specific for the sequences deleted in the td-RSV mutants, we would have a reagent that would approximate a specific probe for the transforming gene of RSV.

The strategy for doing this was straightforward in principal, but difficult in practice (Fig. 1A). In essence, single-stranded, radiolabeled DNA fragments were synthesized from a template of wild-type RSV RNA, then hybridized to td-RSV RNA to remove unwanted components by hydroxylapatite-chromatography, leaving the sarc-specific DNA. Ramareddy Guntaka first put this protocol into motion with some encouraging results. But it was ultimately the ministrations of Dominique Stehelin that produced a sarc probe that met rigorous standards: nearly complete annealing to RSV RNA, no significant annealing to td-RSV RNA (Fig. 1B), and representation of

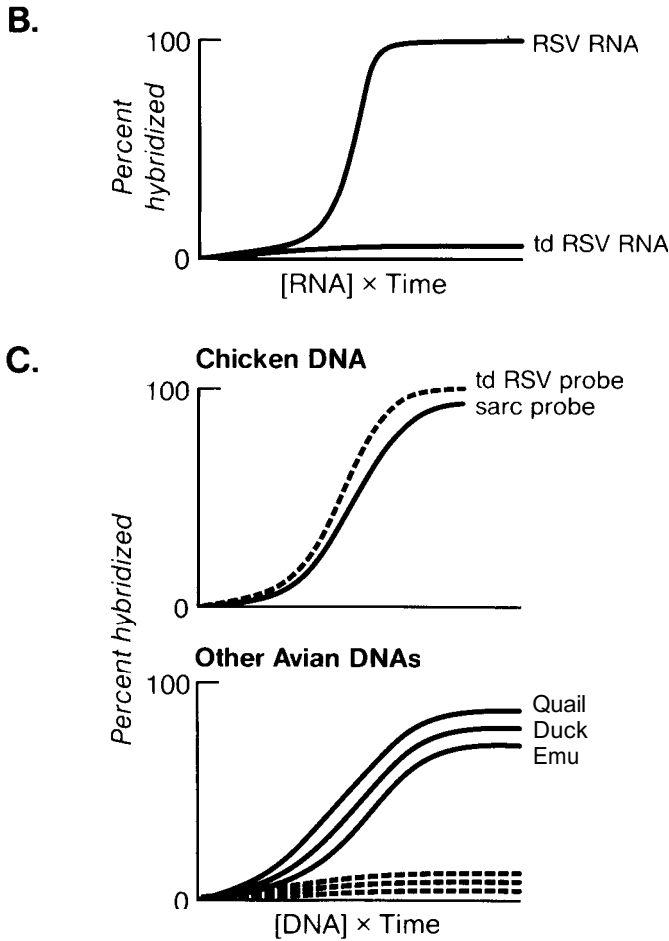


*Figure 1.* Schematic summary of initial experiments with sarc probe (see refs. 22 and 23 for primary data and further details). Panel A: Radiolabeled sarc-specific DNA was prepared by subtractive hybridization of the products of reverse transcription of RSV RNA to RNA from a transformation-defective deletion mutant of RSV (td RSV). The basis for the strategy is described in the text. Thin lines represent RNA, thick lines represent DNA, jagged portions represent sarc sequences (i.e. those present in RSV but not in td RSV genomes). HAP, hydroxylapatite. Panel B (next page): sarc DNA is specific for sequences that differentiate RSV and td RSV. The probe was hybridized to RSV RNA and td RSV RNA and results monitored by HAP chromatography. Panel C (next page) : sarc probe (solid curves) anneals to DNA from many species of birds, whereas probe for other components of RSV genome (td RSV probe, dashed curves) anneals poorly to DNA from species other than chicken. The extent of annealing (normalized values shown here) was determined by HAP chromatography.

over 10 percent of the RSV genome, most of the sarc region, in the probe (22).

#### SARC PROBE DETECTS CONSERVED SEQUENCES IN AVIAN DNA

When Stehelin incubated sarc probe with normal chicken DNA, it annealed extensively (as, of course, did probe made from td-RSV RNA) (Fig. 1C). The results, unambiguously exciting, were still fully consistent with the original oncogene-virogene hypothesis. So we were yet more excited when the next results seemed to violate it: although the "virogene" probe from td-RSV annealed poorly to DNAs from several other avian species, the sarc probe annealed extensively, even to DNA from the Australian emu — a rattite (we



learned, from our Berkeley colleague, Allan Wilson) at a great evolutionary distance from chickens (23). The extent and fidelity of the hybrids formed with sarc probe indicated that its homologs in normal cells had diverged during avian evolution at a rate similar to that of cellular genes used in the few earlier forays into molecular evolution, suggesting that the sequences had been conserved for at least 100 million years.

From these findings, we drew conclusions that seem even bolder in retrospect, knowing they are correct, than they did at the time (23). We said that the RSV transforming gene is indeed represented in normal cellular DNA, but not in the form proposed by the virogene-oncogene hypothesis. Instead, we argued, the cellular homolog is a normal cellular gene, which was introduced into a retroviral genome in slightly altered form during the genesis of RSV. Far from being a noxious element lying in wait for a carcinogenic signal, the progenitor of the viral oncogene appeared to have a function valued by organisms, as implied by its conservation during evolution. Since the viral *src* gene allows RSV to induce tumors, we speculated that its cellular homolog normally influenced those processes gone awry in tumorigenesis, control of cell growth or development.

FIRMING UP THE RESULT: SARC REPRESENTS THE C-SRC  
PROTO-ONCOGENE

Despite the broad claims, the first round of experiments with sarc probe left many worrisome questions unanswered.

The most pressing question, and one foremost in the minds of our critics, seems now both essential and mundane: Was the sarc probe actually detecting a functional, protein-encoding homolog of the viral transforming gene (*v-Src*)? or were the still ill-defined genetic and physical maps of the RSV genome leading us astray? Some support came from geneticists who mapped a large number of the existing transformation mutations of RSV within the region of the viral genome lost during formation of td-RSV deletion mutants (24). More exciting and stronger support came from protein biochemists: Joan Brugge and Ray Erikson discovered that the long-sought product of *v-src* was a protein of about 60,000 daltons (25), one that would require about 1600 nucleotides of coding sequence and could account for most of what was missing from td-RSV. Hermann Oppermann and others (26) then detected a protein in normal cells that seemed virtually indistinguishable from *v-src* protein, confirming the idea that sarc probe was measuring a gene (now called *c-src*) that resembled *v-src*. Ultimately, the molecular cloning and nucleotide sequencing of the RSV genome revealed how fortunate we had been in the design of our probe (27): Most td-RSV mutants lack all of *v-src* and little else.

The second question was more subtle: Did the conservation of *c-src* during avian speciation accurately imply that it was a cellular gene? or might it still represent an inherited viral gene more conserved than other viral elements? Answers came from several quarters, all confirming the arguments based on evolution. Using chicken chromosomes fractionated according to size by Elton Stubblefield in Texas, we found that *c-src* and virogenes are unlinked; the viral genes we could detect were on one or more large chromosomes, but *c-src* was on a small chromosome (28). Steve Hughes then used restriction enzymes to gauge the diversity of sequences in and around viral genes and *c-src* in many individual chickens (29); the pattern generated with sarc probe was monotonous, as would be expected for a conserved cellular gene (and shown to be the case for genes such as globin, ovalbumin, and others). The pattern produced with a probe for viral structural genes, however, suggested variety in number and context, as though they had been introduced into the chicken genome by recent, independent germ line infections. When we examined the transcripts emanating from *c-src* and from virus-related genes, individual chicken embryos contained various amounts and types of viral RNAs but similar quantities of a single, differently-sized species of *c-src* RNA (30). The most powerful evidence for the cellular nature of *c-src* required molecular cloning. For then it was possible to show that the coding sequences of *c-src* were interrupted in many places by introns (31), in the manner recently discovered to be characteristic of cellular genes. In contrast, as described in greater detail below, endogenous virogenes have the insignia of proviruses, being composed of continuous coding domains, flanked by repeated sequences.



The third question was most informative about the mechanism by which *v-src* causes cancer: What accounted for the proposed physiological differences between a beneficial proto-oncogene and a pathogenic viral oncogene derived from it? From the first measurements of *src* gene expression, it was apparent that the viral gene, controlled by a potent viral transcriptional promoter, was expressed much more vigorously than its cellular counterpart (32). But the levels of *v-src* protein required for transformation proved to be lower than non-oncogenic amounts of *c-src* protein (33), implying qualitative differences as well. Mark Collett in Erikson's laboratory (34) and Art Levinson in ours (35) had discovered that *src* proteins are protein kinases, which Tony Hunter and Bart Sefton later showed to be specific for tyrosine residues (36). Saburo Hanafusa's laboratory then defined the subtle structural and physiological differences between the viral and cellular versions of the gene (37): at least three of several aminoacid differences between  $p60^{v-src}$  and  $p60^{c-src}$  enhance the protein-tyrosine kinase activity of the transforming protein. Thus, quantitative and qualitative factors conspire to produce the srconcogene.

Finally, how well conserved is the cellular *src* gene? Early on, Deborah Spector showed that under conditions of reduced stringency most or all the sarc probe could anneal to the genomes of all vertebrates, not just birds (38). Since the implicated mammals included man, these findings helped to create a larger audience for our work, and they raised the possibility that retroviral proto-oncogenes might have a role on human cancer. New technologies ultimately extended the list of organisms that carry *c-src* to include virtually all metazoans — insects (39), worms (40) sponges (41), and hydras (42) — reminders of our evolutionary origins that are at once exhilarating and sobering.

The *src* story remains unfinished. We cannot tell you how *c-src* benefits normal organisms or cells, although recent work implicates *c-src* in both development, especially in the central nervous system (43), and in growth control during mitosis (44). We do not know the physiological targets for the *src* kinase, although numerous phosphotyrosine-containing proteins have been identified (45). And we do not know how the enzymatic activity of p60 is regulated, although phosphorylation is important (46). Nevertheless, the *src* paradigm has stimulated our field to move in several directions: to identify many new viral oncogenes and their cellular progenitors (47), to characterize a stunning variety of oncogenic proteins (48), to make unexpected connections with elements of growth regulatory networks (49), and to describe mutant proto-oncogenes in human tumors (50). These developments are recounted in the accompanying lecture by Mike Bishop. It is my mission to stay with the virus — and especially the provirus.

#### DECIPHERING PROVIRAL STRUCTURE

By the early 1970's the provirus was a well-accepted idea, but the organization of viral DNA and its position within chromosomes were still matters of conjecture. Several peculiarities of viral RNA and the viral life cycle hinted

that proviral DNA must have special attributes (19). First, the priming site for the first strand of viral DNA was near the 5' rather than at the 3' end of viral RNA (51), implying that synthesis must be a complex process and that the provirus must not be a simple copy of viral RNA. Next, a short sequence (R) was found at both ends of viral RNA, and hence present in two copies, but appeared to be copied only once during synthesis of viral DNA (52); how was the second copy of R regenerated? Finally, it was difficult to account for the efficient synthesis of viral RNA without the prospect of a strong transcriptional promoter upstream of the start site; how was that promoter provided?

These problems were solved by the unexpectedly elegant configuration of viral DNA, as worked out mainly by Peter Shank, Steve Hughes, and Hsing-Jien Kung in our group (53) and independently by John Taylor's laboratory in Philadelphia (54). Once again, RSV was the instrument of discovery, and again the results depended upon hybridization with specific probes, this time for terminal regions of the viral genome. In essence, viral genes were found to be flanked in the provirus by long terminal repeats (LTRs) derived from sequences present at both ends of viral RNA (Fig. 2). (The ends of the LTRs correspond to the priming sites for the two DNA strands and thereby helped unravel a strategy of DNA synthesis too convoluted to review here (19).) Because the R sequence is present once in each LTR, it can be reconstituted by transcribing parts of both LTRs. And viral

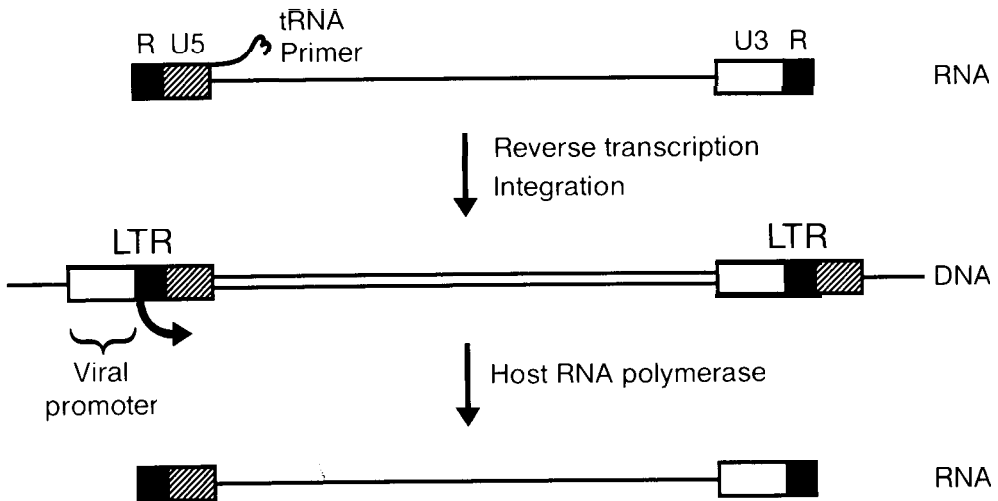


Figure 2. The organization of proviral DNA in comparison to retroviral RNA. The top line shows one subunit of a viral dimeric genome, with host tRNA positioned near the 5' end where it serves to prime synthesis of the first strand of viral DNA. R, short sequence present at both ends of viral RNA; U5 and U3 are sequences unique to the 5' and 3' regions of the RNA that are duplicated during DNA synthesis to form the long terminal repeats (LTRs). The middle line shows a provirus integrated into host cell DNA (single line). The viral coding sequences reside between the LTRs (double line). The region encompassing the viral promoter in U3 of the upstream LTRs is bracketed; the curved arrow denotes the start site and direction of transcription of the provirus by host RNA polymerase. The bottom line shows the composition of the primary viral transcript after 3' processing; the poly(A) tract at the 3' end is not illustrated.

sequences that contain strong transcriptional signals reside upstream of the RNA start site. Mapping of integration sites showed that many regions of the host genome could accommodate a provirus; thus, transcriptional self-sufficiency of the provirus allowed it to function in varied chromosomal contexts.

### PROVIRUSES AS MOBILE ELEMENTS THAT CAUSE INSERTION MUTATIONS

But the structure of the provirus did more than solve some perplexities of the retrovirus life cycle. In general form and even in selected short sequences, proviruses resemble an abundant type of mobile DNA element (Fig. 3), described by now in plants, bacteria, yeast, insects, and many other organisms (55), another arresting example of conservation throughout evolution. Their connection with retroviruses has been strengthened in recent years by discoveries that several such elements are duplicated and relocated by using reverse transcriptases to make new DNA copies from RNA transcripts (56), although they never produce extracellular viruses. These properties also apply to most endogenous proviruses, cloned from the germ lines of many vertebrates (57), reemphasizing the profound differences between inherited virogenes and cellular proto-oncogenes.

One practical consequence of the startling similarities between retroviral proviruses and mobile elements was to consider the possibility that proviruses, like mobile DNA, might cause insertion mutations. In 1978, while

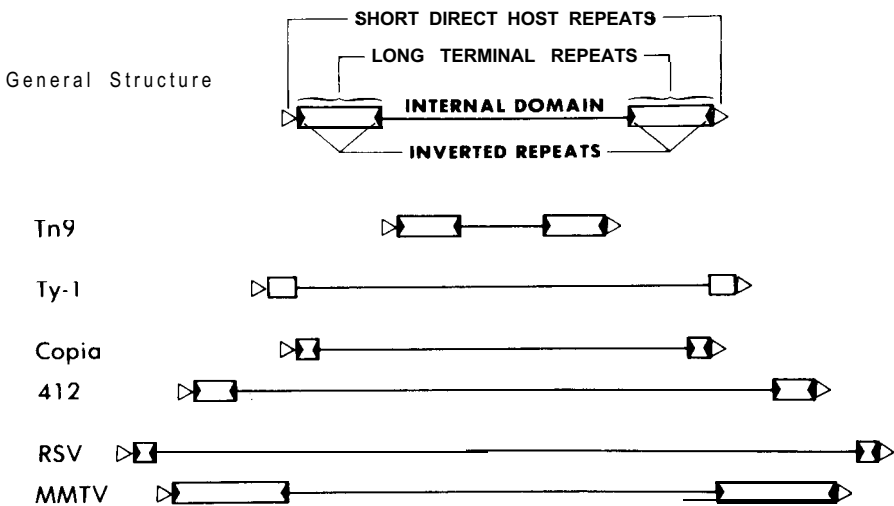


Figure 3. Many mobile DNAs are organized like proviruses. The figure demonstrates some common features of many transposable elements, including LTRs (rectangular boxes), inverted repeats within the LTRs (closed triangles), and short duplications in host DNA generated during insertion (open triangles). The illustrated mobile elements included retroviral proviruses (RSV and mouse mammary tumor virus MMTV), retrotransposons of *Drosophila* (copia and 412) and budding yeast (Ty1), and a conventional transposon of *E.coli* (Tn9). (Reprinted with permission of Academic Press; see ref. 55).

on sabbatical in Mike Fried's laboratory at the Imperial Cancer Research Fund, I designed an experiment to test this idea. John Wyke provided me with a rat cell line transformed by a single RSV provirus, which could serve as a target for insertion mutation by proviruses introduced by superinfection with mouse leukemia virus (MLV). By sifting through many clones of cells that had lost their transformed properties after infection with MLV, Suzanne Ortiz and I found two that contained an MLV provirus inserted at different sites within the pre-existing RSV provirus, interfering with the expression of *v-src* (58).

This experiment established the principle that retroviruses could serve as insertional mutagens to inactivate genes. It also had the heuristic benefit of stimulating us to think about insertion mutations that acted in a dominant fashion by activating gene expression. Greg Payne was then attempting to explain how avian leukosis virus (ALV), a virus virtually indistinguishable from td-RSV and lacking any evidence of a viral oncogene, could nevertheless induce tumors (most commonly B-cell lymphomas) within several weeks after infection of susceptible chickens (59). Might ALV proviruses occasionally integrate adjacent to a cellular proto-oncogene and augment expression through a viral LTR? Greg's evidence for this idea (60), however provocative, was nearly drowned out by the commotion caused by Hayward, Neel, and Astrin's discovery (61) that ALV DNA in B cell lymphomas was adjacent to *c-myc* — a known progenitor of a retroviral oncogene (62) — and that the viral LTR was driving *c-myc* expression.

ALV-induced tumors taught us several new principles: Retroviruses can induce neoplasia by insertionally activating proto-oncogenes (63); proviruses and their target genes can be variously arranged with similar effects on transcription (64); and proto-oncogenes do not need to be transduced to participate in oncogenesis. The last was an especially important point that presaged the later outpouring of mutant proto-oncogenes in human tumors unassociated with any virus (50).

#### USING PROVIRUSES AS TRANSPOSON TAGS FOR NOVEL PROTO-ONCOGENES: THE INT-1 STORY

However important, ALV has failed to introduce us to any proto-oncogenes not already known as forefathers of retroviral oncogenes. For this, we made use of another retrovirus without a viral oncogene, the mouse mammary tumor virus (MMTV). Like RSV, MMTV has a venerable history (65). Found in the milk of inbred mice with a high incidence of mammary cancer over fifty years ago in Holland (66) and at the Jackson Laboratories in Maine (67), MMTV was the first mammalian retrovirus to be discovered; it remains the only efficient viral agent of mammary carcinoma, and thus a model for one of the most common of human cancers.

MMTV-induced mammary tumors are quasi-clonal growths of virus-infected cells (68). To ask whether the tumor cells result from insertion of viral DNA near a heretofore unknown proto-oncogene, Roel Nusse examined many tumors to find one with only a single new provirus; he then

cloned that provirus and its flanking cellular DNA in *E.coli*. An unfamiliar gene, which we called *int-1*, was nearby, and it was expressed in that tumor and several others with nearby insertions, but not in normal mammary glands (69).

But this was not sufficient to implicate *int-1* as a oncogene. First there was the circumstantial force of repetition: over three-quarters of mammary tumors in the C3H mouse strain harbor insertion mutations in the *int-1* locus. Then Tony Brown did what nature had not done, by placing the *int-1* gene within a retroviral genome; the resulting virus alters the growth and morphology of cultured mammary cells (70). Finally, Ann Tsukamoto followed a strategy pioneered by Ralph Brinster and Richard Palmiter and by Philip Leder (71) and introduced the *int-1* gene, linked to an MMTV LTR, into the mouse germ line (72). All the transgenic mice, male or female, experience dramatic hyperplasia of the mammary epithelium, and most of the females develop mammary carcinoma within six months. This is about as close as we can come to fulfilling Koch's postulates for a genetic disease: by placing the virally-mutated form of the gene into the germ line — ironically, much as envisioned to occur naturally in the virogene-oncogene hypothesis — we have recreated the disease.

I cannot leave our transgenic mice without making a more general point. In California and many other places, misguided efforts to abolish the use of laboratory animals seriously threaten medical science. If Peyton Rous had been denied his chickens, our field would have no past; if all of us are now denied mice and other animals, it will have little future.

#### A TENTATIVE SCHEME FOR TRANSDUCTION OF PROTO-ONCOGENES

*int-1* is but the first entry on a now substantial list of proto-oncogenes discovered as loci repeatedly activated by proviruses in tumors (63). Thus, retroviruses usher in the genetic cast in the drama of cancer in two ways: by transduction and insertion mutation. Not surprisingly, the two phenomena appear to be mechanistically related: insertion mutation is probably the first step in the sequence of events that occasionally spawns a viral oncogene as its end product. What we can predict, but not yet fully substantiate by direct observations, is that two recombination events are required for transduction (Fig. 4; 73). The first occurs during proviral integration, placing viral DNA upstream from the activated cellular gene that will be acquired. The second occurs during virus replication in the tumor that results from the insertion mutation; the second step joins viral sequences to cellular sequences derived from the downstream region of the gene. We suppose that more or less in this fashion a close relative of ALV acquired a slightly mutated version of a chicken's *src* gene nearly a century ago, and set us on a path we are still travelling.

#### THE PROSPECTS FOR RETROVIROLOGY

The story thus far confirms David Baltimore's statement of thanksgiving

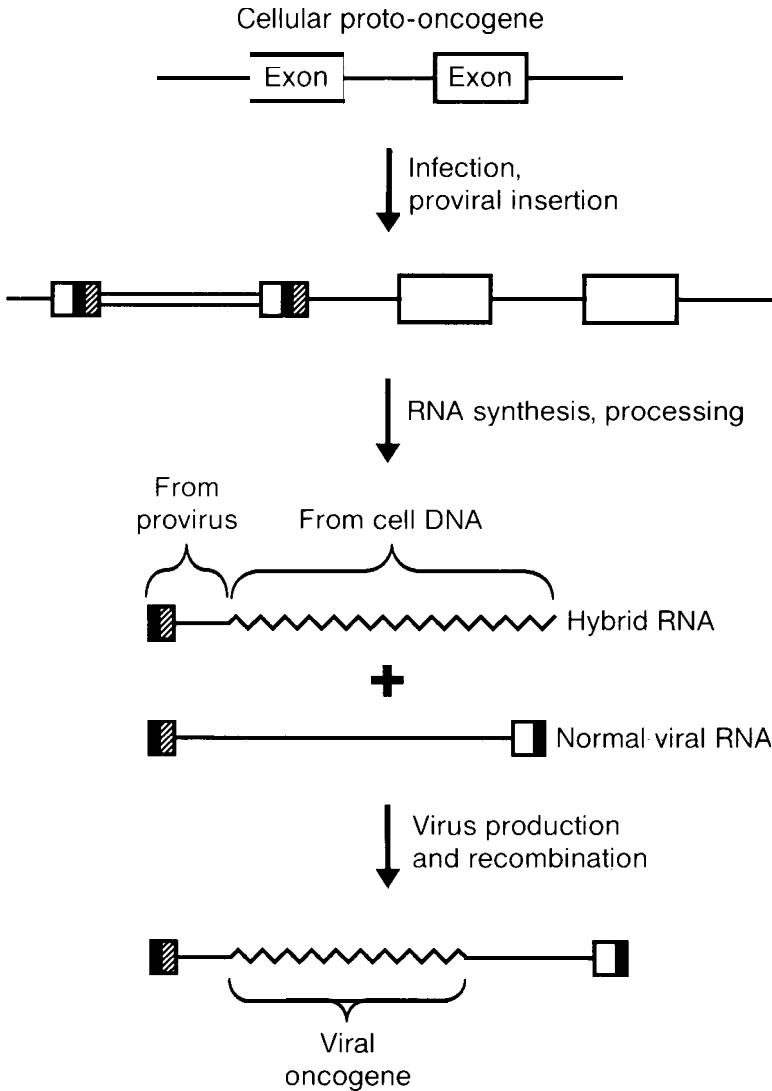


Figure 4. Model for transduction of cellular proto-oncogenes to form retroviral oncogenes. Exons of a proto-oncogene are located downstream of a retroviral provirus recently introduced by infection and denoted as in Fig. 2. Virus-host chimeric RNA, a product of the proviral insertion mutation, recombines with normal viral RNA during virus replication to join viral sequences downstream of the cellular sequences. For further details, see Ref. 73.

(74): "a virologist is among the luckiest of biologists because he can see into his chosen pet down to the details of all its molecules." Because retroviruses, our chosen pets, are such remarkable agents, it has been enough to train our sights on two brief questions — how do retroviruses grow? how do retroviruses cause cancer? — to have extended our concerns outward to the cellular host, as well as to have focused them inward upon the viruses themselves (75). As a result, we have entered into some of the liveliest

arenas in modern biology: the genetic basis of cancer, the transposition of DNA through RNA intermediates, the control of gene expression in eukaryotes, and the molecular evidence for evolution.

At this point, the study of oncogenes and proto-oncogenes has attained a degree of maturity that allows it to be conducted with astonishing little virology. Yet retroviruses remain vital tools for the isolation of important new oncogenes; witness in the past few years the discoveries of the *jun* and *crk* genes (76). Likewise, since the discovery of reverse transcriptase nearly two decades ago, seemingly exhaustive attention has been given to the life cycle of retroviruses (19), yet many central features are just now coming into view (75). Cell surface receptors for viral attachment and entry have been recently identified and show a remarkable range of biochemical properties (77); the proviral integration reaction has been recapitulated in vitro with nucleoprotein complexes (78), allowing a description of integrative precursors and intermediates (79); retroviruses have been recognized as pliable genetic vectors (80) that may one day be used clinically to correct gene deficiencies, in the manner used in nature to transport host-derived oncogenes; many unexpected aspects of viral gene expression have been discovered, including translational frameshifting during the synthesis of reverse transcriptase (81) and complex viral regulatory genes that govern the behavior of two classes of human retroviruses (82); and the principles of virus assembly are emerging through physical and genetic assaults on viral structural proteins and proteases (83). These inherently fascinating problems have now taken on a special urgency, because we are all threatened by the world-wide dissemination of a lethal human retrovirus, the human immunodeficiency virus (84). Thus retroviruses continue to challenge our intellects in ways that may help us grapple with major diseases, cancer and now AIDS, while also revealing fundamental features of the lives of our cells.

#### ACKNOWLEDGEMENTS

My indebtedness is large: to the many coworkers and colleagues mentioned here, and to many more unnamed, for communal labor, ideas, and criticism; to several institutions (especially the National Institutes of Health, the American Cancer Society, and the University of California, San Francisco) for nurturing my career with financial support and other resources; and to my family, for affectionate tolerance. My approach to science has been most profoundly influenced by Ira Pastan, who showed me that laboratory life was preferable to the clinic; by Peter Vogt, who taught me viral genetics when I needed it; and by Mike Bishop, who has been a generous and challenging colleague for nearly two decades. My gratitude would be incomplete if I did not especially mention the friendship and assistance of Suzanne Ortiz, Nancy Quintrell, Jean Jackson, Leon Levintow, Warren Levinson, and Don Ganem.

## REFERENCES

1. P. Rous, *J. Exp. Med.*, **132**, 397 (1911); *Science* **157**, 24 (1967).
2. F. Jacob, J. Monod, *J. Mol. Biol.* **3**, 318 (1961); F. Jacob, A. Ullmann, J. Monod, *ibid* **13**, 704 (1965).
3. R. L. Perlman, B. DeCrombrugge, I. Pastan, *Nature* **223**, 810 (1969); R. Perlman, B. Chen, B. DeCrombrugge, M. Emmer, M. Gottesman, H. Varmus, I. Pastan, in *Cold Spring Harbor Symp. Quant. Biol.* **35**, pp. 419-423 (1970).
4. G. Stent in *A Passion for Science*, L. Wolpert, A. Richards, Oxford University Press, pp. 109-119 (1988).
5. H. E. Varmus, R. L. Perlman, I. Pastan, *J. Biol. Chem.* **245**, 2259 (1970); H. E. Varmus, R. L. Perlman, I. Pastan, *J. Biol. Chem.* **245**, 6366 (1970).
6. H. M. Temin, *Nat. Cancer Inst. Monograph* **17**, 557 (1964); H. M. Temin, *Science* **192**, 1075 (1976).
7. J. P. Bader, *Virology* **26**, 253 (1965).
8. R. J. Huebner, G. J. Todaro, *Proc. Natl. Acad. Sci.* **64**, 1087 (1969).
9. H. M. Temin, S. Mizutani, *Nature* **226**, 1211, (1970).
10. D. Baltimore, *Nature* **226**, 1209 (1970).
11. G. S. Martin, *Nature* **227**, 1021 (1970).
12. L. Gelb, D. Kohne, M. Martin, *J. Mol. Biol.* **57**, 129 (1971).
13. H. E. Varmus, W. E. Levinson, J. M. Bishop, *Nat. New Biology* **223**, 19 (1971).
14. H. E. Varmus, R. A. Weiss, R. Friis, W. E. Levinson, J. M. Bishop, *Proc. Natl. Acad. Sci.* **69**, 20 (1972).
15. H. E. Varmus, S. Heasley, J. M. Bishop, *J. Virol.* **14**, 895 (1974).
16. P. E. Neiman, *Science* **178**, 750 (1972).
17. M. Hill, J. Hillova, *Nat. New Biol.* **237**, 35, (1972); M. Hill, J. Hillova, *Virology* **49**, 309 (1972).
18. H. E. Varmus, P.K. Vogt, J. M. Bishop, *J. Mol. Biol.* **74**, 613 (1973); H. E. Varmus, P. K. Vogt, J. M. Bishop, *Proc. Natl. Acad. Sci.* **70**, 3067 (1973).
19. H. Varmus, R. Swanstrom, Chapter 5 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982) pp. 369-512; H. Varmus, R. Swanstrom, Chapter 5S *ibid*, (1985), pp. 75-134.
20. P. K. Vogt, *Virology* **46**, 939 (1971).
21. G. S. Martin, P. H. Duesberg, *Virology* **47**, (1972); L.-H. Wang, P. Duesberg, K. Beemon, P. K. Vogt, *J. Virol.* **16**, 1051 (1975).
22. D. Stehelin, R. V. Guntaka, H. E. Varmus, J. M. Bishop, *J. Mol. Biol.* **101**, 349 (1976).
23. D. Stehelin, H. E. Varmus, J. M. Bishop, P. K. Vogt, *Nature* **260**, 170 (1976).
24. A. Bernstein, R. MacCormick, G. S. Martin, *Virology*, **70**, 206 (1976).
25. J. S. Brugge, R. L. Erikson, *Nature* **269**, 346 (1977).
26. M. S. Collett, J. S. Brugge, R. L. Erikson, *Cell*, **15**, 1363, (1978); H. Opper-  
mann, A. D. Levinson, H. E. Varmus, L. Levintow, J. M. Bishop, *Proc. Natl. Acad. Sci.* **76**, 1804 (1979); L. R. Rohrschneider, R. N. Eisenman, C. R. Leitch, *Proc. Natl. Acad. Sci.* **76**, 4479 (1979).
27. A. P. Czernilofsky, A. D. Levinson, H. E. Varmus, J. M. Bishop, E. Tischer, H. M. Goodman, *Nature* **287**, 198 (1980); T. Takeya, H. Hanafusa, *J. Virol.* **44**, 12 (1982).
28. T. Padgett, E. Stubblefield, H. E. Varmus, *Cell* **10**, 649 (1977); S. H. Hughes, E. Stubblefield, F. Payvar, J. D. Engel, J. B. Dodgson, D. Spector, B. Cordell, R. T. Schimke, H. E. Varmus, *Proc. Natl. Acad. Sci.* **76**, 1348 (1979).
29. S. H. Hughes, F. Payvar, D. Spector, R. T. Schimke, H. L. Robinson, J. M. Bishop, H. E. Varmus, *Cell* **18**, 347 (1979).
30. D. Spector, B. Baker, H. E. Varmus, J. M. Bishop, *Cell* **3**, 381 (1978); D.



- Spector, K. Smith, T. Padgett, P. McCombe, D. Roulland-Dussoix, C. Moscovici, H. E. Varmus, J. M. Bishop, *Cell* **3**, 371 (1978).
31. R. C. Parker, H. E. Varmus, J. M. Bishop, *Proc. Natl. Acad. Sci.* **78**, 5842 (1981); T. Takeya, H. Hanafusa, R. P. Junghans, G. Ju, A. M. Skalka, *Mol. Cell. Biol.* **1**, 1024 (1981); D. Shalloway, A. D. Zelenetz, G. M. Cooper, *Cell* **24**, 531 (1981).
  32. J. M. Bishop, H. Varmus, Chapter 9 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982) pp. 999-1108.
  33. R. P. Parker, H. E. Varmus, J. M. Bishop, *Cell* **37**, 131 (1984); E. B. Jakobovits, J. E. Majors, H. E. Varmus, *Cell* **38**, 757 (1984).
  34. M. S. Collett, R. L. Erikson, *Proc. Natl. Acad. Sci.* **75**, 2021 (1978).
  35. A. Levinson, H. Oppermann, L. Levintow, H. E. Varmus, J. M. Bishop, *Cell* **15**, 561 (1978).
  36. T. Hunter, B. M. Sefton, *Proc. Natl. Acad. Sci.* **77**, 1311 (1980).
  37. T. Takeya, H. Hanafusa, *Cell* **32**, 881 (1983).
  38. D. H. Spector, H. E. Varmus, J. M. Bishop, *Proc. Natl. Acad. Sci.* **75**, 4102 (1978).
  39. B.-Z. Shilo, R. A. Weinberg, *Proc. Natl. Acad. Sci.* **78**, 6789 (1981); M. A. Simon, B. Drees, T. Kornberg, J. M. Bishop, *Cell* **42**, 831 (1985).
  40. A. Kamb, M. Weir, B. Rudy, H. E. Varmus, C. Kenyon, *Nature* **337**, 364 (1989).
  41. A. Barnekow, M. Schartl, *Mol. Cell. Biol.* **4**, 1179 (1984).
  42. T. C. Bosch, T. F. Unger, D. A. Fisher, R. E. Steele, *Mol. Cell. Biol.* **9**, 4141 (1989).
  43. J. S. Brugge, P. C. Cotton, A. E. Quesada, J. N. Barrett, D. Nonner, R. W. Kenne, *Nature* **316**, 554 (1985); P. F. Maness, M. Aubry, C. G. Shores, L. Frame, K. H. Pfenninger, *Proc. Natl. Acad. Sci.* **85**, 5001 (1988); R. Martinez, B. Mathey-Provot, A. Bernards, D. Baltimore, *Science* **237**, 411 (1987).
  44. D. O. Morgan, J. M. Kaplan, J. M. Bishop, H. E. Varmus, *Cell* **57**, 775 (1989); S. Shenoy, J.-K. Choi, S. Bagrodia, T. D. Copeland, J. L. Maller, D. Shalloway, *Cell* **57**, 763 (1989).
  45. T. Hunter, J. A. Cooper, *Annu. Rev. Biochem.* **54**, 897 (1984).
  46. T. Hunter, *Cell* **49**, 1 (1987).
  47. J. M. Bishop, H. Varmus, Chapter 9S in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Springs Harbor, NY, 1985), pp. 249-356.
  48. H. E. Varmus, J. M. Bishop, Eds., *Cancer Surveys — Proteins Encoded by Oncogenes*, Volume 5, Number 2, (Oxford University Press, England, 1986).
  49. H. E. Varmus, Chapter 9 in *Molecular Basis of Blood Diseases*, G. Stamatoyannopoulos, A. W. Nienhuis, P. Leder, P. W. Majerus, Eds. (W. B. Saunders, 1987), pp. 271-346.
  50. H. E. Varmus, *Ann. Rev. Genetics*, **18**, 553 (1984); J. M. Bishop, *Science* **235**, 305 (1987).
  51. J. M. Taylor, R. Illmensee, *J. Virol.* **16**, 552, (1975).
  52. J. M. Coffin, W. A. Haseltine, *Proc. Natl. Acad. Sci.* **74**, 1908, (1977); M. S. Collett, P. Dierks, J. F. Cahill, A. J. Faras, J. T. Parsons, *Proc. Natl. Acad. Sci.* **74**, 2389 (1977); W. A. Haseltine, J. M. Coffin, T. C. Hageman, *J. Virol.* **30**, 375 (1979); R. P. Junghans, S. Hu, C. A. Knight, N. Davidson, *Proc. Natl. Acad. Sci.* **74**, 477 (1977); E. Stoll, M. A. Billeter, A. Palmenberg, C. Weissman, *Cell* **12**, 57 (1977).
  53. S. Hughes, P. K. Vogt, P. R. Shank, D. Spector, H.-J. Kung, M. L. Breitman, J. M. Bishop, H. E. Varmus, *Cell* **15**, 1397 (1978); P. R. Shank, S. Hughes, H.-J. Kung, J. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, H. E. Varmus, *Cell* **15**, 1383 (1978); H. E. Varmus, P. R. Shank, S. Hughes, H.-J. Kung, S. Heasley, J. Majors, P. K. Vogt, J. M. Bishop, *Cold Spring Harbor Symp. Quant. Biol.* **43**, 851 (1979).

54. T. W. Hsu, J. L. Sabran, G. E. Mark, R. V. Guntaka, J. M. Taylor, *J. Virol.* **28**, 810 (1978); J. L. Sabran, T. W. Hsu, C. Yeater, A. Kaji, W. S. Mason, J. M. Taylor, *J. Virol.* **29**, 170 (1979).
55. H. E. Varmus, Chapter 10 in *Transposable Elements*, J. Shapiro, Ed., (Academic Press, NY, 1983), pp. 411-503; H. E. Varmus, P. O. Brown, in *Mobile DNA* M. Howe, D. Berg, Eds., (ASM Publications, New York, 1989) pp. 35-56.
56. D. Baltimore, *Cell* **40**, 481 (1985); H. E. Varmus, *Nature* **314**, 583 (1985).
57. J. Stoye, J. Coffin, Chapter 10 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1985), pp. 357-404.
58. H. E. Varmus, N. Quintrell, S. Ortiz, *Cell* **25**, 23 (1981).
59. N. Teich, J. Wyke, T. Mak, A. Bernstein, W. Hardy, Chapter 8 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), pp. 785-998.
60. G. S. Payne, S. A. Courtneidge, L. B. Crittenden, A. M. Fadly, J. M. Bishop, H. E. Varmus, *Cell* **23**, 31 (1981).
61. W. S. Hayward, B. G. Neel, S. M. Astrin, *Nature* **290**, 475 (1981); B. G. Neel, W. S. Hayward, H. L. Robinson, J. Fang, S. M. Astrin, *Cell* **23**, 323 (1981).
62. D. Sheiness, J. M. Bishop. *J. Virol.* **31**, 514, (1979); M. Roussel, S. Saule, C. Lagrou, C. Rommens, H. Beug, T. Graf, D. Stehelin, *Nature* **281**, 452 (1979).
63. H. E. Varmus, *Cancer Surv.* **2**, 301 (1982); R. Nusse, A. Berns, Chapter 3 in *Cellular Oncogene Activation*, G. Klein, Ed., (Marcel Dekker, Inc., New York, NY), 1988.
64. G. S. Payne, J. M. Bishop, H. E. Varmus, *Nature* **295**, 209 (1982).
65. J. Hilgers, P. Bentvelzen, *Adv. Cancer Res.* **26**, 143, (1978).
66. R. Korteweg, *Genetics* **18**, 350 (1946).
67. J. J. Bittner, *Science* **84**, 162, (1936).
68. J. C. Cohen, P. R. Shank, V. Morris, R. Cardiff, H. E. Varmus, *Cell* **16**, 333 (1979).
69. R. Nusse, H. E. Varmus, *Cell* **31**, 99 (1982); R. Nusse, A. van Ooyen, D. Cox, Y. K. T. Fung, H. Varmus, *Nature* **307**, 131 (1984).
70. A. M. C. Brown, R. S. Wildin, T. J. Prendergast, H. E. Varmus, *Cell* **46**, 1001 (1986).
71. T. A. Stewart, P. K. Pattengale, P. Leder, *Cell* **38**, 627 (1984); R. D. Palmiter, R. L. Brinster, *Annu. Rev. Genetics*, **20**, 465 (1986); S. Cory, J. M. Adams, *Annu. Rev. Immunology* **6**, 25 (1988); D. Hanahan, *Science* **246**, 1265 (1989).
72. A. S. Tsukamoto, R. Grosschedl, R. C. Guzman, T. Parslow, H. E. Varmus, *Cell* **55**, 619 (1988).
73. H. E. Varmus, *Science* **216**, 812 (1982); J. M. Bishop, *Annu. Rev. Biochem.* **52**, 301 (1983); R. Swanstrom, R. C. Parker, H. E. Varmus, J. M. Bishop, *Proc. Natl. Acad. Sci.* **80**, 2519 (1983).
74. D. Baltimore, *Science* **192**, 632 (1976).
75. H. E. Varmus, *Science* **240**, 1427 (1988).
76. Y. Maki, T. J. Bos, C. Davis, M. Starbuck, P. K. Vogt, *Proc. Natl. Acad. Sci.* **84**, 2848 (1987); P. K. Vogt, T. J. Bos, R. F. Doolittle, *Proc. Natl. Acad. Sci.* **84**, 3316 (1978); B. J. Mayer, M. Hamaguchi, H. Hanafusa, *Nature* **332**, 272 (1988).
77. P. J. Maddon, A. G. Dagleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, R. Axel, *Cell* **47**, 333 (1986); J. S. McDougal, M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, J. K. A. Nicholson, *Science* **231**, 382 (1986); L. M. Albritton, L. Tseng, D. Scadden, J. M. Cunningham, *Cell* **57**, 659 (1989).
78. P. O. Brown, B. Bowerman, H. E. Varmus, J. M. Bishop, *Cell* **49**, 347 (1987).
79. T. Fujiwara, K. Mizuuchi, *Cell* **55**, 497 (1988); P. O. Brown, B. Bowerman, H. E. Varmus, J. M. Bishop, *Proc. Natl. Acad. Sci.* **86**, 2525 (1989).

80. J. Coffin, Chapter 4 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1985). pp. 17-74; E. Gilboa, *RioEssays* **5**, 252 (1987).
81. T. Jacks, H. D. Madhani, F. R. Masiarz, H. E. Varmus, *Cell* **55**, 447 (1988); T. Jacks, M.D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, H. E. Varmus, *Nature* **331**, 280 (1988).
82. H. Varmus, *Genes & Development* **2**, 1055 (1988); B. R. Franza, B. R. Cullen, F. Wong-Staal, Eds., *The Control of Human Retrovirus Gene Expression*, (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988).
83. C. Dickson, R. Eisenman, H. Fan, Chapter 6 in *RNA Tumor Viruses*, R. Weiss, N. Teich, H. Varmus, J. Coffin, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, 1985), pp. 135-146; A. M. Skalka, *Cell* **56**, 911 (1989).
84. *The Science of AIDS: Readings from Scientific American Magazine*, W. H. Freeman & Co., New York, 1988.