Articles

Retroviruses

HAROLD VARMUS

First brought to scientific attention as infectious cancer-causing agents nearly 80 years ago, retroviruses are popular in contemporary biology for many reasons. (i) The virus life cycle includes several events—in particular, reverse transcription of the viral RNA genome into DNA, orderly integration of viral DNA into host chromosomes, and utilization of host mechanisms for gene expression in response to viral signals—which are broadly informative about eukaryotic cells and viruses. (ii) Retroviral oncogenesis usually depends on transduction or insertional activation of cellular genes, and isolation of those genes has provided the scientific community with many of the molecular components now implicated in the control of normal growth and in human cancer. (iii) Retroviruses include many important veterinary pathogens and two recently discovered human pathogens, the causative agents of the acquired immunodeficiency syndrome (AIDS) and adult T cell leukemia/lymphoma. (iv) Retroviruses are genetic vectors in nature and can be modified to serve as genetic vectors for both experimental and therapeutic purposes. (v) Insertion of retroviral DNA into host chromosomes can be used to mark cell lineages and to make developmental mutants. Progress in these and other areas of retrovirus-related biology has been enormous during the past two decades, but many practical and theoretical problems remain to be solved.

RETROVIRUSES ARE SPECIAL ENTRIES ON THE MENU OF biological systems surveyed in this issue. This is so because retroviruses necessarily intersect with many other "systems" (for example, humans, other primates, transgenic animals, *Drosophila*, and yeast) through infection of cells, through strong similarities with mobile genetic elements that reside in eukaryotic chromosomes, or through intimate associations with cellular genes that are instrumental in retrovirus-induced cancers. Furthermore, retroviruses are unusual parasites in that they insinuate themselves into the life-styles of their hosts in revealing and often unprecedented ways: by converting their genes from RNA to a DNA form; by incorporating viral DNA stably into chromosomes of somatic or germ cells; by mutating, and even capturing, cellular genes; by rarely impairing, and often potentiating, the growth of their host cells; and by entrusting gene expression to host mechanisms under the direction of viral signals.

The objective of this selective review is to introduce readers to several of these themes, to evoke the flavor of the discipline of retrovirology, and to raise some questions that might attract the next generation of disciples. Retroviruses will be viewed here from three perspectives potentially attractive to those seeking exciting experimental prospects: (i) as models for the study of fundamental biological problems, including transfer of genetic information, DNA recombination, regulated gene expression, growth control, and macromolecular assembly; (ii) as problems posed by their pathogenic potential in human and animal hosts, where they cause diseases such as AIDS and many forms of cancer; and (iii) as tools for genetic manipulations ranging from gene therapy to mutagenesis.

The author is an American Cancer Society Professor of Molecular Virology in the Department of Microbiology and Immunology and the Department of Biochemistry and Biophysics at the School of Medicine of the University of California, San Francisco, CA 94143.

Growth and Development of the Retrovirus Community

Viruses of the type we now call retroviruses were among the earliest known viruses, first discovered about 80 years ago as filterable agents that cause cancers in chickens (*1*). For many years, however, they had a small following in the scientific community, due, in part, to the lack of reliable cell culture and biochemical techniques and, in part, to skepticism in some quarters about the significance of viruses that had no apparent counterparts in mammals. These attitudes began to shift with the discoveries of viruses, later proved to be retroviruses, that cause mammary carcinomas and leukemias in mice (*2*) and with the development of quantitative assays for chicken sarcoma and murine leukemia viruses in cultured cells (*3*). By the late 1960s and 1970s, the retrovirus community exhibited nearly logarithmic growth as major milestones were passed: the discovery of reverse transcriptase (*4*), the discovery of proviruses transmitted in the germ line (*5*), and the discovery of cellular progenitors of retroviral oncogenes (*6*). In the past few years, the growth of the retroviral community has been further accelerated by several connections between retroviruses and human diseases: the discoveries of human retroviruses that cause adult T cell leukemia/lymphoma (*7*) and AIDS (*8*), and the identification of human oncogenes, related to retroviral oncogenes, that are active in human cancers (*9*). These advances have galvanized widespread interest in retroviruses and their oncogenes in all branches of the medical community, among politicians and public interest groups, in biotechnology firms and their entrepreneurial supporters, and even in the public at large.

No longer a cottage industry, retrovirology has merged with several other disciplines as a consequence of some remarkable discoveries during this decade. (i) The structure of the provirus revealed that retroviruses belonged to a larger class of mobile genetic elements, called retrotransposons (or retroposons), important to investigators working on many eukaryotic organisms, especially yeast and *Drosophila* (*10–12*). (ii) Reverse transcription was assigned a central role in the replication of other viruses [hepatitis B (*13*) and cauliflower mosaic viruses (*14*)] and in the transposition and generation of other kinds of eukaryotic DNA (*15*). (iii) Once

recognized as naturally occurring vectors for host-derived oncogenes, retroviruses were studied as gene vectors by medical geneticists, developmental biologists, and practically anyone who wished to determine the phenotypic consequences of expressing a cloned gene in a cultured animal cell (*16*). (iv) With the attribution of biochemical functions to the products of retroviral oncogenes, investigation of viral tumorigenesis became closely linked to the study of growth factors, their receptors, signal transducers, protein kinases, and transcriptional regulators (*17*). Several of these disciplinary fusions will resurface in later discussions.

The Essential Facts About Retroviruses: The 3-Minute Course

Retroviruses resemble other animal viruses in several respects, but differ from all others in containing an RNA genome that replicates through a DNA intermediate (*4*, *10*–*12*). The extracellular virus particle is composed of a genome (single-stranded RNA) wrapped in a core of viral protein that is, in turn, surrounded by an envelope studded with viral glycoproteins and derived from the membrane of the previous host cell (Figs. 1 and 2A). Although multiplication occurs only within cells and depends on cellular functions, an infecting retrovirus also brings along an organized collection of viral enzymes and RNA designed to direct the synthesis of a double-stranded DNA copy of the RNA genome (reverse transcription) and the precise joining of that DNA to the host chromosome (integration).

The life cycle. Retroviruses attach to cells with the help of normal cell surface proteins specifically recognized by viral envelope proteins, and they probably enter through a mechanism, receptor-mediated endocytosis, that cells have evolved to ingest beneficial extracellular substances such as growth factors (*18*). Entry initiates the conversion of a quiescent, enveloped particle into an enzymatically active nucleoprotein complex that performs its idiosyncratic functions, reverse transcription and integration, with little or no help from the host cell (Fig. 2A). Once its provirus is ensconced in a cell chromosome, however, a retrovirus becomes highly dependent on its host—for replication of the provirus as part of its chromosomal context, for transcription of the provirus by RNA polymerase II, for processing of RNA transcripts by mechanisms normally used to cap, polyadenylate, and splice host RNAs, and for translation of resulting messenger RNAs by host polyribosomes (Fig. 2, A and B). These routine cellular functions are regulated by viral signals that determine the efficiency of transcription and splicing and mediate the occasional bypass of termination codons during translation, as discussed later. Finally, new virus particles are assembled and released. Still poorly characterized recognition signals draw together RNA and core proteins from the cytoplasm to associate with envelope glycoproteins embedded in plasma membranes, and a virus-encoded protease later cleaves viral polyproteins into the smaller components found in mature virus particles.

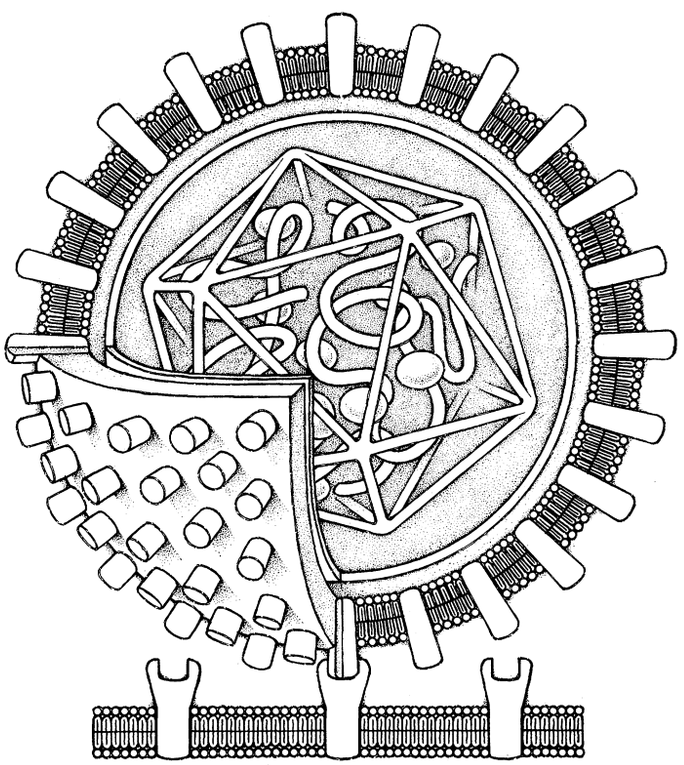

Fig. 1. Schematic view of a retrovirus particle. Two identical single strands of viral RNA and viral enzymes (reverse transcriptase, integrase, and protease) are drawn within an icosahedral viral core, and the core is surrounded by an envelope that is derived from host membranes enriched with viral glycoprotein. Interaction of envelope glycoprotein with a host-encoded cell surface receptor is shown at the bottom. [Adapted from (*93*); copyright 1987 *Advances in Oncology*]

The genome. The viral genome, as found in virus particles, is a complex of two identical chains of RNA, making retroviruses diploid, which is an oddity among viruses. Each viral RNA molecule is base-paired with a specific host transfer RNA that primes DNA synthesis, another curious feature of retroviruses. All retroviral genomes are organized in a standard format, best appreciated in comparison with a DNA version of the viral genome, the provirus integrated within host DNA (Fig. 2B). Sequences that regulate the structural transformations of the viral genome during the life cycle are clustered near the ends of the RNA: signals for initiation and progression of DNA synthesis, for integration, for transcription of the provirus into RNA, for RNA processing, and for packaging RNA into progeny particles. During the intricate maneuvers used to synthesize viral DNA, sequences present once near the ends of viral RNA (U3 and U5) are duplicated to generate long terminal repeats (LTRs), several hundred base pairs in length, at the ends of proviral DNA (Fig. 2B). The LTRs encompass many of the regulatory signals in the viral nucleotide sequence, and they are distinctive features that unite proviruses structurally with the broad collection of eukaryotic transposable elements known as retrotransposons (Fig. 3). Between these regulatory regions are coding sequences (open reading frames) for the major structural proteins of the virus particle (the *gag* frame encodes the core proteins, the *env* frame the envelope glycoproteins); for the enzymes found in particles (a protease, reverse transcriptase, and integrase, at least two of which are encoded by the *pol* frame); and for proteins with specialized, intracellular functions, exhibited only by those retroviruses endowed with oncogenes or regulatory genes.

Genetic behavior. During the virus life cycle, several interesting genetic and quasi-genetic phenomena may occur, especially if cells are infected by more than one virus: production of heterozygotic dimeric genomes, formation of pseudotypes at high frequencies (particles with core proteins and genome provided by one virus and envelope proteins by another), frequent deletions and nucleotide substitutions, and recombination between related, coinfecting viruses. [Recombination between retroviruses is surprisingly efficient, but its mechanistic basis has not been resolved (*19*).] Another, more specialized genetic attribute of retroviruses, their ability to cause insertion mutations, is fundamental to several of the important interactions between these agents and their hosts. The mutations

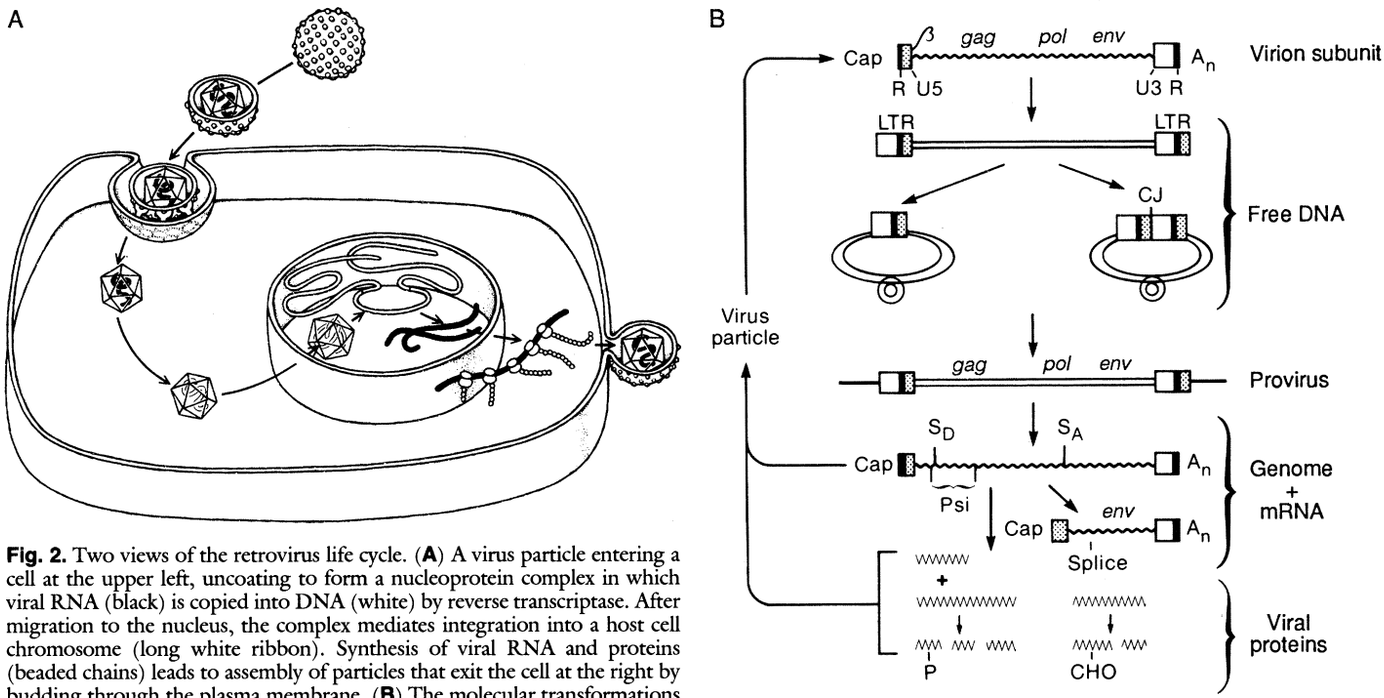

Fig. 2. Two views of the retrovirus life cycle. (**A**) A virus particle entering a cell at the upper left, uncoating to form a nucleoprotein complex in which viral RNA (black) is copied into DNA (white) by reverse transcriptase. After migration to the nucleus, the complex mediates integration into a host cell chromosome (long white ribbon). Synthesis of viral RNA and proteins (beaded chains) leads to assembly of particles that exit the cell at the right by budding through the plasma membrane. (**B**) The molecular transformations of the indicated species of the viral genome during the life cycle. Cap, capped nucleotide at 5′ end of viral RNA; A_n, polyadenylic acid at 3′ end of viral RNA; R, repeated sequence at ends of viral RNA; U3 and U5, unique sequences duplicated during DNA synthesis; LTR, long terminal repeat; CJ, circle junction, site of joining of ends of linear DNA; S_D and S_A, splice donor and acceptor sites, respectively; Psi, signal for packaging of viral RNA; P and CHO, modifications of viral proteins by phosphorylation and glycosylation, respectively. The figure is intentionally ambiguous about the immediate precursor to the provirus. [Adapted from (*11*, *12*, *31*)]

may cause either a recessive loss of function due to gene disruption or a dominant gain of function due to stimulation of expression of genes adjacent to an insertion site.

Biological behavior. Although united by their life cycle and the conserved features of their genomes, retroviruses are unusually diverse in their biological manifestations. Found in all vertebrates in which they have been sought, they can be transmitted horizontally in the form of infectious extracellular particles or genetically in the form of endogenous proviruses integrated within the germ line. The physiological consequences of infection range widely. In some instances, retroviruses produce no apparent effects, even when virus is produced in large amounts. In other cases, retroviruses potentiate growth of cultured cells or cause cancer in the intact host. A few retroviruses are deleterious to cells and produce a variety of destructive diseases, most notably immunodeficiency syndromes such as AIDS.

Retroviruses as Model Systems for Studying Eukaryotic Biology

Retroviruses as exemplars of information flow from RNA to DNA. The very name "retrovirus" embodies the biochemical property for which this class of viruses is most famous: the capacity to copy its RNA into DNA (reverse transcription). Work with these viruses first convinced the scientific community that transfer of information in biological systems was not limited to conventional transcription (DNA to RNA), translation (RNA to protein), and replication (DNA to DNA and RNA to RNA).

Although reverse transcription was first encountered in the retrovirus life cycle, it is hardly unique to retroviruses (*20*); it is now recognized as a widespread phenomenon in eukaryotic cells and viruses (*13–15*). Indeed, as much as 10% of the eukaryotic genome may be composed of products of reverse transcription (*21*). Moreover, reverse transcription seems to have been a logically necessary event at a crucial stage in evolution, the transition from the earlier RNA-dominated world to our DNA-dominated one (*22*).

Despite the many settings in which reverse transcription is now believed to occur, retroviral reverse transcriptases are still the only ones that can be studied in a satisfactory manner. This is due to the ease with which the enzymes can be solubilized and purified from retroviral particles (*23*), the availability of retroviral reverse transcriptases made in *Escherichia coli* by recombinant DNA technology (*24*), and the existence of several mutants in the viral *pol* gene (*10*). (The reverse transcriptases of hepatitis B and cauliflower mosaic viruses have not been solubilized from particles, those encoded by retrotransposons have not been harvested in sufficient amounts for serious biochemistry, and those responsible for synthesizing other components of vertebrate genomes have not been identified.)

Retroviral reverse transcriptases display many unexpected properties (*11*, *12*, *23*). They use RNA as natural primers, including the host transfer RNA base-paired near the 5′ end of the viral genome. They are "jumping polymerases" that transfer nascent strands between templates at least twice during synthesis of retroviral DNA. They have a second enzymatic activity, located in a different domain of the protein, that digests RNA to oligonucleotides once it has been copied into DNA [ribonuclease (RNase) H]. Finally, they will copy virtually any RNA template, once provided with a suitable DNA or RNA primer, making them popular reagents for nucleotide sequencing and for cloning complementary DNA copies of mRNAs. Still, major issues about these unusual and important enzymes are unresolved: virtually no structural work has been done, the active site for polymerization has not been defined, and few useful inhibitors have been identified.

An appreciation of the workings of retroviral reverse transcriptases has been central to the discoveries of reverse transcription in

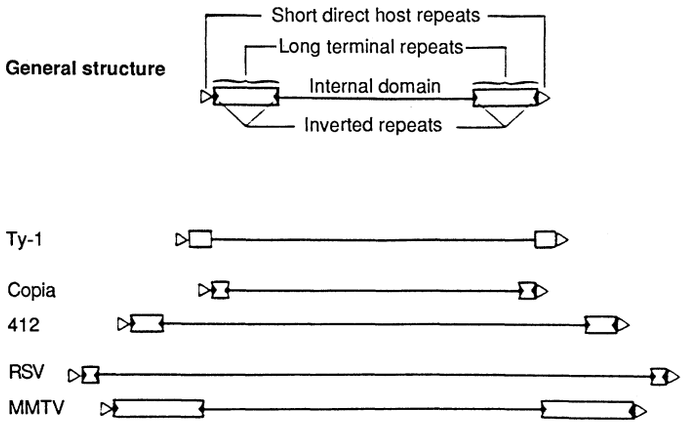

Fig. 3. Representative retrotransposons of yeast (Ty-1) and *Drosophila* (copia, 412) compared with proviruses of Rous sarcoma virus (RSV) and mouse mammary tumor virus (MMTV). Symbols are explained at top. [Adapted from (*11*); used with permission, copyright 1983 Academic Press]

the life cycles of other viruses and in the transposition cycle of eukaryotic transposons. Because a minus strand copy of virion RNA is always made before the plus strand, retroviral DNA synthesis appears asymmetrical, in contrast to traditional semiconservative DNA synthesis. A similar asymmetry of viral strands is apparent in livers infected by hepatitis B viruses: minus DNA strands are much more abundant than plus strands (*25*). On the basis of this provocative clue, stronger evidence for reverse transcription was sought and obtained (*13*). Retrovirus-based principles have also been helpful in the design of experiments showing that the yeast transposon Ty migrates by reverse transcription of an RNA copy of the element (*26*) and in the characterization of retrotransposition in *Drosophila* (*27*).

Retroviruses provide an efficient system for specialized eukaryotic recombination. Information about recombination in vertebrate organisms is scanty in general, although a specialized system for somatic recombination rearranges some genes in immune cells (*28*) and homologous recombination is known to occur at high frequency between extrachromosomal DNAs (*29*) and at low frequency with chromosomal DNA (*30*). Integration of viral DNA into host chromosomes by a specialized and precise recombinational mechanism provided by the virus is a cardinal feature of the retrovirus life cycle (*11*, *12*, *31*), and one that differentiates retroviruses from several other viruses (such as simian virus 40) whose DNAs are integrated infrequently and haphazardly (*32*). Indeed, integration of retroviral DNA is the recombination event in higher eukaryotes that occurs with greatest efficiency and precision, and it is the most amenable to biochemical and genetic analysis. Moreover, the analogous structures of endogenous proviruses and retrotransposons imply that the insertion of these elements uses the same mechanism.

A few facts about retroviral integration are now well established: (i) the structure of the product, proviral DNA, is invariant, with viral DNA always joined to host DNA 2 bp from the ends of the LTRs (*33*); (ii) specific sequences near the ends of the LTRs are necessary for the reaction (*34*) and can be considered analogs of bacteriophage attachment (*att*) sites; and (iii) the integration reaction requires a viral protein, the integrase, that is encoded near the 3′ end of the *pol* gene but is not involved in reverse transcription (*35*). Still, many important aspects of retroviral integration remain in dispute, including the nature of the immediate precursor to the provirus, the organization of viral ingredients, the preferential use of sites in host chromosomes, and the enzymological characteristics of the reactions.

These issues are now more easily approached because the integration reaction can be studied in a cell-free system, with murine leukemia virus (MLV) DNA as the integrating species and naked bacteriophage λ DNA as a target (*36*). The reaction is driven by a viral nucleoprotein machine derived from infected cells and composed of recently synthesized viral DNA associated with viral proteins from parental particles. This system and refinements of it may ultimately provide a picture of retroviral integration that rivals the view of integration of temperate bacteriophage DNAs into the *E. coli* chromosome.

Retroviruses as guides to the molecular basis of cancer. The cancer-causing properties of retroviruses probably provide the most common motivations for choosing to work with these agents. Oncogenic retroviruses, isolated from such vertebrates as fish, chickens, rodents, cats, subhuman primates, and humans, induce sarcomas (tumors of mesenchymal origin), various kinds of leukemias, and, less often, epithelial malignancies (the most common human cancers), including carcinomas of the breast, kidney, and liver (*10*). Given their small genomes and the regularity with which many induce characteristic forms of cancers in convenient laboratory animals, retroviruses represent seductively simple instruments with which to ask how something as complex as a normal animal cell can be converted into a cancer cell.

Almost all oncogenic retroviruses seem to fall into two camps [an important exception, human T cell leukemia virus (HTLV), is discussed below]. One group, typified by Rous sarcoma virus (RSV), carries a viral oncogene responsible for the swift induction of tumors in animals and the efficient transformation of cells in culture (*37*). The others, exemplified here by avian leukosis virus (ALV) and mouse mammary tumor virus (MMTV), lack a viral oncogene, do not transform cells in culture, but regularly cause tumors after long latency through what appears to be a multistep process (*38*).

Both groups of viruses cause tumors, at least in part, through the agency of cellular genes (Fig. 4) (*37–39*). Viruses like RSV have captured (transduced) host genes through a mechanism that probably begins with a proviral insertion mutation, ultimately placing the captured genes under the control of viral signals and mutating their coding sequences as well. Viruses without oncogenes stimulate expression of cellular genes through adjacent proviral insertions that override normal control elements and sometimes alter the structure of the gene products. Because both the transduced and insertionally activated genes contribute to cancerous change, they are called oncogenes, and their normal progenitors, proto-oncogenes. Evidence to date generally supports the premise that proto-oncogenes are important regulators of cell growth or development (*17*, *39*). Many of the proto-oncogenes discovered through the use of retroviruses are also sometimes targets for nonviral, somatic mutations believed to lead to human cancer (*9*).

Retroviral transduction and molecular cloning of provirally activated genes have together been responsible for isolation of the vast majority of proto-oncogenes, which now number about fifty (see Table 1 for examples). The profound influence of these genes upon the study of eukaryotic cells is apparent from the following historical synopses.

■ The cellular origin of v-*src*. Among retroviral oncogenes, the v-*src* gene of RSV was especially susceptible to genetic and biochemical maneuvers in the era before molecular cloning, because RSV is the only retrovirus with a transduced oncogene that can replicate without a helper virus. Temperature-sensitive and deletion mutants of v-*src* established that virus multiplication could be dissociated from neoplastic transformation and that transformation required continued expression of the viral oncogene (*40*, *41*). With molecular probes defined through the use of deletion mutants, v-*src* was shown to be closely related to, and presumably derived from, a normal and

highly conserved cellular gene, c-*src* (*6*), thereby establishing a paradigm for more than 20 other retroviral oncogenes (*37*, *39*).

■ A family of protein kinases. Subsequent discoveries that the protein products of v- and c-*src* are protein kinases (*37*, *42*), which specifically phosphorylate tyrosine residues (*43*), drew attention to the idea that tyrosine phosphorylation might be central to growth control, even though tyrosine phosphate constitutes less than 1% of the amino acid–derived phosphate in cellular proteins (*44*). The simple assays developed to detect kinase activities in immune complexes containing *src* proteins and the amino acid sequence motifs emblematic of the active domains of protein-tyrosine kinases have helped identify well over a dozen such kinases, including several that are transmembrane receptors for polypeptide growth factors (*44*). Some of these receptors have proven to be themselves the products of proto-oncogenes: the epidermal growth factor receptor gene is the progenitor of the avian virus oncogene, v-*erb*B (*45*), and the receptor for the macrophage growth factor, CSF-1, is encoded by the progenitor of the feline virus oncogene, v-*fms* (*46*). The ligand for yet another receptor with protein-tyrosine kinase activity, the platelet-derived growth factor (PDGF) receptor, is partially encoded by the *sis* proto-oncogene (*47*).

■ An oncogene explains a mutant chromosome. An odd version of chromosome 22, called the Philadelphia chromosome (Ph^1), was one of the first visible signs of a reproducible genetic lesion to be noted in human leukemia cells (*48*). More than 20 years later, c-*abl*, the cellular homolog of the oncogene of Abelson-MLV, normally found on chromosome 9 where it encodes a relatively inactive protein-tyrosine kinase, was shown to be broken and fused to another gene on chromosome 22, forming both Ph^1 and a hybrid protein with augmented kinase activity (*49*).

■ *ras* oncogenes in human tumors. Nearly a decade ago, human tumors and cells transformed by chemical mutagens were found to contain active oncogenes through experiments in which DNA from such cells and tumors was used to induce oncogenic properties in a tissue culture cell line (*50*). Subsequently the genes were identified in most cases as mutant versions of *ras* genes (*51*), proto-oncogenes first discovered as the progenitors of the *ras* oncogenes of murine sarcoma viruses (*37*). Like *src* proteins, *ras* proteins have a measurable biochemical function, guanosine triphosphate (GTP) binding and hydrolysis, and they also belong to a large family of proteins that includes GTPases (G proteins), which convey extracellular signals to adenylate cyclases (*52*).

■ An oncogene implicated in transcriptional control. The *jun* oncogene, recently discovered in an avian sarcoma virus, is related to the yeast *GCN4* gene, a regulator of transcription, on the basis of both structural and functional tests (*53*). Because *GCN4* protein was known to bind the same DNA sequence as does a mammalian transcriptional activation complex, AP-1, the cellular homolog of v-*jun* seemed a good candidate to encode a component of AP-1; recent evidence argues strongly that it does (*54*).

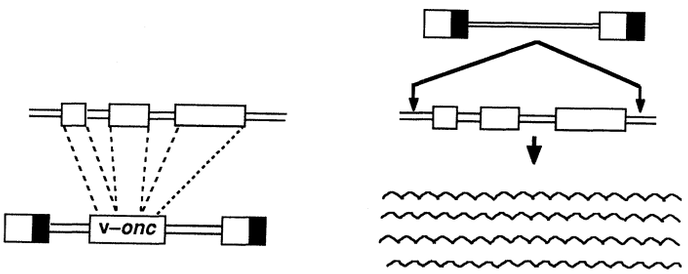

Fig. 4. Simplified representations of the two common mechanisms by which retroviruses harness cellular proto-oncogenes to cause cancers. The figure on the left indicates that a typical viral oncogene (v-*onc*) is derived from exons of a cellular gene by transduction; the figure on the right indicates that proviral insertion mutations on either side of the exons of a cellular gene may cause augmented expression of the gene.

■ Genetic rearrangements of c-*myc*. ALV, a cause of B cell lymphomas in chickens, was the first retrovirus shown to cause insertion mutations during tumor induction (*55*). The significance of the mutations was established by the identity of the activated gene: c-*myc* (*56*), a gene already labeled a proto-oncogene because several retroviruses carry it in transduced form (*37*). Thus retroviruses are central to both arms of the argument that first implicated a specific cellular gene in an oncogenic process: the progenitor of a retroviral oncogene is also the target for mutation by a provirus without an oncogene. The c-*myc* insertion mutations proved to be paradigms for other kinds of genetic rearrangements—chromosomal translocations and gene amplification—that affect c-*myc* and its close relatives, *N-myc* and *L-myc*, with particular frequency (*9*).

■ Development and *int*-1. The first gene found activated by MMTV proviral insertions in mammary cancers (*57*), *int*-1, is not related to any oncogenes transduced by retroviruses. The most striking features of this proto-oncogene suggest a role in development: in mice, the gene is normally expressed in only two places, the neural tube in midgestational embryos and postmeiotic cells in testes (*58*); in *Drosophila*, the homolog of *int*-1 has been identified as *wingless*, a gene required for a normal gradient of cells in larval body

Table 1. Categories of proto-oncogenes and modes of retroviral activation. The table provides a few examples of proto-oncogenes (most discussed in the text) whose protein products have been located within or outside the cell and assigned physiological and biochemical functions. The two columns at the right indicate whether the genes have been naturally transduced by retroviruses and whether they have been encountered as targets for proviral insertion mutations. Further details and references can be found in the text and in (*9*), (*17*), and (*36*). PDGF, platelet-derived growth factor; EGF, epidermal growth factor; CSF-1, colony stimulating factor–1.

Location of protein product	Proposed function	Example	Retroviral transduction	Insertional activation
Secretory vesicles and extracellular space	Growth factor, ligand for membrane receptor	*sis*/PDGF	+	−
		int-1	−	+
Plasma membrane				
Transmembrane	Growth factor receptor with protein-tyrosine kinase activity	*erb*B/EGF receptor	+	+
		fms/CSF-1 receptor	+	+
Inner face	Protein-tyrosine kinase	*src*	+	−
		abl	+	−
	Signal transduction, GTP binding and hydrolysis	*ras*	+	+
Cytoplasm	Protein-serine (threonine) kinase	*mos*	+	+
Nucleus	DNA binding, transcriptional regulation	*jun*	+	−
		fos	+	−
		myc	+	+

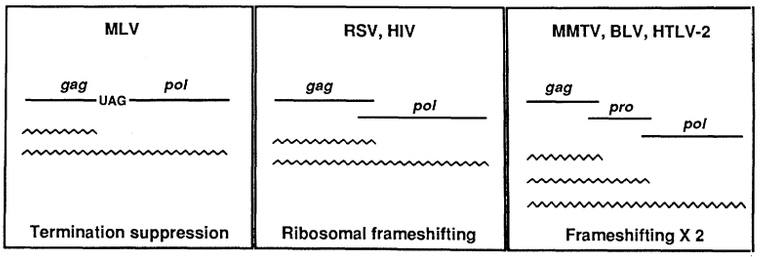

Fig. 5. Strategies for expression of retroviral *pol* genes. Arrangements of the *gag*, *pol*, and *pro* reading frames of some representative retroviral genomes (straight lines) are shown in relation to their protein products (wavy lines). Underlying reading frames are −1 with respect to the frames above. Suppression and frameshifting events occur with frequencies ranging from 5 to 25%, so that *gag-pol or gag-pro-pol* polyproteins are about 5% as abundant as *gag* proteins. MLV, murine leukemia virus; RSV, Rous sarcoma virus; HIV, human immunodeficiency virus, MMTV, mouse mammary tumor virus; BLV, bovine leukemia virus; HTLV, human T cell leukemia virus.

segments (*59*). The *int*-1 protein is likely to act as an intercellular signal for both normal development and neoplastic growth (*60*).

These brief tales are meant to suggest the multiplicity of ways in which genes brought to light by retroviruses are shaping the study of oncogenesis, growth control, and development. These genes provide an unfulfilled opportunity to describe in molecular terms how cells behave when responding to signals for normal growth or when violating restraints during neoplastic growth. To do this, it will be necessary to learn more about the biochemical properties of the proteins involved and the identities of their relevant biochemical targets.

Could the genes we now call proto-oncogenes have been found without the assistance of retroviruses? It is possible that growth factor receptors would ultimately have led to genes like *src*, that transcription factors would have led to *jun*, and that *Drosophila* genes implicated in development would have led to genes like *int*-1. But progress by such routes, and perception of the relationship to cancer, would have been painfully slow.

Retroviruses as probes for host regulatory mechanisms. The utility of retroviruses for understanding properties of their host cells is particularly obvious during viral gene expression, when the provirus is dependent on host machinery for transcription, RNA processing, translation, and protein modification (Fig. 2). The simplicity of their genomes and their harmonious existence within their hosts made retroviruses especially attractive reagents for working on such topics in the years before molecular cloning. Although any cellular gene can now, in principle, be isolated for such purposes, retroviruses continue to be widely used, in part because of historic precedent, and in part because a retroviral provirus presents a remarkable opportunity to examine adaptation of a parasite to its host. Thus the host provides the machinery, and the provirus encodes the signals that regulate expression, sometimes in surprising ways. This formula has inspired study of many aspects of retroviral gene expression; a few examples illustrate its attractions.

■ Transcription. Signals that modulate retroviral transcription are lodged mainly in the U3 region of the LTR and were among the first eukaryotic promoters and enhancers to be carefully studied (*11*, *12*). The regulatory elements are recognized by host transcription factors (*61*), determine how well a virus will grow in different cell types (*11*, *12*), and influence the oncogenic spectrum of each retrovirus (*38*, *62*). Sequences in the MMTV LTR are binding sites for the glucocorticoid receptor, and the MMTV promoter is stimulated by the hormone-receptor complex (*63*). The MMTV LTR was the first transcriptional initiator shown to be subject to primary regulation by glucocorticoid hormones (*64*), it is still commonly used to examine the mechanism of regulation, and it is a popular promoter for achieving inducible expression of heterologous genes in eukaryotic cells (*65*).

■ Splicing. Even retroviruses that make only a single spliced subgenomic RNA must regulate the ratio of spliced RNA and its precursor, since the precursor is also used as both genomic RNA and mRNA (see Fig. 2B). The signals for maintaining appropriate levels are not known, but they may be differently interpreted in different hosts. Thus in avian cells less than half of RSV RNA is spliced to make two subgenomic RNAs, but in mammalian cells nearly all is spliced to form one subgenomic species (*66*). In addition, some retroviruses [human immunodeficiency virus (HIV), HTLV, and their close relatives] have more complex splicing patterns and encode viral proteins that influence the ratio of spliced to unspliced RNAs (*67*, *68*).

■ Translation. The mechanisms used for synthesis of retroviral *pol* proteins illustrate the capacity of the eukaryotic translational apparatus to do unexpected things in response to retroviral signals. All retroviruses express their *pol* genes as *gag-pol* fusions, but *gag* and *pol* are separated by a stop codon in MLV; read in different, briefly overlapping frames in RSV and HIV; and separated by a third frame (*pro*, for the viral protease) in MMTV and HTLV (Fig. 5). Rather than commandeer the host's splicing apparatus to create mRNAs with a single open reading frame for *gag-pol* (or *gag-pro-pol*) proteins, retroviruses have instead exploited previously unrecognized potentials of vertebrate ribosomes to insert an amino acid occasionally in response to the nonsense codon at the end of MLV *gag* (*69*) or to shift reading frames at defined sites and frequencies when translating the other viral RNAs (*70*). Neither of these phenomena have been encountered during the translation of cellular mRNA and would, of course, be deleterious if allowed to occur frequently and without purpose. But they have clear benefits for retroviruses: structural (*gag*) proteins can be made in large amounts and catalytic (*pro* and *pol*) proteins in relatively small amounts; and *pol* products can be incorporated into viral cores through attached *gag* components.

What is the basis of such translational control? All of the retroviruses frameshifts move the ribosome into the −1 reading frame in response to at least two sets of instructions in the viral RNA: a short sequence at the frameshift site, and secondary structure downstream of the site (*71*). Although no cellular genes have yet been found to use the frameshifting potential of eukaryotic ribosomes, at least one other class of viruses, the coronaviruses, take advantage of it (*72*), and at least one retrotransposon, the Ty element of yeast, mediates frameshifting in the +1 direction (*73*).

Retroviruses as structural models. Retroviruses are excellent models for thinking about complex interactions among macromolecules in eukaryotic cells. The retrovirus life cycle, like that of several other animal viruses, is rich with such interactions. At the outset, a retroviral glycoprotein must recognize a cell surface protein, bind to it, and mediate uptake of the virus particle into the cell. How do these events occur? Do they differ from other internalization processes? What are the structural transformations of the particle that accompany virus entry and activate reverse transcription of viral RNA? What is the organization of the nucleoprotein complex that makes viral DNA and mediates its integration? Later in the life cycle, viral RNA, transfer RNA, and products of *gag*, *pol*, and *env* must correctly assemble into particles. What are the rules that govern virus assembly? In particular, how do *gag* proteins interact to form cores? How does RNA get into cores? How is the protease activated to process polyproteins into mature components? How does the

core associate with a region of plasma membrane enriched with envelope glycoproteins? In this area, there are few accomplishments to recount: it is a true frontier.

Retroviruses as Pathogens in Humans and Animals

To this point, we have considered several ways in which retroviruses have been informative about general problems in eukaryotic biology. But retroviruses have an unusual property as a biological system: they are problems themselves, because they are causative agents of disease, including lethal diseases of humans.

■ HIV and AIDS. Discovery of a retrovirus as the cause of AIDS (*8*) has had many effects on this discipline. Precepts about retroviruses are now matters of public health, funding for retroviral research in both public and private sectors is growing rapidly, and "retrovirus" is in the lexicon of common speech. But there have also been important effects on the retrovirologist's view of his or her own science generated by the discovery of HIV: a renaissance of interest in the retrovirus life cycle, the pathogenesis of cytotoxic infections, and the immune response to retroviruses.

Several factors have contributed to the revitalized study of the life cycle. Most obvious is the incentive to identify steps in replication at which intervention might be successful, despite the generally dismal experience with antiviral drugs. (i) Identification of the cell surface protein, CD4, as the receptor (or a major component of the receptor) for HIV (*74*) permits a deliberate assault on the initial step in replication and offers a potent ligand for virus particles (*75*). (ii) Although the only drug to show clear clinical benefit thus far, azidothymidine, attacks the "obvious" step, reverse transcription (*76*), recent perceptions about retroviral integration, transcription, translation, proteolysis, and assembly encourage a search for new drugs to act at other steps as well. (iii) The HIV genome has been discovered to encode no less than five novel proteins in addition to those encoded by *gag*, *pol*, and *env*: one protein is a positive regulator of HIV gene expression, acting mainly to increase levels of viral RNA (*77*); another influences the relative abundance of various spliced and unspliced HIV mRNAs (*67*); a third augments the infectiousness of HIV (*78*); and yet another inhibits virus growth and resembles host G proteins (*79*). The mechanisms by which such proteins regulate virus production are largely unknown, but of obvious importance.

Because most retroviruses are not cytopathic and because there has been little incentive to develop vaccines against retroviral diseases in most animals, studies of retroviral pathogenesis (other than oncogenesis) and of the immune response to retroviral infection have lagged behind work on the molecular biology of these viruses. AIDS has dramatically changed these attitudes. The immune response must be understood in order to develop clinical tests for HIV and for the consequences of infection; to devise better strategies for vaccination in the face of discouraging levels of neutralizing antibodies in infected patients (*80*) and disappointing vaccine trials in chimpanzees (*81*); and to evaluate hypotheses about pathogenesis, taking into account both antiviral and autoimmune components and the complications of infecting cells that are themselves central to an immune response (*82*).

■ HTLV. The first isolates of HTLVs from patients with an uncommon but virulent leukemia (*7*) vindicated a decade of frustrating and sometimes embarrassing efforts to find human oncogenic retroviruses to match their counterparts in animals (*83*). The HTLVs have proven to be curious agents, difficult to understand as pathogens and difficult to study as infectious viruses. Unlike the common oncogenic retroviruses of animals, the HTLVs neither carry host-derived oncogenes nor activate cellular proto-oncogenes by insertion mutation. Instead, their oncogenic action has been provisionally ascribed to an open reading frame that lies between the *env* gene and the 3'LTR and encodes a protein that acts as a positive effector of transcription from the HTLV LTR and from certain cellular promoters (*84*). However, models for tumorigenesis must also account for the prolonged latency and infrequent occurrence of disease in infected people, the lack of viral gene expression in primary tumor tissue, and the dearth of direct transformation assays for individual viral genes (*85*).

Although the number of cases of HTLV-associated leukemia and lymphoma is relatively small, many people are infected, especially in Japan and the Caribbean and among intravenous drug abusers in this country and Europe (*86*). Thus better understanding of the biological properties of the HTLVs and their relatives (bovine leukemia virus and simian T cell leukemia virus) is urgently needed.

Retroviruses as Tools for Studying Development, Delivering Genes, and Curing Diseases

Retroviruses can also be used as technical devices for genetically altering host cells and organisms. When put to such purposes, retroviruses themselves are not usually at the heart of the scientific question; instead they need to be understood only as far as necessary to achieve other objectives: marking cells with a recognizable provirus, causing mutations with interesting phenotypes, expressing a favorite gene in a desired cell type, or correcting a genetic deficiency.

The incentive to use retroviruses as genetic vectors originated with the perception that retroviruses with viral oncogenes are naturally occurring genetic vectors (Fig. 4). The ways in which oncogenes are incorporated into viral genomes and efficiently expressed have now been expropriated by investigators designing vectors of their own, to deliver to chosen cells any of the large collection of genes made available by molecular cloning. In general, this enterprise has been hugely successful: experiments in many fields of biology now depend on retrovirus vectors to deliver genes to cultured cells and occasionally to animals. The vectors can be grown to high titers, they often carry two genes in various arrangements, and they are available in models that do or do not initiate an infection that spreads to surrounding cells (*16*).

It is anticipated that retroviral vectors will ultimately be used to correct human genetic deficiencies (*16*). However, reliable expression of transmitted genes has yet to be achieved after infection of hematopoietic stem cells, the usual targets in current strategies, and it is not yet known whether infected cells will persist in the host in numbers adequate to ameliorate symptoms. The safety of retrovirus vectors has also not been fully evaluated. Improved design of vectors intended for human gene therapy thus remains a major technical challenge, but one that promises relief from any disease caused by a recessive mutation in a gene available for delivery.

In the meantime, retroviruses are becoming important tools in developmental biology because proviruses can stably and benignly mark cells for tracing lineages and can initiate insertion mutations with developmental consequences. (i) Cells have been marked with MLV proviruses by infecting preimplantation embryos to learn when cells become committed to a single lineage or organ (*87*); by infecting hematopoietic stem cells to show that a single cell can serve as the source of all blood cells (*88*); and by infecting retinas to identify the heterogeneous descendants of a single cell (*89*). (ii) Natural or experimental infection of the mouse germ line with MLV has occasionally produced insertion mutations with especially interesting effects: an endogenous provirus on chromosome 9 is respon-

sible for the light hair pigmentation characteristic of *dilute* mice (*90*), and an insertion into an intron of a collagen gene after infection of the mouse germ line produced a recessive mutation lethal to midgestational embryos (*91*). (iii) Retroviral infection of embryonic stem cells in culture may ultimately permit production of mouse strains with a wide variety of important genetic lesions. Some success has already been achieved with a known X-linked gene, *hprt*, as a mutational target for MLV; metabolic selection of cells in which *hprt* was disrupted recently led to production of HPRT-deficient mice from the mutant stem cells (*92*).

A Final Perspective on Retrovirology

This review has stressed the several facets of retroviruses that have attracted people to them in increasing numbers over the past two decades: retroviruses as potent models for understanding many aspects of both normal and cancerous eukaryotic cells; as important pathogens in humans and animals; as technical devices for gene delivery, mutagenesis, and lineage marking; and, most simply, as inherently fascinating microbes. Although the success of retrovirology is apparent from the sheer numbers of people who now work with these viruses, a more persuasive measure is the major impact the discipline has had upon the way we now think about many important topics in contemporary biology: the pathways for transfer of genetic information in eukaryotic cells, the causes of cancer and other major diseases, the genes and proteins that contribute to normal growth and development, and the regulation of gene expression.

Despite the extraordinary productivity of retrovirology in recent years, questions in all branches of the discipline seem larger and more numerous than ever. The major outlines of the replicative cycle are firmly drawn, but mechanisms of central events—virus entry, integration, regulated expression, and assembly—are just now coming into view and have assumed a greater urgency because of AIDS. A monumental list of oncogenes and proto-oncogenes has been assembled, but the biochemical activities crucial to growth control and neoplasia await discovery. Important human pathogens have been identified among retroviruses, but strategies for prevention and cure are still desperately needed. Strong evidence for the utility of retroviruses as genetic reagents is in hand, but insertional mutagenesis is not yet a simple experimental device, and gene therapy with retroviral vectors is not yet ready for clinical trials. Such deficiences are the legacy of progress, and an invitation to the future.

REFERENCES AND NOTES

1. P. Rous, *J. Exp. Med.* **13**, 397 (1911); V. Ellermann and O. Bang, *Fizentralbl. Bakteriol.* **46**, 595 (1980).
2. J. J. Bittner, *Science* **84**, 162 (1936); L. Gross, *Proc. Soc. Exp. Biol. Med.* **76**, 27 (1951).
3. H. M. Temin and H. Rubin, *Virology* **6**, 669 (1958).
4. H. M. Temin and S. Mizutani, *Nature* **226**, 1211 (1970); D. Baltimore, *ibid.*, p. 1209.
5. R. J. Huebner and G. J. Todaro, *Proc. Natl. Acad. Sci. U.S.A.* **64**, 1087 (1969); D. R. Lowy, W. P. Rowe, N. Teich, J. W. Hartley, *Science* **174**, 155 (1971); W. P. Rowe, *J. Exp. Med.* **136**, 1272 (1972).
6. D. Stehelin, H. E. Varmus, J. M. Bishop. P. K. Vogt, *Nature* **260**, 170 (1976).
7. B. J. Poiesz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7415 (1980).
8. F. Barré-Sinoussi *et al.*, *Science* **220**, 868 (1983); R. C. Gallo *et al.*, *ibid.* **224**, 500 (1984); J. A. Levy *et al.*, *ibid.* **225**, 840 (1984).
9. J. M. Bishop, *ibid.* **235**, 305 (1987); H. E. Varmus, *Annu. Rev. Genet.* **18**, 553 (1984).
10. R. A. Weiss, N. Teich, H. Varmus, J. Coffin, Eds., *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982, 1985), vol. 1 and vol. 2, respectively.
11. H. E. Varmus, in *Mobile Genetic Elements*, J. Shapiro, Ed. (Academic Press, New York, 1983), pp. 411–503.
12. ——— and R. Swanstrom, in (*10*), vol. 1, pp. 369–512.
13. J. Summers and W. S. Mason, *Cell* **29**, 403 (1982).
14. R. Hull and S. N. Covey, *Trends Biol. Sci.* **8**, 119 (1982).
15. A. M. Weiner, P. L. Deininger, A. Efstratiadis, *Annu. Rev. Biochem.* **55**, 631 (1987).
16. E. Nichols, *Human Gene Therapy* (Harvard Univ. Press, Cambridge, MA, 1988); J. Coffin, in (*10*), vol. 2, pp. 17–73.
17. H. E. Varmus and J. M. Bishop, Eds., *Cancer Surv.* **5** (1986); H. E. Varmus, in *Molecular Basis of Blood Diseases*, G. Stamatoyannopoulos, A. W. Nienhuis, P. Leder, P. W. Majerus, Eds. (Saunders, Philadelphia, 1987), pp. 271–346.
18. C. A. Mims, *J. Infect. Dis.* **12**, 199 (1986).
19. J. M. Coffin, *J. Gen. Virol.* **42**, 1 (1979); E. Hunter, *Curr. Top. Microbiol. Immunol.* **79**, 295 (1978).
20. H. E. Varmus, *Sci. Am.* **257**, 56 (September 1987).
21. H. M. Temin, *Mol. Biol. Evol.* **2**, 455 (1986).
22. J. D. Watson, V. Hopkins, T. Roberts, J. Steitz, A. Weiner, *Molecular Biology of the Gene* (Benjamin-Cummings, Menlo Park, CA, ed. 4, 1987), vol. 2, chap. 28, pp. 1098–1163.
23. I. M. Verma, *Biochim. Biophys. Acta* **473**, 1 (1977).
24. M. J. Roth, N. Tanese, S. Goff, *J. Biol. Chem.* **260**, 9326 (1985); N. Tanese, J. Sodroski, W. Haseltine, S. Goff, *J. Virol.* **59**, 743 (1986); A. Hizi, C. McGill, S. H. Hughes, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 1218 (1988).
25. W. S. Mason, C. Aldrich, J. Summers, J. M. Taylor, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3997 (1982).
26. D. J. Garfinkel, J. D. Boeke, G. R. Fink, *Cell* **42**, 507 (1985); J. D. Boeke, D. J. Garfinkel, C. A. Styles, G. R. Fink, *ibid.* **40**, 491 (1985).
27. A. J. Flavell and D. Ish-Horowicz, *ibid.* **34**, 415 (1983); I. R. Arkhipova *et al.*, *ibid.* **44**, 555 (1986).
28. S. Tonegawa, *Sci. Am.* **253**, 122 (October 1985); L. Hood, M. Kronenberg, T. Hunkapiller, *Cell* **40**, 225 (1985); S. Akira, K. Okazaki, H. Sakano, *Science* **238**, 1134 (1987).
29. R. S. Kucherlapati, E. M. Eves, K.-Y. Song, B. S. Morse, O. Smithies, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3153 (1984); C. T. Wake, F. Vernaleone, J. H. Wilson, *Mol. Cell. Biol.* **5**, 2080 (1985); P. K. Bandyopadhyay, S. Watanabe, H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3476 (1984).
30. K. R. Thomas, K. R. Folger, M. R. Capecchi, *Cell* **44**, 419 (1986); F. L. Lin, K. Sperle, N. Sternberg, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1391 (1985); R. M. Liskay and J. L. Stachelek, *Cell* **35**, 157 (1983).
31. H. E. Varmus and P. Brown, in *Mobile DNA Elements*, M. Howe and D. Berg, Eds. (American Society of Microbiology, Washington, DC, in press).
32. M. Botchan, W. Topp, J. Sambrook, *Cell* **9**, 269 (1976); G. Ketner and T. J. Kelly, Jr., *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1102 (1976).
33. S. H. Hughes *et al.*, *Cell* **15**, 1397 (1978); T. W. Hsu, J. L. Sabran, G. E. Mark, R. V. Guntaka, J. M. Taylor, *J. Virol.* **28**, 810 (1979).
34. A. T. Panganiban and H. M. Temin, *Nature* **306**, 155 (1983); J. Colicelli and S. P. Goff, *J. Mol. Biol.* **199**, 47 (1988).
35. L. A. Donehower and H. E. Varmus, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6461 (1984); P. Schwartzberg, J. Colicelli, S. P. Goff, *Cell* **37**, 1043 (1984); A. T. Panganiban and H. M. Temin, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7885 (1984).
36. P. O. Brown, B. Bowerman, H. E. Varmus, J. M. Bishop, *Cell* **49**, 347 (1987).
37. J. M. Bishop and H. E. Varmus, in (*10*), vol. 1, pp. 999–1108.
38. N. Teich, J. Wyke, T. Mak, A Bernstein, W. Hardy, *ibid.*, pp. 785–998.
39. J. M. Bishop, *Annu. Rev. Biochem.* **52**, 301 (1983).
40. G. S. Martin, *Nature* **227**, 1021 (1970); P. K. Vogt, *Virology* **46**, 939 (1971).
41. M. Linial, in (*10*), vol. 1, pp. 649–783.
42. M. S. Collett and R. L. Erikson, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2021 (1978); A. D. Levinson, H. Oppermann, L. Levintow, H. E. Varmus, J. M. Bishop, *Cell* **15**, 561 (1978).
43. T. Hunter and B. M. Sefton, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1311 (1980).
44. T. Hunter and J. A. Cooper, *Annu. Rev. Biochem.* **54**, 897 (1985).
45. J. Downward *et al.*, *Nature* **307**, 521 (1984).
46. C. J. Sherr *et al.*, *Cell* **41**, 665 (1985).
47. R. F. Doolittle *et al.*, *Science* **221**, 275 (1983); M. D. Waterfield *et al.*, *Nature* **304** 35 (1983).
48. P. C. Nowell and D. A. Hungerford, *Science* **132**, 1497 (1960).
49. J. Groffen *et al.*, *Cell* **36**, 93 (1984); J. B. Konopka, S. M. Watanabe, O. N. Witte, *ibid.* **37**, 1035 (1984).
50. C. Shih, B.-Z. Shilo, M. P. Goldfarb, A. Dannenburg, R. A. Weinberg, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5714 (1979).
51. C. J. Tabin *et al.*, *Nature* **300**, 143 (1982); C. J. Der, T. G. Krontiris, G. M. Cooper, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3637 (1982).
52. M. Barbacid, *Annu. Rev. Biochem.* **56**, 779 (1987).
53. P. K. Vogt, T. J. Bos, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3316 (1987); K. Struhl, *Cell* **50**, 841 (1987).
54. D. Bohmann *et al.*, *Science* **238**, 1386 (1987).
55. B. G. Neel, W. S. Hayward, H. L. Robinson, J. Fang, S. M. Astrin, *Cell* **23**, 323 (1981); G. S. Payne *et al.*, *ibid.*, p. 311.
56. W. S. Hayward, B. G. Neel, S. M. Astrin, *Nature* **290**, 475 (1981).
57. R. Nusse and H. E. Varmus, *Cell* **31**, 99 (1982).
58. A. Jakobovits, G. M. Shackleford, H. E. Varmus, G. R. Martin, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7806 (1986); G. M. Shackleford and H. E. Varmus, *Cell* **50**, 89 (1987); D. G. Wilkinson, J. A. Bailes, A. P. McMahon, *ibid.*, p. 79.
59. F. Rijsewijk *et al.*, *Cell* **50**, 649 (1987); C. V. Cabrera, M. C. Alonso, P. Johnston, R. G. Phillips, P. A. Lawrence, *ibid.*, p. 659; N. Baker, *EMBO J.* **6**, 1765 (1987).
60. J. Papkoff, A. M. C. Brown, H. E. Varmus, *Mol. Cell. Biol.* **7**, 3978 (1987); G. Morata and P. Lawrence, *Dev. Biol.* **56**, 227 (1977).
61. G. Nabel and D. Baltimore, *Nature* **326**, 711 (1987); N. A. Speck and D. Baltimore, *Mol. Cell. Biol.* **7**, 1101 (1987); K. A. Jones, J. T. Kadonaga, P. A. Luciw, R. Tjian, *Science* **232**, 755 (1986).

62. P. A. Chatis, C. A. Holland, J. W. Hartley, W. P. Rowe, N. Hopkins, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 4408 (1983); D. Celander and W. A. Haseltine, *Nature* **312**, 159 (1984).
63. K. Yamamoto, *Annu. Rev. Genet.* **199**, 209 (1985).
64. G. Ringold, K. R. Yamamoto, G. M. Tomkins, J. M. Bishop, H. E. Varmus, *Cell* **6**, 299 (1975).
65. E. B. Jakobovits, J. E. Majors, H. E. Varmus, *ibid.* **38**, 757 (1984).
66. W. S. Hayward, *J. Virol.* **24**, 47 (1977); S. R. Weiss, H. E. Varmus, J. M. Bishop, *Cell* **12**, 983 (1977); N. Quintrell, S. H. Hughes, H. E. Varmus, J. M. Bishop, *J. Mol. Biol.* **143**, 363 (1980).
67. M. B. Feinberg, R. F. Jarrett, A. Aldovini, R. C. Gallo, F. Wong-Staal, *Cell* **46**, 807 (1986); J. Sodroski *et al.*, *Nature* **321**, 412 (1986).
68. J. Inoue, M. Yoshida, M. Seiki, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 3653 (1987); M. Hidaki, J. Inoue, M. Yoshida, M. Seiki, *EMBO J.* **7**, 519 (1988).
69. Y. Yoshinaka, I. Katoh, T. D. Copeland, S. Oroszlan, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1618 (1985).
70. T. Jacks and H. E. Varmus, *Science* **230**, 1237 (1985); T. Jacks, K. Townsley, H. E. Varmus, J. Majors, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4298 (1987); R. Moore, M. Dixon, R. Smith, G. Peters, C. Dickson, *J. Virol.* **61**, 480 (1987).
71. T. Jacks, F. R. Masiarz, H. D. Madhani, H. E. Varmus, unpublished data.
72. I. Brierley *et al.*, *EMBO J.* **6**, 3779 (1987).
73. J. Clare and P. Farabaugh, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2829 (1985); W. Wilson, M. H. Malim, J. Mellor, A. J. Kingsman, S. M. Kingsman, *Nucleic Acids Res.* **14**, 7001 (1986).
74. A. G. Dalgleish *et al.*, *Nature* **312**, 763 (1984); D. Klatzmann *et al.*, *ibid.* p. 767; P. J. Maddon *et al.*, *Cell* **47**, 333 (1986); J. S. McDougal *et al.*, *Science* **231**, 382 (1986).
75. Q. J. Sattentau and R. A. Weiss, *Cell* **52**, 631 (1988).
76. H. Mitsuya and S. Broder, *Nature* **325** 773 (1987).
77. C. A. Rosen *et al.*, *ibid.* **319**, 555 (1986); B. R. Cullen, *Cell* **46**, 973 (1986); A. G. Fisher *et al.*, *Nature* **320**, 367 (1986).
78. A. G. Fisher *et al.*, *Science* **237**, 888 (1987); K. Strebel *et al.*, *Nature* **328**, 728 (1987).
79. P. A. Luciw, H. Oppermann, J. M. Bishop, H. E. Varmus, *Mol. Cell. Biol.* **4**, 1260 (1984); B. Guy *et al.*, *Nature* **330**, 266 (1987).
80. R. A. Weiss *et al.*, *Nature* **316**, 69 (1985).
81. P. W. Berman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
82. A. S. Fauci, *Science* **239**, 617 (1988).
83. R. Weiss, in (*10*), vol. 1, pp. 1205–1281.
84. J. G. Sodroski, C. A. Rosen, W. A. Haseltine, *Science* **225**, 381 (1984); M. Seiki, J. I. Inoue, T. Takeda, M. Yoshida, *EMBO J.* **5**, 561 (1986); W. C. Greene *et al. Science* **232**, 877 (1986).
85. I. S. Y. Chen, W. Wachsman, J. D. Rosenblatt, A. J. Cann, *Cancer Surv.* **5**, 329 (1986).
86. Y. Hinuma *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6476 (1981); R. S. Tedder *et al.*, *Lancet* **ii**, 125 (1984); R. C. Gallo, *Sci. Am.* **255**, 88 (December 1986).
87. P. Soriano and R. Jaenisch, *Cell* **46**, 19 (1986).
88. J. E. Dick, M. C. Magli, D. Huszar, R. A. Phillips, A. Bernstein, *ibid.* **42**, 71 (1985); I. R. Lemischka, D. H. Raulet, R. C. Mulligan, *ibid.* **45**, 917 (1986).
89. J. Price, D. Turner, C. Cepko, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 156 (1987).
90. N. A. Jenkins, N. G. Copeland, B. A. Taylor, B. K. Lee, *Nature* **293**, 370 (1981); N. G. Copeland, K. W. Hutchison, N. A. Jenkins, *Cell* **33**, 379 (1983).
91. R. Jaenisch *et al.*, *Cell* **32**, 209 (1983).
92. M. R. Kuehn, A. Bradley, E. J. Robertson, M. J. Evans, *Nature* **326**, 295 (1987).
93. H. E. Varmus, *Adv. Oncol.* **3**, 3 (1987).
94. I thank members of my laboratory for helpful discussions of topics reviewed here, P. Brown, for pertinent collaboration on (*31*), M. Bishop for long-standing shared interests in retroviruses, L. Levinton for textual criticism, N. Hardy for the drawings in Figs. 1 and 2A, J. Marinos for Fig. 4 and help with the manuscript, and T. Jacks for Fig. 5. Our work is supported by grants from NIH.

Research on Bacteria in the Mainstream of Biology

BORIS MAGASANIK

The study of the genetics, biochemistry, and physiology of bacteria during the last 40 years has provided the concepts and methods for the study of cells of all types at the molecular level. Although much is already known about the mechanisms bacteria use to regulate the expression of their genes, a great deal more remains to be discovered that will have relevance to both prokaryotic and eukaryotic cells. Similarly, the study in bacteria of the transactions of DNA, of the synthesis and function of the cell membrane, of differentiation, and of the interaction with eukaryotic cells will undoubtedly produce results of general importance. The advantages of using bacteria for these studies include their simple noncompartmented structure, the accessibility of their genetic material, and the possibility of correlating the expression of a gene in the intact cell with its expression in a system composed of highly purified components. Finally, the comparative study of a wide variety of microorganisms may result in a better understanding of the evolution of prokaryotes and eukaryotes and lead to a comprehensive theory of cell biology.

The author is at the Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

Already I know much, but would like to know all.—Goethe's Faust

IN THEIR INTRODUCTORY CHAPTER TO THE RECENTLY PUBlished treatise *Escherichia coli and Salmonella typhimurium, Cellular and Molecular Biology* (*1*), Schaechter and Neidhardt conclude with the statement: "Not everyone is mindful of it, but all cell biologists have two cells of interest: the one they are studying and *Escherichia coli*" (*2*). This view correctly reflects the great contribution the study of this prokaryotic organism has made to the current concepts of the biology of eukaryotic microbial, plant, and animal cells. Yet, less than 50 years ago, in 1954, Kluyver and Van Niel, two eminent microbiologists, found it necessary to devote five lectures at Harvard University to convince their audience that the study of microbes could make a major contribution to biology (*3*). As late as 1942, J. S. Huxley expressed the view that bacteria may lack a genetic system analogous to that of higher organisms (*4*). It was only in 1943, when Luria and Delbrück reported the results of their experiments on the statistics of mutation in *E. coli,* that it was clearly shown that changes in the phenotype that had been observed in bacteria were not due to a direct effect of the environment, but arose from spontaneous genetic alteration followed by Darwinian selection (*5*). In 1944, the identification by Avery and his collaborators of the material responsible for the transformation of cells of *Streptococcus pneumoniae* as DNA, whose presence in the nuclei of higher cells was well established, confirmed the concept of the unity