Genetic Engineering of Novel Genomes of Large DNA Viruses

Bernard Roizman and Frank J. Jenkins

Summary. Analyses of the function of specific genes and sequences of large DNA viruses such as herpesviruses and poxviruses present special problems because of the size of their genomes (120 to 250 kilobase pairs). Various methods for engineering site-specific insertions or deletions based on the use of selectable markers have been developed and applied for the elucidation of the function of specific DNA sequences, the identification of genes nonessential for virus growth in cell culture, and the expression of foreign genes. These methods should also make possible the construction of viral vectors capable of delivering genes specifying antigens for the prevention of infectious diseases in humans and animals.

Viral genomes encode a vast variety of information. In addition to coding sequences for proteins, these genomes encode specific signals for the initiation and termination of transcription, the regulation of gene product abundance, and the modification of genome structure necessary for gene expression and genome replication. One central objective of studies on the molecular biology of viruses infecting human and animals is the elucidation of the functions encoded in the genomes of the viruses. A second objective, no less important, is the specific modification of the viral genomes for use as vaccines or as vectors of genes whose products provide protection against infectious disease.

Identification of nucleotide sequence function, unlike that of gene products, requires mutagenesis by nucleotide substitution, deletion, or insertion. In some instances the function of specific nucleotide sequences within viral genomes can be deduced from studies of isolated domains of the genome in vitro or in cells. However, the effects of the mutations observed in vitro and in cells must be confirmed by testing viral genomes in which these modifications have been introduced. Construction of novel genomes carrying insertions, deletions, or substitutions is therefore essential not only for the attenuation of viruses or the construction of virus vectors but also for analyses of function.

With small DNA viruses that can be cloned in their entirety (*1*), or DNA viruses that can be cleaved by various restriction endonucleases into a small number of fragments that can reassemble in the proper order with relatively good efficiency (*2, 3*), the construction of novel genomes carrying the desirable modifications may be tedious but not impossible. This is not the case, however, for genomes of such large DNA viruses as poxviruses (200 kilobase pairs) or herpesviruses (120 to 250 kbp). The sheer size of the genomes renders the elegant techniques developed for genetic manipulation of papovaviruses (5 kbp) and adenoviruses (35 kbp) totally inappropriate.

In this article, we describe techniques designed for the genetic manipulation of large genomes such as those of poxviruses and herpesviruses. Because these techniques have been applied to solve specific questions regarding the function of domains of the genomes of the herpes simplex virus 1 (HSV-1) and because the properties of the novel genomes generated by these techniques are relevant for an assessment of the potential of these techniques, it is both convenient and desirable to begin with a brief description of these genomes and the information they encode.

Bernard Roizman is a Joseph Regenstein Distinguished Service Professor of Virology and Frank J. Jenkins is a postdoctoral fellow at the Marjorie B. Kovler Viral Oncology Laboratories, University of Chicago, Chicago, Illinois 60637.

The HSV-1 Genome: Structure and Information Content

The genome. The HSV-1 genome extracted from virions is linear, double-stranded DNA approximately 150 kbp in size (*4*). It is convenient to view the genome as consisting of two covalently linked components designated as L (long) and S (short) (Fig. 1). Both L and S components consist of largely unique sequences (U_L and U_S) flanked by inverted repeats (*5, 6*). The inverted repeat sequences of the L component are each 9 kbp in size and have been designated as *ab* and *b'a'*, whereas those of the S component are each 6.5 kbp in size and have been designated as *c'a'* and *ca* (*6*). The *a* sequence is the only one shared by both L and S components. It is approximately 200 to 500 bp in size, depending on the number of reiterations of small sequences contained in its domain (*7, 8*). Only one *a* sequence is located at the terminus of the S component, whereas from one to more than five may be present at the terminus of the L component and at the junction between the L and S components (*7, 9, 10*). The sequence arrangement of the HSV-1 genome may be represented by

$$a_nb\text{-}U_L\text{-}b'a'_mc'\text{-}U_S\text{-}ca$$

where the primes indicate sequence inversion and *n* and *m* are the number of copies of *a* sequences.

A remarkable property of the HSV-1 genome is that the L and S components can invert relative to each other. As a consequence, viral DNA extracted from virions or from infected cells consists of four equimolar populations differing solely with respect to the relative orientation of the two components (*11*).

To date, three origins have been mapped in the HSV-1 genome. Of these, two are located within the inverted repeats (*c* sequence) of the S component, whereas the third is located in the middle of the L component (*12–17*). It is not clear whether the three function together or sequentially and whether all three are essential. The terminal *a* sequence of the S component is the *cis*-acting site for packaging of the mature viral DNA into preformed capsids (*14, 18*).

The functions of the inverted repeats and of the inversions of the L and S components is but one of the problems approached experimentally by construction of novel, genetically engineered genomes.

Gene arrangement and regulation. By enumerating the relatively abundant novel proteins appearing in infected

cells, it was originally estimated that the HSV-1 genome encodes approximately 50 polypeptides (*19*). Current estimates suggest that the number is closer to 70. Two observations are of particular significance. First, although herpesvirus genes carry transcriptional and translational signals similar to those of other DNA viruses infecting higher eukaryotic cells, the messenger RNA's arising from a vast majority of the genes are not spliced (*20*). Second, the information density is lower than that encoded in the genomes of smaller viruses, such as the papovaviruses.

The 50 genes specifying abundant products form coordinately regulated groups whose expression is sequentially ordered in a cascade fashion (*21, 22*). The five α genes, which make up the first group to be expressed, are operationally defined as the genes transcribed in the infected cells in the absence of de novo, viral protein synthesis. The β genes are operationally defined by their requirement for functional α gene products and by a total lack of sensitivity to inhibitors of viral DNA synthesis on their expression. The β genes are heterogeneous with respect to temporal patterns of their expression. The β_1 genes are generally expressed earlier than the β_2 genes, and in some cell lines they are difficult to differentiate from α genes. The onset of expression of β genes coincides with a decline in the rate of expression of α genes and irreversible shutoff of host protein synthesis. The γ genes are also heterogeneous. The γ_1 genes are expressed earlier and are differentiated from β genes solely by their requirement for viral DNA synthesis for maximal expression. In contrast, the γ_2 genes are expressed later in infection and stringently require viral DNA synthesis for their expression (*23*).

The α proteins are translocated into the nucleus. Ample evidence supports the view that the α4 gene product is a large multifunctional protein required for transcription of both β and γ genes (*24–26*). The products of the α22 and α27 genes play a role in the expression of γ genes (*27–29*), but their function—as well as the functions of the two other α genes (α0 and α47)—is less well understood than that of α4. The β genes identified to date appear to play a major role in the replication of viral DNA, in the shutoff of the expression of α genes [reputed to be a function of the major DNA-binding protein ($\beta_1$8) (*30*)], and in the induction by as yet unknown means of γ gene expression. The γ genes appear to be structural components of the virion. The

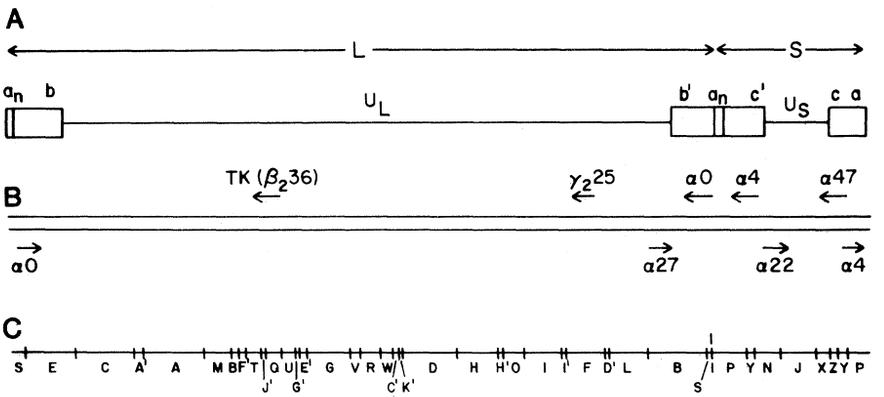

Fig. 1. Schematic of the HSV-1 genome. (A) Locations of unique sequences (U_L and U_S), terminal repeat sequences ($a_n b$ and ca), and inverted internal repeat sequences ($b' a_n c'$). (B) Locations of certain α, β, and γ genes; the arrows represent direction of transcription and the numbers refer to the infected cell polypeptide (ICP) number of the corresponding gene. (C) Bam HI restriction endonuclease cleavage sites on the HSV-1 (F) genome.

mechanism of the shutoff of β gene expression by γ gene products is not understood, but it does not appear to be due to competition for transcriptional or translational factors (*22*). At least one γ gene product introduced into the cells during infection causes a transient, but not essential, shutoff of host macromolecular metabolism (*31, 32*). Another γ gene product abundantly represented in the virion appears to have a major role in the induction of α gene expression immediately after infection (*27, 33–38*).

A pattern of gene clustering has been observed, but it is not totally consistent. The α genes are located in or near the inverted repeats, and two genes (α0 and α4), located entirely within the repeats, are diploid (Fig. 1) (*24, 39, 40*). All α genes are transcribed from independent promoters that share AT-rich homologs required for the induction of these genes by the γ *trans*-inducing protein and GC-rich enhancer-like elements (*41*). Most of the β and γ genes are dispersed throughout the genome, but small clusters of genes with related functions have been noted.

General strategy of virus multiplication. Following infection and multiplication at the portal of entry, the virus may infect nerve endings and ascend to the trigeminal and dorsal root ganglia and remain latent in neurons for the life of the host (*42*). Periodically, physical or emotional trauma or fever causes the latent virus to multiply, descend to or near the portal of entry, and initiate lesions apparent as vesicles and known as fever blisters.

Infection of cells is initiated by the fusion of the membrane covering the virus—the envelope—with the plasma membrane of the cell (*43*). The capsid is released into the cytoplasm and is transported to a nuclear pore, where the DNA is released into the nucleus (*44*). In the nucleus the viral DNA is transcribed by host RNA polymerase II (*45*) and ultimately specifies the α, β, and γ proteins described above. Viral DNA is replicated by a viral DNA polymerase (*46*).

Late in infection, the nascent viral DNA appears to be present in head-to-tail concatemers from which unit-length viral DNA is excised and packaged into capsids (*47, 48*). The capsids containing viral DNA acquire a new surface protein (*49, 50*), adhere to the underside of the nuclear membrane, and become enveloped by patches in the membrane containing viral membrane proteins (*51, 52*). The enveloped particles are transported to the cell surface through the endoplasmic reticulum, which is also modified by the insertion of the viral membrane proteins (*52*).

A characteristic of the larger DNA viruses that differentiates them from the smaller papovaviruses and adenoviruses is that they encode numerous enzymes involved in DNA metabolism. Among these enzymes are a DNA polymerase (*46*), a deoxynucleotide kinase (thymidine kinase) (*53*), ribonucleotide reductase (*54, 55*), deoxyribonuclease (DNase) (*46*), and uridine triphosphatase (UTPase) (*56, 57*). Not all of the gene products involved in viral DNA metabolism have been identified, but many do have counterparts in host cells. Because cells in culture express some gene products that can substitute for the products of viral genes, it has been possible, in some instances, to inactivate or delete the corresponding viral gene (*27, 58*). One objective of genetic engineering of novel HSV genomes is to identify viral genes whose function is similar to that of host genes and can be deleted.

Principles of Genetic Engineering of Novel HSV Genomes

Basic strategy. Deproteinized HSV DNA is infectious (*59–61*). Cotransfection of intact DNA with a molar excess of a mutagenized DNA fragment will result in progeny virus in which the mutagenized fragment recombines and replaces homologous DNA sequences (*62*). The recombination frequency varies, depending on the size and location of nonhomologous sequences and the size of homologous flanking sequences. For example, a 1-kbp insert located at the end of a 2-kbp DNA fragment is less likely to recombine than 5 kbp inserted in the middle of the same fragment. Recombinants can be isolated by screening for either DNA or expression of a gene contained in the insert (for example, β-galactosidase), but the work is tedious and time-consuming.

Another problem is that viruses modified by insertional mutagenesis may grow more slowly and yield less progeny than the wild-type parent. In consequence, the slower growing recombinants may be overgrown by the parent, wild-type virus. The isolation of desirable recombinants is greatly facilitated by the use of a selectable marker. Under selective pressure only the recombinants carrying the selectable marker are able to grow and are readily isolated by plaque purification.

Although several selectable markers are available (*63*), the one chosen by this laboratory is the HSV thymidine kinase (TK) gene. This gene was chosen because procedures are available both for and against the expression of the selectable marker, thereby permitting the selection of TK^- and TK^+ recombinant progeny (*12, 27*).

Properties of the HSV TK. Cells in culture have two pathways for the synthesis of thymidine monophosphate (TMP). The major pathway involves the conversion of deoxyuridine monophosphate to TMP by thymidylate synthetase. The second, or scavenger, pathway involves phosphorylation of thymidine by TK. Several herpesviruses and the poxviruses encode a TK. This enzyme appears to be important for normal virus growth in experimental animals (*64*), but it is not essential for growth in cell cultures because mutants lacking a functional TK gene have been known for many years (*58*). The host TK enzyme is also not essential for cell growth, provided that the primary pathway for TMP synthesis is not obstructed by metabolic inhibitors.

The HSV TK differs in two important respects from its cellular counterpart. First, the name is a misnomer because the viral enzyme phosphorylates deoxypurines and deoxypyrimidines whereas the host TK does not phosphorylate deoxycytidine or deoxypurines (*53*). The second and more important difference is that the substrate range of the viral TK for deoxynucleotide analogs is rather wide and includes analogs that bear little resemblance to deoxynucleotides, whereas the substrate range of the host TK is restricted to a small number of analogs (for example, bromodeoxyuridine) that closely resemble thymidine.

In consequence, analogs that are phosphorylated by the viral TK (for example, arabinosylthymine and acyclovir), but not by the host TK, will not affect uninfected TK^+ cells and will permit the multiplication of TK^- viruses but not that of TK^+ viruses in these cells. Arabinosylthymine and acyclovir can be used, therefore, to select TK^- viruses in both TK^+ and TK^- cell lines, but bromodeoxyuridine can be used for this purpose only in TK^- cells inasmuch as this analog is phosphorylated by the host TK.

Obstruction of the thymidylate synthetase pathway for TMP biosynthesis by such inhibitors as methotrexate or aminopterin is not deleterious in TK^+ cells but will result in the depletion of the thymidine triphosphate (TTP) pool and destruction of TK^- cells. Because of the depletion in the TTP pool, TK^- cells overlaid with medium containing methotrexate or aminopterin will permit the growth and plaque formation of TK^+ viruses but not of TK^- viruses. TK^+ viruses are therefore readily selected in TK^- cells overlaid in HAT medium, which contains hypoxanthine, aminopterin (or methotrexate), and thymidine

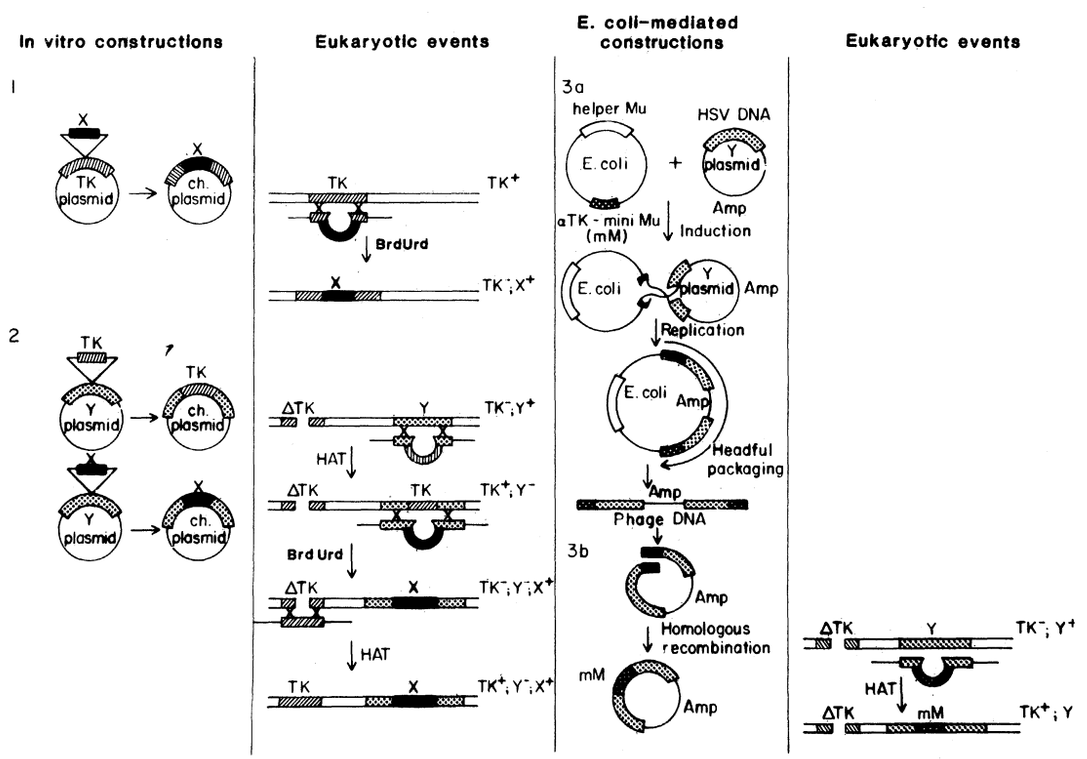

Fig. 2. Schematic outline of procedures used to construct novel HSV genomes (see explanation in text).

(hypoxanthine and thymidine being used to compensate for other metabolic products whose synthesis is blocked by these drugs).

It should be pointed out that the selection of TK^- recombinants is generally more efficient than that of TK^+ recombinants. Although spontaneous TK^- mutants are readily obtainable, the frequency of such mutations in a cloned viral population is 10^{-4} to 10^{-5}—that is, lower than the recombination rate between viral DNA and a DNA fragment containing an insertion or deletion flanked by at least 2 kbp of DNA homologous to contiguous DNA sequences in the viral genome. Furthermore, in cells doubly infected with a TK^+ parent and a TK^- recombinant, both viruses are destroyed by the analog phosphorylated by the viral TK.

The TK^+ selection suffers from two problems. First, the uninfected TK^- cells survive only a few days in the presence of methotrexate or aminopterin, and debilitated TK^+ recombinants may not have sufficient time to spread and form plaques. Furthermore, in cells that are doubly infected with the TK^- parent and TK^+ recombinant, the TK^+ parent acts as a helper and both viruses grow.

Construction of novel genomes. Figure 2 illustrates three strategies used for construction of recombinant genomes based on the use of TK as a selectable marker. The first strategy (Fig. 2-1) involves the insertion of a sequence into the transcribed domain of the TK gene. The DNA fragment carrying the TK gene with the insert is then amplified by cloning in *Escherichia coli* and cotransfected with intact HSV wild-type DNA. TK^- recombinants carrying the insert are then selected from among the progeny of transfection as described above. This technique allows rapid selection of recombinant progeny, but it restricts the insertions to the domain of the TK gene. The preferred, but not unique, sites for insertional mutagenesis of the TK gene (Fig. 3) are the Bgl II site in the transcribed noncoding sequences and the Sac I site in the coding sequences (*12, 27*).

The second strategy permits the construction of recombinants carrying insertions or deletions at specifically selected sites. In the first step of this procedure, sequences are deleted near one end of the domain of the TK gene. The DNA fragment containing the TK gene carrying the deletion is cotransfected with intact wild-type DNA, and TK^- progeny carrying the deletion are then selected from among the progeny of transfection. The TK^- recombinant constructed for these studies—HSV-1(F)Δ305—contains a 700-bp deletion extending downstream from the site of transcription initiation ($^+1$) beyond the Sac I site ($^+540$). Although this recombinant makes a truncated 19-kilodalton protein that reacts with antibody to TK (*65*), it is unable to convert thymidine to TMP.

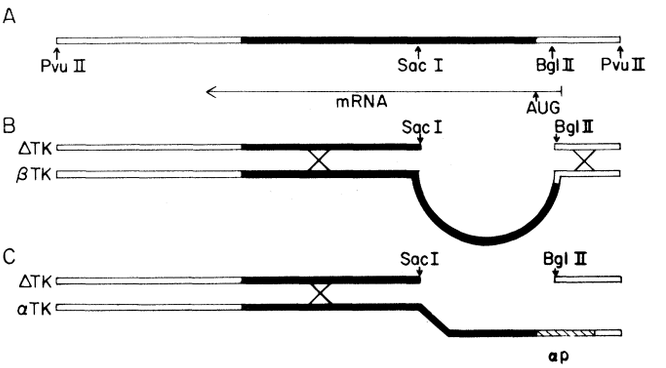

Fig. 3. (A) Schematic of Pvu II DNA fragment from HSV-1 containing the wild-type TK gene; the horizontal arrow represents the location of the TK messenger RNA transcript, and the solid line represents the coding region for the TK protein. (B) Homologous recombination between the deleted TK gene (ΔTK) in HSV-1(F) Δ305 (top line) and an intact wild-type TK gene (βTK) (bottom line). (C) Illustration of nonhomology between the 5′ end of the deleted TK gene and the 5′ end of the chimeric αTK gene used for insertional mutagenesis at sites other than the natural TK gene; αp refers to the location of the α promoter and regulatory region fused to the structural TK gene in the αTK chimeric gene.

In the second step (Fig. 2-2), the TK is inserted into a HSV DNA fragment at a specific site. In initial studies the wild-type TK gene was used (*27*). In subsequent studies, a chimeric gene consisting of the coding sequences of the TK gene fused to the promoter regulatory domain of α genes was used in order to reduce the probability that the TK gene will recombine at its natural location and to restore the integrity of the wild-type gene (*66, 67*).

Because the deletion in the natural TK gene extends both upstream and downstream from the Bgl II site, whereas the new promoter regulatory domain extends upstream from the Bgl II site (Fig. 3), the probability that the chimeric gene will restore TK activity at the natural site is greatly reduced; in fact, such recombinants have not been observed. Viral TK^+ recombinants carrying the TK insert at the desired site are selected by plating the progeny of transfection on TK^- cells maintained in HAT medium. An appropriate deletion or insertion is made in the target sequence at the site of insertion of the TK gene to delete or insert a specific HSV or foreign gene in the domain of the target sequence. This DNA fragment is then amplified and cotransfected with intact DNA of the recombinant carrying the TK gene inserted in the target sequence. Viral TK^- recombinants carrying the deletion or insertion in the target gene are then selected from among the progeny of transfection as described above.

When it is desirable to assess the effect of the deletion in the target sequence only, the native TK gene can be restored by cotransfecting the DNA of this recombinant with DNA fragments carrying the entire domain of the TK gene. The TK^+ progeny of this transfection should differ from the wild-type virus only with respect to the deletion or insertion in the target sequence (Fig. 2-2).

This strategy permits the selection of viable recombinants carrying insertions or deletions at specific sites. The major disadvantages are that the location of the target sequence within the cloned DNA fragment must be known. Furthermore, this sequence must contain a suitable restriction endonuclease cleavage site for the insertion of the TK gene. For example, if the objective is to inactivate by insertional mutagenesis a gene that overlaps in part with the domain of another gene either at its 3′ or 5′ terminus, a suitable site would be upstream or downstream from the overlapping regions. In some instances, no suitable sites are found or the precise domain of the target gene is not known.

The third strategy (Fig. 2-3, a and b), a modification of the second, involves the use of the mini-Mu derivative of the transducing bacteriophage Mu (*67, 68*). The mini-Mu prophage contains the left and right ends of the bacteriophage Mu plus the ner, A and B genes, a temperature-sensitive *c* repressor gene of Mu, and a selectable marker. The mini-Mu prophage is defective and requires a helper Mu for replication, but the presence of the left and right ends—plus the A and B genes of Mu—allow for transposition within an *E. coli* cell.

For use in insertional mutagenesis of HSV genomes, a functional TK gene

under the control of the HSV α4 gene promoter and regulatory region was inserted into the mini-Mu to yield an αTK mini-Mu (αTK-mM) (*67*). The αTK-mM was then used to lysogenize an *E. coli* strain containing a helper Mu prophage with a temperature-sensitive repressor *c* gene. Growth of the double lysogen *E. coli* strain at 30°C prevents replication and transposition of both the helper Mu and αTK-mM prophages.

The use of the αTK-mM system for insertional mutagenesis requires three steps. In the first step (Fig. 2-3a), the double lysogen *E. coli* strain is transfected with plasmid DNA containing the target HSV-1 DNA fragment. Induction by shift up to 42°C of the bacterial cells results in replication and transposition of both the helper Mu and αTK-mM. As a result of transposition, the αTK-mM will integrate into many different sites within the plasmid DNA and produce a cointegrate structure integrated into the *E. coli* DNA consisting of the plasmid DNA flanked by single copies of the αTK-mM.

In the case of the cointegrate structures, packaging begins with the leftmost copy of the αTK-mM and proceeds through the plasmid DNA and into the second copy of the αTK-mM. If the cointegrate structure is less than 38 kbp, then flanking *E. coli* DNA sequences will be packaged. Cointegrate structures greater than 38 kbp result in deletions of the second αTK-mM copy. The αTK-mM is 9.6 kbp in size. Therefore, recombinant plasmids 18 kbp or more in length will cause deletion of part or all of the second copy of the αTK-mM. As is evident below, at least a portion of the second copy of the αTK-mM is essential, and therefore the recombinant plasmid must be less than 28 kbp in size.

The lysates of the double lysogen transfected with plasmids containing HSV DNA and induced by shift up to 42°C should contain phages with defective genomes consisting of the recombinant plasmid DNA's linearized at random sites and flanked by direct copies of the αTK-mM. The second step (Fig. 2-3b) involves infection of an *E. coli* Rec(A^+) strain lysogenized by a helper Mu prophage. The presence of the helper Mu prevents the cointegrate structure from transposing to the *E. coli* DNA. Homologous recombination between the flanking copies of the αTK-mM results in the production of plasmid molecules in which the αTK-mM is randomly inserted into the recombinant plasmid. *E. coli* carrying the plasmid molecules can then be readily grown by selection for the antibiotic resistance marker contained in the αTK-mM. The isolated plasmid DNA consists of a pool of plasmid molecules containing an αTK-mM inserted at many different sites within the HSV DNA insert and nonessential vector sequences.

In the third step (Fig. 2-3c), the amplified αTK-mM plasmids are cotransfected with intact TK^- HSV DNA, and the progeny of the transfection are then plated on TK^- cells under HAT medium. The TK^+ progeny carrying the αTK-mM insert may be heterogeneous inasmuch as the αTK-mM may be expected to insert at any site within the target sequences of HSV DNA that is not essential for growth in the cells in which the selection is done.

The αTK-mM system offers several advantages. Foremost, it obviates the need for inserting the selectable marker at a specific and often inaccessible site. The production of recombinant plasmids containing randomly inserted αTK-mM is rapid—less than 48 hours. The αTK-mM is an excellent probe for sequences and genes nonessential for growth in the cells used for the selection of recombinant viruses. Furthermore, as is the case of the inserted TK gene, the αTK-mM can be replaced by recombination with an insertion or deletion in the target sequence.

Application of Genetic Engineering to Construction of Novel HSV Genomes

The techniques described above have been used to insert HSV and foreign sequences into HSV genomes and to probe HSV-1 DNA for domains not essential for growth in cell culture. Among the central issues to be considered are the capacity of the genome as a vector for HSV or foreign genes, stability of the recombinants, and privileged sites characterized by spontaneous rearrangements.

Minimum and maximum sizes of HSV genomes that can be packaged. Current studies by Frenkel and associates (*14, 18*) and others (*69*) indicate that packaging of the genome requires the presence of terminal *a* sequences and is not by a simple headful mechanism. The capsids are capable of packaging small head-to-tail concatemers of defective genomes (*18*), but capsids containing concatemers significantly smaller than the wild-type genome are not enveloped. One hypothesis that could explain this observation is that capsids containing standard-length DNA become modified and bind an additional protein (*50*) and that capsids packaging less than a minimum-size DNA do not become modified and fail to bind this protein.

The minimum size of the packaged genome is not known but the smallest nondefective genome packaged efficiently is that of the I358 recombinant in which a 2-kbp sequence containing the TK gene replaced 15 kbp of DNA that included a portion of the unique sequences and nearly the entire internal inverted repeat sequence (*70*). The largest genomes packaged to date have included an αTK-mM 9.6 kbp in size (*67*). Taken together, these results indicate that the 2-kbp sequence containing the TK gene in an I358 recombinant could be replaced by an insert at least 24 kbp in size.

Sites and sequences not essential for growth in cell culture. The sites and sequences shown to date to be nonessential for viral replication include (Fig. 1) (i) various portions of the domain of the TK gene (*27*), (ii) the sequences located between the 3′ terminus of the α27 and α0 genes (*67, 70*), (iii) the sequences (approximately 15 kbp) located between the 3′ terminus of the α27 gene and the promoter-regulatory domain of the α4 gene and constituting unique sequences of the L component and nearly the entire internal inverted repeat sequence $b'a_mc'$ (*67, 70*), (iv) the sequences located between the terminal *a* sequence and the 3′ terminus of the α4 gene at both ends of the S component (66), and (v) the sequences flanking the origins of DNA synthesis of the S component (*71*) and portions of the domains of the α22 gene (*27*).

The recombinants carrying inserts or deletions in these regions of HSV DNA vary in their capacity to grow. The recombinants R325 and R328 carrying 500- and 100-bp deletions in the coding sequences of the α22 gene grow as well as the wild-type parent in Vero and HEp-2 cells but poorly in human embryonic lung cells and in various rodent cell lines. Studies on these recombinants suggest that a host factor is able to complement the deleted gene in Vero and HEp-2 cell lines but not in the restrictive cells (*29*). The I358 recombinant described above yields approximately one-tenth as much virus as the wild type in Vero and HEp-2 cells and considerably less virus in rodent cell lines or human embryonic cell lines. A characteristic of this virus is that its DNA is frozen in one arrangement; that is, the L and S components do not invert relative to each other.

The insertion of the TK gene, the selectable marker, into appropriate sites

of target sequences has been the rate-limiting step in the analyses of HSV DNA for sequences and genes not essential for virus growth in the cells used for selection of recombinant genomes. The development of the αTK-mM system for insertional mutagenesis will considerably facilitate these studies.

Insertion of HSV and foreign sequences into HSV genome. The sequences and genes inserted into HSV-1 DNA to date include (i) HSV DNA sequences inserted for the purpose of identifying putative *cis*-acting sites; (ii) HSV genes inserted for the purpose of identifying their function by complementing the authentic, mutagenized copy; and (iii) foreign genes inserted to determine whether the HSV genome can serve as a vector for the expression of foreign genes.

Specifically, to identify the promoter-regulatory domains of the α genes, a fragment containing the capping site and upstream sequences of both copies of the α4 gene was inserted into the Bgl II site (nucleotide +50) of the TK gene in the proper transcriptional orientation. In the resulting recombinants, the chimeric TK gene was regulated as an α gene (*27*). Similar insertion of a fragment carrying a putative γ_2 promoter-regulatory sequence converted the natural βTK into a γ_2 gene (*72*). Insertion of DNA fragments containing the *a* sequence into the Bgl II site of the TK gene led to the identification of this sequence as the site-specific recombination site for the inversion of the L and S components (*12*).

The HSV genes specifically inserted into the genome were designed to complement mutations in the native gene (*73*) or to express a mutated gene (*74*). The foreign genes inserted into the genome were the *S* gene of hepatitis B virus (*75*), chick ovalbumin gene (*76*), and the EBNA 1 and EBNA 2 genes of Epstein-Barr virus (*77*). The processing of products of foreign genes inserted into HSV-1 vectors has not been studied extensively; normally, secreted proteins are secreted from cells infected with HSV-1 vectors carrying their genes (*75*).

Requirements for expression of foreign genes. Host genes and genes of other viruses linked to their natural promoters are expressed poorly if at all in HSV vectors (*78*). When linked to promoter-regulatory domains of HSV genes, the foreign genes are expressed and regulated as the authentic HSV genes whose promoter was "borrowed" (*75*). HSV-2 genes are expressed in HSV-1 vectors under their own promoters, and it is conceivable that genes of other herpesviruses may also be expressed in HSV-1 vectors under their own promoters.

Stability of genetically engineered HSV-1 genomes. Several major types of rearrangement in genetically engineered strains have been noted. These include the following:

1) Cell-specific restrictions. An example of this phenomenon is recombinant R316 constructed by the insertion of Bam HI N fragment of HSV-1 (Fig. 1) into the Bgl II site of the TK gene (Fig. 3) and converting the natural (β) TK into an α-regulated gene (*27*). This recombinant is stable in Vero cells; in various other cell lines and in experimental animals, mutants carrying various size deletions in the inserted Bam HI N fragment emerge and rapidly overgrow the R316 recombinant (*29*). No explanation for this bizarre, cell-specific restriction has emerged so far.

2) Equalization of genetically related duplicated genes. Insertion of a gene closely related but nonidentical to that resident in the HSV genome may result in the accumulation of identical recombinant sequences as a consequence of genetic recombination and segregation (*79*).

3) Insert-dependent defective genomes. Insertion of a second copy of a sequence in the same orientation relative to the first copy may result in the generation of head-to-tail reiterations of novel subsets of defective HSV genomes in which all sequences between the two copies of the reiterated sequence had been deleted. Defective genomes require an origin of DNA synthesis for amplification and a terminal *a* sequence for packaging by factors supplied by the helper virus. In addition, the length of the subset must be such that an integral number of tandem reiterations fall within the minimum- and maximum-size DNA that is both packaged and enveloped (*80*). The insertion of a second copy of sequences in the inverted orientation relative to the native copy has resulted, in some instances, in the inversion of the sequences flanked by the inverted repeats (*12, 79*). Whether a particular set of inverted repeats will cause inversions is totally unpredictable and may reflect recombinational "hot spots" within the inserted sequence.

4) Privileged sites. These sites are defined as most likely to result in major rearrangements of the HSV genome. Thus, insertion of sequences at or near the termini of the inverted repeats may result in the spontaneous deletion of most of the inverted repeat sequence $b'a_n'c'$ (Fig. 1) (*67, 70*). To date, insertions inside the inverted repeats have not caused such deletions (*66, 81*). Insertions within one of the inverted repeat sequences of the S component ($c'a'$ or ca) (Fig. 1) result in the duplication of the insertion in the other repeat (*66*). The duplication is almost obligatory if the insert is close to the *a* sequence but diminishes in frequency with distance.

Construction of Novel Genomes of Other Large DNA Viruses

The strategies for construction of novel genomes described in this article have been used for the insertion and expression of foreign genes in vaccinia virus. Like HSV, vaccinia virus encodes a TK gene, and both the vaccinia and HSV TK genes have been used to insert foreign genes linked to vaccinia virus promoters (*82–86*). The extensive experience obtained with vaccinia virus in the course of its use as a vaccine for the eradication of smallpox has been cited as the basis for its use as a vector of vaccines for the prevention of a wide variety of infectious diseases. The genomic capacity of vaccinia virus to carry foreign genes is at least equal to that of HSV.

Conclusions

Various techniques based on the use of selectable markers are now available for the analysis of the function of specific domains of large DNA genomes and for the construction of vectors capable of delivering vaccines for the prevention of infectious diseases in humans and animals.

References and Notes

1. K. W. C. Peden, J. M. Pipas, S. Pearson-White, D. Nathans, *Science* **209**, 1392 (1980).
2. Q. S. Kapoor and G. Chinnadurai, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2184 (1981).
3. J. E. Mertz, J. Carbon, M. Herzberg, R. W. Davis, P. Berg, *Cold Spring Harbor Symp. Quant. Biol.* **39**, 69 (1974).
4. E. D. Kieff, S. L. Bachenheimer, B. Roizman, *J. Virol* **8**, 125 (1971).
5. P. Sheldrick and N. Berthelot, *Cold Spring Harbor Symp. Quant. Biol.* **39**, 667 (1974).
6. S. Wadsworth, R. J. Jacob, B. Roizman, *J. Virol.* **15**, 1487 (1975).
7. E. S. Mocarski and B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 7047 (1981).
8. A. J. Davison and N. M. Wilkie, *J. Gen. Virol.* **55**, 315 (1981).
9. M. M. Wagner and W. C. Summers, *J. Virol.* **27**, 374 (1978).
10. H. Locker and N. Frenkel, *ibid.* **32**, 424 (1979).
11. G. S. Hayward, R. J. Jacob, S. C. Wadsworth, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4243 (1975).
12. E. S. Mocarski and B. Roizman, *Cell* **31**, 89 (1982).
13. N. Frenkel, H. Locker, W. Batterson, G. Hayward, B. Roizman, *J. Virol.* **20**, 527 (1976).
14. D. A. Vlazny and N. Frenkel, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 742 (1981).
15. N. D. Stow, *EMBO J.* **1**, 863 (1982).

16. _____ and E. C. McMonagle, *Virology* **130**, 427 (1983).
17. S. K. Weller *et al.*, *Mol. Cell. Biol.* **5**, 930 (1985).
18. D. A. Vlazny, A. Kwong, N. Frenkel, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1423 (1982).
19. R. W. Honess and B. Roizman, *J. Virol.* **12**, 1346 (1973).
20. E. K. Wagner, in *The Herpesviruses*, B. Roizman, Ed. (Plenum, New York, 1985), vol. 3, pp. 45–104.
21. R. W. Honess and B. Roizman, *J. Virol.* **14**, 8 (1974).
22. _____, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1276 (1975).
23. A. J. Conley, D. M. Knipe, P. C. Jones, B. Roizman, *J. Virol.* **37**, 191 (1981).
24. D. M. Knipe, W. T. Ruyechan, B. Roizman, I. W. Halliburton, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3896 (1978).
25. D. M. Knipe, W. T. Ruyechan, R. W. Honess, B. Roizman, *ibid.* **76**, 4534 (1979).
26. R. A. F. Dixon and P. A. Schaffer, *J. Virol.* **36**, 189 (1980).
27. L. E. Post and B. Roizman, *Cell* **25**, 227 (1981).
28. C. Greene and P. Schaffer, *Abstracts of the Ninth International Herpesvirus Workshop* (1984), p. 11.
29. A. E. Sears, I. W. Halliburton, B. Meignier, S. Silver, B. Roizman, *J. Virol.*, in press.
30. P. J. Godowski and D. M. Knipe, *ibid.* **47**, 478 (1983).
31. M. L. Fenwick and M. J. Walker, *J. Gen. Virol.* **41**, 37 (1978).
32. G. S. Read and N. Frenkel, *J. Virol.* **46**, 498 (1983).
33. S. Mackem and B. Roizman, *ibid.* **43**, 1015 (1982).
34. _____, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 4917 (1982).
35. _____, *J. Virol.* **44**, 939 (1982).
36. W. Batterson and B. Roizman, *ibid.* **46**, 371 (1983).
37. M. E. M. Campbell, J. W. Palfreyman, C. M. Preston, *J. Mol. Biol.* **180**, 1 (1984).
38. P. E. Pellett, J. L. C. McKnight, F. J Jenkins, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
39. R. J. Watson, C. M. Preston, B. Clements, *J. Virol.* **31**, 42 (1979).
40. S. Mackem and B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7122 (1980).
41. T. M. Kristie and B. Roizman, *ibid.* **81**, 4065 (1984).
42. T. J. Hill, in *The Herpesviruses*, B. Roizman, Ed. (Plenum, New York, 1985), vol. 3, pp. 175–240.
43. C. Morgan, H. M. Rose, B. Mednis, *J. Virol.* **2**, 507 (1968).
44. W. Batterson, D. Furlong, B. Roizman, *ibid.* **45**, 397 (1983).
45. F. Constanzo, G. Campadelli-Fiume, L. Foa-Tomas, E. Cassai, *ibid.* **21**, 996 (1977).
46. H. M. Keir and E. Gold, *Biochim. Biophys. Acta* **72**, 263 (1963).
47. R. J. Jacob, L. S. Morse, B. Roizman, *J. Virol.* **29**, 448 (1979).
48. T. Ben-Porat and S. Tokazewski, *Virology* **79**, 292 (1977).
49. W. Gibson and B. Roizman, *J. Virol.* **10**, 1044 (1972).
50. D. K. Braun, B. Roizman, L. Pereira, *ibid.* **49**, 142 (1984).
51. R. W. Darlington and L. H. Moss III, *Prog. Med. Virol.* **11**, 16 (1969).
52. J. Schwartz and B. Roizman, *J. Virol.* **4**, 879 (1969).
53. S. Kit and D. R. Dubbs, *Virology* **26**, 16 (1965).
54. G. H. Cohen, *J. Virol.* **9**, 408 (1972).
55. D. Huszar and S. Bacchetti, *ibid.* **37**, 580 (1981).
56. F. Wohlrab and B. Francke, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 100 (1980).
57. V. G. Preston and F. B. Fisher, *Virology* **138**, 58 (1984).
58. D. R. Dubbs and S. Kit, *ibid.* **22**, 493 (1964).
59. D. Lando and M. L. Rhyiner, *C. R. Acad. Sci.* **269**, 527 (1969).
60. F. L. Graham, G. Velihaisen, N. M. Wilkie, *Nature (London) New Biol.* **245**, 265 (1973).
61. P. Sheldrick *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3621 (1973).
62. W. T. Ruyechan, L. S. Morse, D. M. Knipe, B. Roizman, *J. Virol.* **29**, 677 (1979).
63. J. Hubenthal-Voss and B. Roizman, unpublished observations.
64. R. B. Tenser and M. E. Dunston, *Virology* **99**, 417 (1979).
65. M. Ackermann, M. Sarmiento-Batterson, B. Roizman, in preparation.
66. J. Hubenthal-Voss and B. Roizman, *J. Virol.* **54**, 509 (1985).
67. F. J. Jenkins, M. J. Casadaban, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 4773 (1985).
68. B. A. Castilho, P. Olfson, M. J. Casadaban, *J. Bacteriol.* **158**, 488 (1984).
69. N. D. Stow, E. C. McMonagle, A. J. Davison, *Nucleic Acids Res.* **11**, 8205 (1983).
70. K. L. Poffenberger, E. Tabares, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2690 (1983).
71. P. Mauromara-Nazos, J. Hubenthal-Voss, B. Roizman, unpublished observations.
72. S. Silver and B. Roizman, *Mol. Cell. Biol.* **5**, 518 (1985).
73. G. T.-Y. Lee, K. L. Pogue-Geile, L. Pereira, P. G. Spear, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 6612 (1982).
74. M. G. Gibson and P. G. Spear, *J. Virol.* **48**, 396 (1983).
75. M.-F. Shih, M. Arsenakis, P. Tiollais, B. Roizman, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5867 (1984).
76. M. Arsenakis, K. L. Poffenberger, B. Roizman, in preparation.
77. M. Hummel *et al.*, in preparation.
78. C. Tackney, G. Cachianes, S. Silverstein, *J. Virol.* **52**, 606 (1984).
79. K. L. Pogue-Geile, G. T.-Y. Lee, P. G. Spear, *ibid.* **53**, 456 (1985).
80. K. L. Poffenberger and B. Roizman, *ibid.* **53**, 587 (1985).
81. F. J. Jenkins and B. Roizman, unpublished observations.
82. M. Mackett, G. L. Smith, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 7415 (1982).
83. G. L. Smith, M. Mackett, B. Moss, *Nature (London)* **302**, 490 (1983).
84. G. L. Smith, B. R. Murphy, B. Moss, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7155 (1983).
85. M. Mackett, G. L. Smith, B. Moss, *J. Virol.* **49**, 857 (1984).
86. S. Gillard, D. Spehner, P. Drillien, *ibid.* **53**, 316 (1985).
87. Supported in part by grants from the National Cancer Institute (CA08494 and CA19264), United States Public Health Service, American Cancer Society (MV2T), and Institut Merieux. F.J.J. is a postdoctoral trainee (CA09241) of the National Cancer Institute.

Arabidopsis thaliana and Plant Molecular Genetics

Elliot M. Meyerowitz and Robert E. Pruitt

It is worth understanding the molecular genetics of plants not only because such understanding has practical value in the improvement of crops but also because studies of plants offer an opportunity to gain insight into various basic life processes unique to plants. Plants do not use the same hormones as animals; nor do they use hormones in the same way that animals use them. Most or all of the cells in a plant may both produce and respond to plant hormones at some time. Plants respond to stress differently than animals. Plants also respond to light in diverse and subtle ways, with photosynthesis being only one of the responses of plants to light that are not found in animals. Even the basic developmental processes of plants have features that distinguish them from those of animals. Among these features are the absence of cell migration in plant development and the fact that each flowering plant has certain parts (the meristems) that remain embryonic and can produce adult organs, including germ cells, throughout the life of the plant. Further, individual differentiated cells taken from the vegetative parts of plants can dedifferentiate and regenerate to form entire, fertile plants.

Although it is sensible and necessary to study crop plants for purposes of crop improvement, the crop plants now used for basic classical and molecular genetic studies have disadvantages for some of the types of experimentation used in this work. Classical genetics depends on researchers being able to raise many successive generations of organisms in large numbers. Typical crop plants have generation times of several months, and they require a great deal of field space for growth in large numbers. The genetics of some of these plants is also made more difficult by polyploidy or allopolyploidy. The ease with which recombinant DNA work can be done with any organism depends in part on the size of the organism's nuclear genome; the smaller the genome, the less work is required to screen recombinant DNA libraries and thus to isolate any particular gene. The genomes of the plants presently used for recombinant DNA

Elliot M. Meyerowitz is an associate professor and Robert E. Pruitt is a graduate student in the Division of Biology, California Institute of Technology, Pasadena 91125.