HSV GENE FUNCTIONS: WHAT HAVE WE LEARNED THAT COULD BE GENERALLY APPLICABLE TO ITS NEAR AND DISTANT COUSINS?

B. ROIZMAN*#

The Marjorie B. Kovler Viral Oncology Laboratories, University of Chicago, 910 East 58th Street, Chicago, IL 60637, USA

Summary. – Herpes simplex virus 1 (HSV-1) encodes at least 84 polypeptides to perform two functions: to enable viral replication and to create the environment in which the entry of the virus into host cells, synthesis of virion components, assembly and egress are optimized. Whereas the former are indispensable for viral replication, the latter, numbering 47, can be deleted without a major effect on viral replication in cells in culture. Of particular interest are gene products whose function is either to modify cellular proteins (set 1) or to block entirely their function (set 2). An example of set 1 is the infected cell protein No. 0 (ICP0), a promiscuous transactivator of genes introduced into cells by infection or transfection. In its nuclear phase this protein binds to cyclin D3, extends its life by many hours, and sequesters it in nuclear structures known as ND10. In its cytoplasmic phase, ICP0 binds the translation elongation factor EF-1δ. Another viral protein, the U_L13 protein kinase, hyper-phosphorylates EF-1δ. ICP0 and the protein kinase stimulate protein synthesis and cause the cell to induce the synthesis of pre-S phase cellular proteins the virus needs for its replication. The $\gamma_1$34.5 protein, a prototype of set 2, also has multiple functions. One, mapped at its carboxyl terminus, blocks the effects of double-stranded RNA-dependent protein kinase R (PKR) that is activated by all wild-type and mutant viruses examined to date. PKR phosphorylates eIF-2α and shuts off protein synthesis. $\gamma_1$34.5 protein binds protein phosphatase 1 and redirects it to dephosphorylate eIF-2α. Although PKR is activated in wild-type-infected cells, protein synthesis is unaffected. HSV-1 encodes in addition at least two proteins, ORF O and ORF P that are repressed during productive infection. The ORF P protein localizes in spliceosomes and blocks the synthesis of viral proteins derived from spliced mRNA. The ORF O protein binds ICP4, the major regulatory protein, and prevents it from binding to DNA. The role of ORF O and ORF P proteins in the establishment of latency is uncertain. A significant discovery that has emerged from these studies is that viral proteins can perform several functions that may be totally unrelated.

Introduction

All herpesviruses have an identical mission: to multiply and spread, and to use latency as a mechanism whereby they perpetuate themselves in their respective hosts. All herpesviruses contain DNA genomes encoding numerous genes and all of them seem to encode a lot of genes that can be deleted without affecting the ability of the virus to grow in cultured cells. We know more about HSV-1 and HSV-2 than about any other herpesvirus and it seems appropriate that HSV-1, the oldest known herpesvirus also serves as a point of reference in the analyses of the function of its cousins, both near and distant.

HSV-1 spreads by physical contact between infected and uninfected individuals. The virus multiplies at the portal of entry (mouth or genitals) and both infecting and progeny viruses can infect sensory nerve endings and be transported by retrograde axonal flow to the neuronal nucleus. In sensory neurons the virus may multiply or establish a latent infection. In the course of productive infection, the virus expresses most of its genes, yields infectious progeny, and destroys the host cell. In latent infection, the virus expresses a very small fraction of its sequences and does not appear to impair the functions of its host. It is convenient to consider the viral genome first, followed by viral gene functions in productive and latent infections, respectively.

*E-mail: bernard@cummings.uchicago.edu; fax: +1-773-702-1631.

#Invited lecture.

I. The viral genome

Genographics

The HSV genome consists of two stretches of unique sequences, long (U_L) and short (U_S), flanked by inverted repeats (for references see review by Roizman and Sears, 1996). The inverted repeats flanking U_L (*ab* and *b'a'*, 9 kbp each) and those flanking U_S (*a'c'* and *ca*, >6 kbp each) together account for nearly 10% of the genome. The DNA sequences present in at least 1 copy per genome sum up to approximately 138 kbp. Approximately 15 kbp are repeated once, and the terminal a sequence of approximately 0.5 kbp may be repeated in excess of 10 times in some strains. At least 84 of the several hundred open reading frames (ORFs) identified in the viral genome are expressed – 12 more than originally predicted by McGeoch *et al.* (1988) (Roizman and Sears, 1996; Roizman, 1996). Overall, the genes are densely packed: while it could be predicted that the promoter domains of many genes would map within coding sequences of other genes, or that the ends of ORFs would overlap, the surprising and perhaps distinctive aspects of the HSV genographics is that several genes are totally antisense to each other. Another curious aspect of the HSV genography is that several ORFs yield proteins whose sequences are partially collinear. In each instance, the larger ORF encodes in its 5'-domain a promoter which initiates transcription and, ultimately translation from a downstream in-frame initiator methionine. Two examples are relevant. Thus the U_L26 ORF encodes a protein whose N-terminal domain acts as a protease. A second protein is encoded by the 3'-domain of the U_L26 designated U_L26.5. This protein has the same amino acid sequence as the carboxyl terminal domain of the U_L26 protein but it has a different function. A similar situation exists for the a22 and U_S1.5 ORFs (Roizman and Sears, 1996; Roizman, 1995). The advantage to the virus is that the synthesis of the these proteins can be regulated individually.

Unlike the situation prevailing in cells infected with other members of the family *Herpesviridae*, very few proteins are the products of spliced mRNAs. Two sets of key proteins made in this fashion should be noted. Of the six α proteins, 3 are made from spliced mRNAs. The coding domains of the gene encoding ICP0 consists of 3 exons. Whereas the coding domains of ICP22 and ICP47 consist of a single exon, the promoters of these genes located in inverted repeats are identical and each has an intron. In the case of U_L15, the coding domain consists of two exons and the intron contains the coding domains of two genes, U_L16 and U_L17 (McGeoch *et al.*, 1988).

II. The function of viral gene expression in productive infection

HSV gene expression in productive infection

Grossly, the HSV-1 genes form at least 6 groups whose expression is coordinately regulated and sequentially ordered in a cascade fashion (Roizman and Sears, 1996). In productive infection, the six α genes are expressed first followed by β_1, β_2, γ_1 and γ_2. Thus α genes are expressed in the absence of prior viral protein synthesis. HSV conveniently carries a structural protein designated α-trans-inducing factor (αTIF) or VP16 which interacts with two cellular protein complexes, C1 (also known as HCF) and Oct1 at specific response elements in the promoter of α genes, and activates their transcription (Roizman and Sears, 1996). The expression of β genes requires α proteins but these genes can be expressed in the absence of viral DNA synthesis. Whereas γ_2 genes require viral DNA synthesis for their expression, the γ_1 genes are expressed, but at suboptimal levels in the absence of viral DNA synthesis. At least 2 ORFs are expressed only in the absence of α gene expression and, as detailed later in this text, may be designated as pre-α genes. This gene nomenclature is useful but should not be misleading. The α genes share common promoter elements which identify them unambiguously as α genes (Mackem and Roizman, 1982). The remaining, β_1, β_2, γ_1 and γ_2 genes do not share promoter elements unique to each group. Rather, the timing of their expression is the consequence of the constellation of promoter elements in each gene (Roizman and Sears, 1996).

HSV gene function in productive infection

Systematic sequence-specific deletions in the viral genome indicated that of the 82 genes known to be expressed during productive infection, 37 cannot be deleted without abrogating the capacity of the virus to multiply in susceptible cells. The functions of these genes include (i) entry and egress of virus from infected cells (ii) regulation, both positive and negative, of RNA synthesis and processing, (iii) synthesis of viral DNA, (iv) key virion proteins, and (v) assembly of capsids, packaging of DNA, and envelopment. For the purposes of this review, two proteins are especially noteworthy. ICP4, an α protein, binds to DNA and regulates genes both positively and negatively depending, in part, on the location of its binding site. A key function of ICP27, also an α protein, is to block the splicing of RNA in productively infected cells (Roizman and Sears, 1996).

As many as 47 genes are dispensable for viral replication in at least some cells in culture. The key word „dispensable" is a misnomer since (i) viruses lacking this gene have not been isolated from humans and (ii) mutants lacking these

genes are frequently but not always attenuated in experimental animal systems. The functions of these genes include (i) extending the host range of the virus by enabling efficient entry, sorting and egress, (ii) more extensive regulation of viral gene expression and of the function of viral proteins, (iii) enhancement of the nucleic acid metabolism of the infected cell to favor viral DNA synthesis and repair in nondividing cells, (iv) blocking host response to infection, and (v) modification of cellular environment to render viral replication more efficient by stabilizing some cellular proteins and degrading others (Roizman and Sears, 1996; Roizman, 1996).

A few examples of the function of these accessory proteins are useful. α0, the gene encoding ICP0 maps in the inverted repeats *ab* and *b'a'*. ICP0 in transfection assay acts as a promiscuous transactivator. The protein performs multiple functions. Thus it associates with and causes the disappearance of the nuclear structures known as ND-10. It also binds to an ubiquitin-specific protease, cyclin D3, and the elongation factor EF-1δ (Kawaguchi *et al.*, 1997a,b; Everett *et al.*, 1997). In infected cells, it colocalizes with the ubiquitin-specific protease and cyclin D3 in ND-10 structures and stabilizes the cyclin. Since the function of the bound cyclin D3 is unaffected, the purpose of stabilizing cyclin D3 is to bring the infected cell to the brink of the S phase, but certainly no further. The purpose of the interaction with EF-1δ is less well understood. ICP0 migrates from nuclei to cytoplasm as early as 3 hrs after infection. However, another viral protein, the U_L13 protein kinase hyperphosphorylates EF-1δ, presumably to maintain protein synthesis at a high level late in infection. ICP22, another α protein, also expresses several functions. The key function, however, appears to be efficient protein synthesis, especially of a subclass of late proteins. In the absence of ICP22, the virus grows very poorly in primary cell cultures but at near wild-type levels in continuous simian or human cell lines (Roizman and Sears, 1996).

The $\gamma_1$34.5 gene maps in the inverted repeats between the α0 and the terminal *a* sequence. The gene encodes a protein of 263 amino acids. This protein consists of a 160-residue amino terminal, 3 amino acids (AlaThrPro) repeated 10 times and a carboxyl-terminal domain of 73 amino acids (Roizman and Sears, 1996). The latter domain is homologous to the corresponding domain of GADD34, a conserved mammalian protein induced in cells undergoing differentiation, DNA damage repair, serum deprivation, etc. In-frame mutations or deletions in the amino-terminal domain result in gross attenuation of the virus in that it is unable to replicate in the CNS of experimental animal systems. Deletion of the carboxyl-terminal domain results in the total shutoff of protein synthesis as a consequence of the activation of PKR which in turn phosphorylates the α subunit of the translation initiation factor 2 (eIF-2α) (Roizman and Sears, 1996). Recent studies have shown that PKR is activated in cells infected with both wild-type and mutant viruses. However, in wild-type virus-infected cells the $\gamma_1$34.5 protein acts as accessory phosphatase 1α (PP1) factor. Specifically, it binds to PP1 through a short sequence common to all PP1 accessory factors and redirects PP1 to specifically dephosphorylate eIF-2α at the expense of other substrates of PP1 (6). The carboxyl-terminus of GADD34 can substitute for $\gamma_1$34.5 in precluding the shutoff of protein synthesis but the virus carrying the chimeric gene is avirulent (Roizman and Sears, 1996).

The activation of PKR in cells infected with wild-type virus is not surprising since both strands of the HSV genome are transcribed and transcription-termination signals are not very effective. Earlier studies have shown that >50% of the viral genome is represented in dsRNA formed by self-annealing of RNA extracted from cells late in infection (Roizman and Sears, 1996). Formation of dsRNA could activate PKR. HSV has evolved a novel strategy to block the shutoff of protein synthesis caused by dsRNA activated PKR by dephosphorylating eIF-2α rather than blocking its phosphorylation.

An important side issue is that for a number of genes second site mutations appear to compensate for the deleted gene sequences. Thus in the case of the $\gamma_1$34.5 gene Muhr and Gluzman (1996) reported the isolation of a second site compensatory mutant that restored protein synthesis. In this mutant, the α47 coding sequences were deleted juxtaposing the α promoter of this gene next to the coding sequences of the U_S11 gene. In essence, the mutation converted the gene from a γ_2 into an early gene. Subsequent studies have shown that U_S11 interacts with PKR and blocks phosphorylation of eIF-2α (Cassady and Roizman, 1998).

The examples cited above were deliberately focused on genes to be discussed later in this text. Nevertheless, the lessons apply to most of the genes whose functions have been studied in some detail. The salient features of what we have learned are that (i) most genes encode several functions. These functions are frequently quite diverse and not specifically related to each other, and (ii) while the cornerstone of the virus are the functions which replicate it, its edifice are the accessory genes which enable it to control the environment of the cell in exquisite detail. In many instances, ablation of a specific accessory function seems to have a minimal effect on yield, plaque size, or egress from the infected cells. Nevertheless, each marginal increment increases the ability of the virus to replicate and spread efficiently from cell to cell.

III. Latency

Latent infection

The profuse literature on HSV latency has contributed very little understanding of the events and mechanisms by which latency is established, maintained and terminated.

Central to the problem is the absence of an authentic *in vitro* model of latency. Yet many conclusions regarding latency are drawn from viral infection of cells *in vitro*. The experimental animal model systems are limited in number and diverse in properties. The virus establishes latency in mice, rabbits, guinea pigs and nonhuman primates. The virus reactivates spontaneously in rabbits and, in the case of HSV-2, also in guinea pigs. The most commonly used model is the mouse, but while its value is that it does not reactivate spontaneously with high frequency, it may not resemble latently infected humans in all respects. The available data on key issues may be summarized as follows (Roizman and Sears, 1996):

(i) The mechanisms by which the virus is transported by retrograde flow to the sensory neuron are not known. It may be presumed that on infecting nerve endings, a capsid-tegument structure is transported to the neuronal nuclear pore.

(ii) There is considerable evidence that the HSV genome is contained in an episomal form in latently infected neuron in 10 to 100 copies per genome of the cell harboring latent virus.

(iii) The only portion of the genome expressed during latency is antisense to the $\alpha 0$ and $\gamma_1 34.5$ genes and encompasses virtually the entire inverted repeat sequences flanking the U_L domain. Mappings to this domain are the 2.0 and 1.5 kb RNAs. These appear to be abundant, stable introns which accumulate in nuclei of some, but not all neurons, and do not encode proteins. The precursor or product of the splicing has not been identified. The larger, approximately 8.7 kb RNA is less abundant. The 2.0 and 1.5 kb latency-associated RNAs (LATs) are dispensable. Mutants lacking LATs establish latency although the absolute viral load present in the sensory ganglia may be reduced.

(iv) The mechanism by which HSV establishes latency in sensory neurons is unknown. The hypotheses proposed to date include (a) the presence in neuron of a repressor, as for example other members of the Oct protein family, which blocks transcription of α genes, (b) the absence of a cellular factor, as for example, C1, required for transcription of α genes, or (c) virally induced repressors of viral gene expression. None of these hypotheses has been tested critically or has yielded unambiguous results.

(v) The sequence of events which terminates latency and initiates viral replication is unknown. It has been proposed that stimuli such as damage to sensory nerve endings cause viral DNA synthesis and that reactivation results when the number of copies of viral DNA exceed a threshold level (Roizman and Sears, 1996).

(vi) The fate of the neuron in which latent virus became reactivated is unknown. The notion that on reactivation the virus begins to multiply precisely as in cells in culture and that the progeny is transported anterograde to the periphery is not uniformly accepted since by necessity it would imply that the neuron must succumb to infection. Resistance to the notion that the neuron dies stems from the observation that some individuals shed reactivated virus quasi continuously and yet suffer no anesthesia at the site of the recurrences. The underlying hypothesis, that the neuron survives notwithstanding all of the damage inflicted by viral gene products is untenable. Other explanations, that other neurons supply nerve endings to areas recovering from recurrent lesions and that the virus may survive for a long time in mucous membranes are more likely.

(vii) There is no consensus on the mechanism of the anterograde transport since synthesis of enveloped capsids and transport of virus would inevitably destroy the infected neuron. The hypothesis that a subviral particle is transported anterograde, from the nucleus of the sensory neuron to peripheral cell in contact with the nerve endings is bereft of evidence or precedent.

Viral functions in latently infected cells

The results derived from studies of productive infections in cells in culture are key sources of ontology of HSV and the platform on which predictions regarding the mechanisms of HSV latency can be attempted. The key lessons we have learned from studies of productive infections is that HSV has evolved a voracious appetite for controlling its environment and attempts to accomplish its objectives by more than one pathway. This conclusion is inconsistent with the prevailing view that HSV abdicated its control of cellular environment during latency.

The LATs described above have been viewed as functional end-products. The alternative view is that LATS are a stable product of transcripts expressed either transiently or at a low level and therefore the possibility exists that other ORFs located within the genome domain transcribed during latency are expressed. If this were the case, it would be expected that the function of the gene product would repress the expression of α genes and that the products would not be made or would be nonfunctional during productive infection. Analyses of the transcribed domain indicated the presence of 16 ORFs, designated A through P, and which contained more than 50 codons each. Of the five ORFs tested, two, ORF P and ORF O, were expressed and met the criteria expected of genes which could be expressed during latency (Roizman an Sears, 1996).

Both ORF P and ORF O and their products have been characterized in some detail. A high affinity ICP4 binding site is located at the transcription initiation site of ORF P. ORF P therefore was expressed only under the conditions

where ORF P was nonfunctional or the response element was ablated by mutagenesis. The ORF P gene maps quasi-totally antisense to the $\gamma_1$34.5 gene and expression of ORF P blocks that of the $\gamma_1$34.5 gene. The ORF P protein was found to bind a cellular protein known as p32. It colocalizes with spliceosomes and is pulled down with antibody to splicing factors. In cells infected with viruses carrying a derepressed ORF P, the ORF P protein is rapidly posttranslationally modified. However, with the time-frame of a few hrs after initial infection, the accumulation of ICP22 and of ICP0, the products of spliced mRNA, are grossly reduced whereas the accumulation of ICP4 and ICP27 are unaffected (Bruby and Roizman, 1996).

ORF O overlaps in part with ORF P but in a different reading frame. The first and only methionine codon of ORF O is located in the TATA box of ORF P (Randall *et al.*, 1997). Furthermore, extensive analyses of cells infected with a derepressed ORF P failed to detect an RNA which would correspond to ORF O. An additional problem was posed by the observation that since the ICP4 response element is located considerably downstream of the location of putative transcription initiation site of ORF O, the total repression of ORF O by ICP4 was inconsistent with the requirement that the response element be at or very near the transcription initiation site of the gene repressed by ICP4. In-frame insertional mutagenesis revealed that translation of ORF O initiates at the ORF P initiator methionine and that at a point between the first and 35th codon of ORF P translation switches to the ORF O reading frame. The mechanism of the frame-shift is unclear but within the interval between the initiator and 35th codon there are no obvious splice donor or acceptor sites. The ORF O protein was found to bind ICP4 and to preclude it from binding to its response elements on DNA.

ORF P and ORF O block the synthesis or function of three key α proteins, ICP0, ICP22 and ICP4. Viral replication cannot ensue in the absence of these proteins. Inasmuch as ORF P and ORF O are antisense to the $\gamma_1$34.5 gene, ablation of the coding domains of these genes would also result in the deletion of the $\gamma_1$34.5 gene. Since the latter is required for viral replication in neuronal cells, the experiment would be meaningless. Substitution of the initiator methionine of ORF P and ORF O, experimentally less convincing, does not block the establishment of latency. The conclusion remains therefore that while ORF P and ORF O may play a role in the establishment of latency, they are not *per se* sufficient.

A central impediment to the studies of HSV latency is the absence of an authentic *in vitro* model. A system in which the virus remains silent in cells in culture can be established, but only by applying drugs or conditions that are not required *in vivo*. While the mouse is the least expensive and carrier of sensory ganglia suitable for studies of latency, the results may well be colored by the inherent resistance of sensory neurons to viral replication. In essence, the molecular basis of latency remains a major objective of experimental herpesvirology.

Acknowledgements. These studies were aided by grants from the National Cancer Institute (CA47451 and CA71933), the United States Public Health Service.

References

Bruni R, Roizman B (1996): ORF P – a herpes simplex virus gene repressed during productive infection encodes a protein which binds a splicing factor and reduces synthesis of viral proteins made from spliced mRNA. *Proc. Natl. Acad. Sci. USA* **93**, 10, 423–10, 427,

Cassady KA, Gross M, Roizman B (1998): The second-site mutation in the herpes simplex virus recombinants lacking the $\gamma_1$34.5 genes preclude shutoff of protein synthesis by blocking the phosphorylation of eIF-2α. *J. Virol.* **72**, 7005–7011.

Cassady KA, Gross M, Roizman B (1998): The herpes simplex virus U_S11 protein effectively compensates for the $\gamma_1$34.5 gene if present before activation of protein kinase R by precluding its phosphorylation and that of the α subunit of translation initiation factor eIF-2. *J. Virol.* **72**, 8620–8626.

Everett RD, Meredith M, Orr A, Cross A, Kathoria M, Parkinson J (1997): A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. *EMBO J.* **16**, 566–577.

He B, Gross M, Roizman B (1997): The $\gamma_1$34.5 protein of herpes simplex virus 1 complexes with protein phosphatase 1α to dephosphorylate the a subunit of the eIF-2 translation initiation factor and preclude the shutoff of protein synthesis by double stranded RNA activated protein kinase (PKR). *Proc. Natl. Acad. Sci. USA* **94**, 843–848.

Kawaguchi Y, Bruni R, Roizman B (1997a): Interaction of herpes simplex virus 1 α regulatory protein ICP0 with elongation factor EF-1: ICP0 affects translational machinery. *J. Virol.* **71**, 1019–1024.

Kawaguchi Y, Van Sant C, Roizman B (1997b): Herpes simplex virus-1 α regulatory protein ICP0 interacts with and stabilizes the cell-cycle regulator cyclin D3. *J. Virol.* **71**, 7328–7336.

Randall G, Lagunoff M, Roizman B (1997): The product of ORF O located within the domain of herpes simplex virus 1 genome transcribed during latent infection binds to and inhibits in vitro binding of ICP4 to its cognate DNA site. *Proc. Natl. Acad. Sci. USA* **94**,10379–10384.

McGeoch DJ, Dalrymple MA, Davison AJ, Dolan A, Frame MC, McNab D, Perry, LJ, Scott JE, Taylor P (1988): The complete DNA sequence of the long unique region in the genome of herpes simplex virus type 1. *J. Gen. Virol.* **69**, 1531–1574.

Mohr I, Gluzman Y (1996): A herpesvirus genetic element which affects translation in the absence of viral GADD34 function. *EMBO J.* **15**, 4759–4766.

Roizman B, Sears AE (1996): The replication of Herpes simplex viruses. In Fields, BN, Knipe DM, Howley P, Chanock RM, Hirsch MS, Melnick JL, Monath TP, Roizman B (Eds): *Fields' Virology.* 3rd ed. Lippincott-Raven Press, New York, pp. 2231–2295.

Roizman B (1996): The function of herpes simplex virus genes: a primer for genetic engineering of novel vectors. *Proc. Natl. Acad. Sci. USA* **93**, 11307–11312.