Upstream Promoter Elements of the Herpes Simplex Virus Type 1 Glycoprotein H Gene

KEVIN R. STEFFY AND JERRY P. WEIR†*

Department of Microbiology, University of Tennessee, Knoxville, Tennessee 37996

Received 1 June 1990/Accepted 5 November 1990

To investigate the *cis*-acting sequence elements that are involved in the regulation of herpes simplex virus type 1 late-gene expression, recombinant viruses were constructed that express the *Escherichia coli lacZ* gene from the promoter of the glycoprotein H (gH) gene. Deletion experiments established an upstream boundary for the gH promoter of no more than 83 bp from the start of gH transcription and showed that the promoter sequences did not overlap with coding sequences of the upstream thymidine kinase (*tk*) gene. Sequences of the *tk* gene previously shown to be required for efficient processing of the *tk* transcript were essential for expression from the gH promoter and included a TATA-like element. In addition, the gH TATA element was specifically mutagenized to substitute the TATA elements of immediate-early, early, and other late viral promoters for the gH TATA element. The results indicated that the TATA element was an interchangeable component of herpes simplex virus type 1 promoters and did not regulate temporal expression.

Studies designed to define the DNA sequences involved in regulation and expression of herpes simplex virus type 1 (HSV-1) late genes have shown that sequences upstream from the TATA element of the glycoprotein C (gC) gene can be deleted without affecting expression (3). Other studies have shown that a 15-bp sequence containing the gC TATA element is absolutely required for expression (2) and that deletion of sequences containing the start of transcription and the 5' noncoding region of the gC gene eliminate expression from this late promoter (9). Furthermore, expression from hybrid promoters, consisting of sequences from both the early thymidine kinase gene (*tk*) and late UL42 promoters, indicated that regulatory domains of late promoters are downstream from the TATA element (5). In spite of these efforts, a definition of the sequence elements that distinguish late HSV-1 promoters from other types of viral promoters is lacking.

In an effort to define the promoter elements that are common among HSV-1 late promoters, we have constructed recombinant viruses that express β -galactosidase (β -gal) from several late promoters (10). The glycoprotein H (gH) promoter is of particular interest, not only because it is a late viral promoter, but also because of its proximity to the 3' end of the *tk* gene. In fact, only 90 bp separate the coding sequences of the *tk* gene from the start of gH transcription. It has been suggested that sequences that are part of the *tk* gene overlap with, or are even the same as, sequences that regulate gH expression (7). In particular, a repeated AATAAAA sequence in this region is homologous to sequences important for processing and polyadenylation at the 3' end of many eucaryotic genes and also similar to the TATA element that is present in the promoter region of many eucaryotic genes.

In previously reported experiments, it was found that both A+T-rich sequences and a G+T-rich sequence downstream from them were necessary for efficient processing and polyadenylation of the *tk* mRNA (1, 11). The aim of the present

study was to define the upstream promoter sequences of the gH gene and to determine the extent of their overlap with the *tk* gene. In addition, we have evaluated the specificity of the gH TATA element by substitution of this element with the TATA elements from other viral promoters.

Isolation of the gH promoter. To begin characterization of the *cis*-acting DNA sequences necessary for the expression of the gH gene, the gH promoter was isolated and linked to the *lacZ* gene encoding the procaryotic enzyme β -gal. Sequences between approximately -1400 and +198, relative to the reported start of gH transcription, were inserted into the insertion vector pGal8 and used to generate the recombinant herpesvirus vgHP.1 (Fig. 1) (10). Vero cells were infected with vgHP.1, and β -gal activity was measured in the presence and absence of phosphonoacetic acid (PAA). Expression of β -gal from HSV-1 promoters in recombinant viruses has been shown to faithfully reflect the temporal nature of the HSV-1 promoter used (8). In agreement with previous results (10), β -gal activity was initially detectable at 6 h postinfection and continued to increase for at least 24 h postinfection (data not shown). In the presence of PAA, β -gal activity was barely detectable (Fig. 2), indicating that the *lacZ* gene was expressed as an HSV-1 late gene when under the control of the gH promoter. Similarly, recombinant viruses were made that deleted upstream sequences to -800 (vgHP.2) and -83 (vgHP.3). Quantitation of β -gal activity from cells infected with each recombinant virus revealed that deletion of sequences upstream of -83 had no effect on expression from the gH promoter (Fig. 2). Thus, the gH promoter did not overlap with the coding sequences of the *tk* gene and had an upstream boundary no further than 83 bp from the start of transcription. The sequence of the gH promoter is shown in Fig. 1.

The site of transcription initiation from the gH promoter was analyzed by primer extension experiments. RNA was isolated from cells infected with the recombinant viruses, hybridized with a radioactive labeled primer that was complementary to either the 5' end of the gH mRNA or the β -gal mRNA, and extended with reverse transcriptase. With the gH primer, a single extension product resulted that migrated alongside an A residue (Fig. 3A), and this was designated +1 (Fig. 1), in agreement with previously reported results (7).

* Corresponding author.

† Present address: Department of Cellular Immunology, Walter Reed Army Institute of Research, 9620 Medical Center Dr., Rockville, MD 20850.

FIG. 1. Sequence of the gH promoter from the 3' end of the *tk* coding sequence to the start of gH translation. Both coding sequences are shown as open boxes. A 2.5-kb *EcoRI* DNA fragment of HSV-1(KOS) from map units 0.298 to 0.315 was isolated, cloned into the phage vector M13mp18, and designated 18-gHP. In vitro mutagenesis (4) of this recombinant phage, which contains approximately 1,600 bp upstream and 900 bp downstream from the start of gH transcription, changed the ATGGGG at the start of gH translation to a *SalI* site (GTCGAC). A 1.8-kb fragment was isolated by using the new *SalI* site and the upstream *EcoRI* site and cloned upstream of the β -gal gene in the HSV-1 insertion vector pGal8 (10). This plasmid was designated pgHP.1 and was used to construct the corresponding recombinant virus vgHP.1. The *EcoRI-SalI* promoter fragment was also cloned into M13mp18 to serve as a template for in vitro mutagenesis. Deletion of upstream sequences to -800 was accomplished by digestion of pgHP.1 with *SphI*, followed by religation. Deletion to -83 was accomplished by insertion of an *SphI* site (GCATGC) at positions -89 to -83 (underlined) by in vitro mutagenesis, followed by *SphI* digestion and religation. The *SphI-SalI* fragment (-83 to +198) was cloned into M13mp18 and served as the template for the other described mutations and deletions.

FIG. 2. Expression of β -gal from recombinant viruses. Vero cells were infected with recombinant viruses in the presence or absence of PAA, and β -gal activity was determined at 24 h postinfection as described before (6, 8) and was expressed relative to that of vgHP.3. Numbering of bases is relative to the start of transcription (arrow). The open box represents the downstream A+T-rich sequence; the upstream A+T-rich sequence is represented by a shaded box. The open triangle in gHP.8 represents a 4-base deletion from -76 to -73, and the solid triangle in gHP.10 represents the mutation of bases -40 to -35 to produce an *SphI* site.

FIG. 3. Analysis of RNA from recombinant viruses. RNA was isolated from infected Vero cells, hybridized to either a gH-specific primer (A) or a β -gal primer (B to E), and extended with reverse transcriptase (8). Each extension product (arrowheads) was run alongside a sequencing ladder prepared with the same primer and the appropriate plasmid DNA. The sequence shown on the left is the sequence around the start of gH transcription (* denotes +1). Individual lanes of the sequence reaction are labeled A, C, G, and T, and P is the primer extension lane. Primer extensions shown were obtained by using the gH primer and RNA from HSV-1(F)-infected cells (A), the β -gal primer and RNA from vgHP.2-infected cells (B), vgHP.4-infected cells (C), vgHP.5-infected cells (D), and vgHP-TK-infected cells (E).

Primer extension analysis of the mRNAs encoding β -gal from vgHP.1-, vgHP.2-, and vgHP.3-infected cells revealed that transcription initiated at the same site as for the authentic gH mRNA. The data from a primer extension experiment with the β -gal primer and the RNA from vgHP.2-infected cells is shown in Fig. 3B.

Sequence elements of the gH promoter. In an effort to further identify sequence elements important for expression of gH, specific deletions and mutations were made in the gH promoter sequences upstream from the start of transcription. In one construction, vgHP.8, the sequences from -76 to -73, which resemble the eucaryotic CCAAT consensus sequence, were deleted. When vgHP.8 was used to infect cells, β -gal activity was unaffected (Fig. 2).

The role of the two A+T-rich regions in the gH promoter was investigated by specific deletions. Two recombinant viruses, vgHP.4 and vgHP.5, were constructed which had either the upstream (-45 to -38) or the downstream (-32 to -26) A+T-rich region deleted. Whereas elimination of the upstream A+T-rich sequence had little effect on β -gal activity, elimination of the downstream A+T-rich sequence from -32 to -26 reduced β -gal activity by more than 50% (Fig. 2). Another recombinant virus, vgHP.6, in which both A+T-rich regions were deleted (-45 to -26), had almost no β -gal expression, indicating that an A+T-rich sequence was required for gH promoter activity. Analysis of RNA from infected cells revealed that the β -gal mRNAs produced by vgHP.4 and vgHP.5 initiated at the authentic gH start of transcription (Fig. 3C and D).

The recombinant virus vgHP.10 had the sequence from -40 to -35 changed from AAAGAC to an *Sph*I restriction site (GCATGC). This site-specific mutation had a negligible effect on β -gal expression. However, when this new restriction site was used to delete gH promoter sequences upstream from -35, an increase in β -gal synthesis was observed. It is not yet clear whether any sequences in this region are part of the gH promoter. It is unlikely that elimination of the upstream A+T-rich sequence causes this increase in expression, because this sequence could be deleted without effect (vgHP.4) and because the mutation in

			β -Gal Activity	
-40	-30	-20	no PAA	PAA
AAAGACAGAATAAAACGCACG			100	0.3
gHP3				
AAAGAGACTATATGAGCCACG			96	0.7
gHP-ICP4				
AAAGACGCATATTAAGGTACG			85	0.5
gHP-TK				
AAAGAGGGTATAAATTCCACG			131	2.2
gHP-gC				

FIG. 4. Substitution of TATA elements. Vero cells were infected with recombinant viruses in the presence or absence of PAA, and β -gal activity was determined at 24 h. The 13-base TATA element of the gH promoter and the substituted TATA elements from the ICP4, tk, and gC genes are shown in bold face letters. β -Gal activity is expressed relative to that of vgHP.3.

vgHP.10 had no apparent effect on β -gal expression. Taken together, the data indicated that although the upstream A+T-rich sequence was dispensable, a TATA-like sequence was required for expression from the gH promoter. However, these results also suggested that the location of this element was not fixed. Expression occurred from the gH promoter when only the upstream A+T-rich sequence was present, at least when shifted 6 bases downstream. In this construction, vgHP.5, the remaining A+T-rich sequence was located at -39 to -29, relative to the start of transcription, compared with positions -26 to -32 in vgHP.4. The fact that mRNA transcription was initiated at the same position indicated that there was some flexibility in the location of the TATA element and strongly implied that the start of transcription was determined either by other promoter elements or by the interaction of the TATA sequence with other promoter elements.

Specificity of the gH TATA element. The fact that a sequence could be a functional element of the tk polyadenylation signal as well as the gH promoter suggested that the composition of this element might not be critical for gH promoter expression. Three TATA elements derived from other HSV-1 promoters were substituted for the gH TATA element by in vitro mutagenesis. One recombinant, vgHP-ICP4, carried a 13-base sequence containing the TATA element from the promoter for the immediate-early ICP4 gene (Fig. 4). Expression of β -gal from cells infected with this recombinant virus was unchanged from the level of expression observed in vgHP.3-infected cells. Substitution of the tk promoter TATA element (vgHP-TK) decreased β -gal expression, whereas substitution of the gC TATA element (vgHP-gC) elevated β -gal expression. All three recombinant viruses still expressed β -gal as a viral late gene (Fig. 4). In addition, primer extension experiments revealed that each β -gal mRNA was initiated at the same position as the gH mRNA (Fig. 3E). These results indicated that TATA elements from different HSV-1 promoters were interchangeable and suggested that the TATA element influences the level of expression from a particular promoter rather than determining when that promoter will be active.

Recently, Homa et al. (2) demonstrated that a 15-bp sequence containing the gC TATA element is required for expression from the gC promoter. They also suggested that the gC TATA element was specific for late viral gene expression, since no expression was observed after substitution with a fragment containing the tk TATA element. That

experiment, however, substituted *tk* sequences from -37 to +52, which contained not only the *tk* TATA element but downstream sequences including the *tk* start of transcription, for gC promoter sequences from -144 to +124. In the experiment reported here, only a 13-bp fragment containing the gH TATA element was replaced by the analogous sequence from the *tk* promoter.

Another recent report described the construction of a hybrid promoter with the *tk* TATA element and upstream sequences fused to sequences downstream from the TATA element of the late viral UL42 promoter (5). The resulting promoter had properties of both early and late viral promoters. The accumulated data suggest that the TATA element is an interchangeable, general component of most, if not all, HSV-1 promoters and that it interacts with sequence elements that are further upstream in an early promoter and further downstream in a late promoter.

This work was supported in part by Public Health Service grant AI24471 from the National Institutes of Health and a Junior Faculty Research Award from the American Cancer Society. Kevin R. Steffy was supported by a training grant from the National Institutes of Health.

REFERENCES

1. Cole, C., and T. Stacy. 1985. Identification of sequences in the herpes simplex virus thymidine kinase gene required for efficient processing and polyadenylation. *Mol. Cell. Biol.* 5:2104-2113.
2. Homa, F., J. Glorioso, and M. Levine. 1988. A specific TATA box promoter element is required for expression of a herpes virus type 1 late gene. *Genes Dev.* 2:40-53.
3. Homa, F., T. Otal, J. Glorioso, and M. Levine. 1986. Transcriptional control signals of a herpes simplex virus type 1 late (γ) gene lies within bases -34 to +124 relative to the 5' terminus of the mRNA. *Mol. Cell. Biol.* 6:3652-3666.
4. Kunkel, T., D. Roberts, and R. Zakour. 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods Enzymol.* 54:367-382.
5. Mavromara-Nazos, P., and B. Roizman. 1989. Delineation of regulatory domains of early (β) and late (γ_2) genes by construction of chimeric genes expressed in herpes simplex virus 1 genomes. *Proc. Natl. Acad. Sci. USA* 86:4071-4075.
6. Miller, J. H. 1972. Experiments in molecular genetics, p. 352-355. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. Sharp, J., M. Wagner, and W. Summers. 1973. Transcription of herpes simplex virus genes in vivo: overlap of a late promoter with the 3' end of the early thymidine kinase gene. *J. Virol.* 45:10-17.
8. Weir, J. P., and P. R. Narayanan. 1988. The use of β -galactosidase as a marker gene to define the regulatory sequences of the herpes simplex virus type 1 glycoprotein C gene in recombinant herpesviruses. *Nucleic Acids Res.* 16:10267-10282.
9. Weir, J. P., and P. R. Narayanan. 1990. Expression of the herpes simplex virus type 1 glycoprotein C gene requires sequences in the 5' noncoding region of the gene. *J. Virol.* 64:445-449.
10. Weir, J. P., K. R. Steffy, and M. Sethna. 1990. An insertion vector for the analysis of gene expression during herpes simplex virus infection. *Gene* 89:271-274.
11. Zhang, F., R. Denome, and C. Cole. 1986. Fine-structure analysis of the processing and polyadenylation region of the herpes simplex virus type 1 thymidine kinase gene by using linker-scanning, internal deletion, and insertion mutations. *Mol. Cell. Biol.* 6:4611-4623.