Sequence and transcriptional analysis of the guinea-pig cytomegalovirus DNA polymerase gene

Mark R. Schleiss*

Division of Infectious Diseases, Children's Hospital Research Foundation, 3333 Burnet Avenue, Cincinnati, Ohio 45229, USA

Although the guinea-pig cytomegalovirus (GPCMV) displays a similar pathogenesis to human cytomegalovirus (HCMV), there have unfortunately been few molecular analyses of the GPCMV genome to date. The guinea-pig has proved useful for the testing of drugs active against CMV infection, and insights derived from characterization of the specific virally encoded molecular targets of antiviral therapies would allow this model system to be more fully developed. Because the DNA polymerase serves as an important target for nucleoside antiviral agents active against herpesviruses, experiments were undertaken to identify, clone and sequence the GPCMV DNA polymerase gene (pol). A 3285 bp ORF capable of encoding a 1094 amino acid protein was identified spanning portions of the HindIII Q and P fragments of the genome. This ORF contained extensive homology to other herpesvirus DNA pol genes. Northern blot analyses identified two 3’ coterminal pol-specific mRNAs of 3′9 and 1′9 kb at early times post-infection. Primer extension and nuclease protection analyses mapped the 5’ end of the 3.9 kb transcript to a site 275 bases upstream of the pol initiation codon. Comparison of the GPCMV pol-encoded sequence to those of other herpesvirus polymerases identified non-conservative amino acid substitutions in a domain involved in substrate recognition.
Short communication

Fig. 1. For legend see opposite.
the extent of clinical illness between ganciclovir and placebo groups in the setting of acute GPCMV infection, although the ganciclovir-treated animals had a lower concentration of virus in salivary gland cultures (Fong et al., 1987). The basis for this apparent resistance of GPCMV to ganciclovir is unknown. The recent characterization of the GPCMV glycoprotein B gene (Schleiss, 1994) has established for the first time the orientation of the GPCMV genome. Therefore, as a necessary prerequisite to understanding the molecular basis of the resistance of GPCMV to nucleoside analogues, the following studies were undertaken to identify and characterize the putative GPCMV DNA polymerase (pol) gene.

Since the genes encoding the DNA polymerase and the glycoprotein B collocate in the genomes of all herpesviruses (Kouzarides et al., 1987a), GPCMV sequences immediately downstream of the gB gene were determined, with the prediction that this region would contain the pol gene. As indicated in Fig. 1 (a), the gB/pol locus maps to the HindIII K, Q and P restriction fragments of the viral genome (Gao & Isom, 1984). Indicated restriction endonuclease sites were used to subclone these regions into pUC (United States Biochemical) and pGEM (Promega) vectors. A 4857 bp region extending from a ClaI site in the gB ORF to a BglII site in the HindIII P fragment was sequenced (Fig. 1b). Analysis of the nucleotide sequence revealed the presence of an ORF of 3285 nucleotides capable of encoding a protein of 1094 amino acids with extensive homology to the DNA polymerase proteins of both HCMV (Kouzarides et al., 1987b; Heilbron et al., 1987) and MCMV (Elliott et al., 1991). As with both the HCMV and MCMV pol genes, the promoter for the GPCMV pol lies within the coding sequences of the upstream gB gene. Translation of the pol ORF appears to initiate at the fifth ATG codon in the pol message. Short ORFs of 7 and 13 amino acids are observed in the 5' untranslated leader region of the pol mRNA; two other ATG codons which encode methionine residues in the glycoprotein B ORF are also present in the mRNA leader sequences upstream of the pol initiation codon. Similar short ORFs in the 5' leader sequences of HCMV genes play a role in regulating gene expression (Geballe & Mocarski, 1988). The GPCMV pol ORF contained extensive homology to other herpesvirus DNA pol genes. Comparison of the 1094 amino acid sequence of the GPCMV DNA polymerase protein with the polymerases of other cytomegaloviruses revealed identical amino acids at 559 positions in the MCMV DNA polymerase protein and 558 positions in the HCMV polymerase for an overall identity with each homologue of 51%.

Fig. 1. (a) Map location of the GPCMV pol gene. A detailed restriction map of the GPCMV glycoprotein B/pol gene locus is shown, spanning an approximately 9.5 kb region of the viral genome. The gB and pol mRNAs are transcribed from right-to-left and share a common 3' end. Probes used for mapping pol transcripts are indicated (A and B, arrows); 6.8 kb gB and 3.9 kb pol transcripts are indicated by solid lines; the truncated 1.9 kb pol transcript is indicated by a dashed line. pol and gB ORFs are indicated by stippled and cross-hatched bars, respectively. (b) Nucleotide sequence of a 4857 bp ClaI-BglII fragment containing the GPCMV pol gene. Numbers to the left indicate nucleotide sequence, numbers on the right, amino acid sequence. Restriction sites are indicated above the nucleotide sequence. The cap site marking the initiation of pol transcription is indicated by a bold asterisk (*). The GATATAA sequence and potential cis-regulatory elements 5' of the start site are indicated in bold type. The termination codon for the upstream gB ORF (TAA), the 14 bp palindromic sequence in the pol promoter, and polyadenylation signal sequences are underlined.
To examine transcription of the GPCMV pol gene, RNA was prepared from infected cells at various times post-inoculation and Northern analyses were performed. A 509 nucleotide BamHI–PstI fragment spanning bases 1005–1514 of the pol gene (Fig. 1b) was isolated and used as a template for the synthesis of a 489 nucleotide single-stranded probe complementary to the predicted pol mRNA sequence using the primer 5’ GAAGAGGTTGATGTTGTATC 3’ in a PCR reaction. Similarly, a 441 nucleotide HindIII–BamHI restriction fragment (spanning bases 2905–3346, Fig. 1b) was used as a template in a PCR reaction with a primer 5’ GATTCAGGTA-

GAGTTCGTCC 3’ to synthesize a 228 nucleotide [α-32P]dCTP-labelled single-stranded probe corresponding to a region further downstream in the pol ORF (Sturzl & Roth, 1990). Probes were then hybridized to total cellular RNA. Fig. 2(a) indicates that by 4 h following viral inoculation a transcript of 3.9 kb is identified by the upstream probe; the synthesis of this transcript is unaffected by the inclusion of phosphonoacetic acid (PAA) in the media, defining pol as an 'early' gene (Yin et al., 1990). The 3.9 kb pol transcript is similar in size to that reported for MCMV (Elliott et al., 1991), but is slightly smaller than the 4.7 kb HCMV pol mRNA.

Fig. 3. Nuclease protection mapping of the 3.9 kb GPCMV pol transcript. (a) Constructs used for pol transcript mapping. The arrow indicates the predominant 3.9 kb pol transcript. The pol ORF is indicated by a cross-hatched bar. (b) RNase protection analyses of the 5’ end of the 3.9 kb polymerase mRNA. [32P]CTP-labelled riboprobes antisense to the pol sequences spanning the XhoI and PstI sites (pKTS 183) and the Clal and PstI sites (pKTS 205) were synthesized using SP6 polymerase and hybridized to total whole-cell RNA obtained from cells harvested at 4 and 48 h post-infection. Following RNase digestion, protected fragments were sized on 6% acrylamide–8 M-urea gels. Lanes: M, molecular size markers (100 bp ladder); U, uninfected cell RNA; 4, 4 h infected cell RNA; 48, 48 h infected cell RNA; Pr, riboprobe. A predominant fragment of approximately 550 bases was protected from nuclease digestion by riboprobes generated from both constructs (arrow). (c) For 3’ end mapping, RNA from both 4 h and 48 h post-infection was hybridized to a 900 nucleotide riboprobe generated from pKTS 134, a clone containing two potential polyadenylation signals 3’ of the pol coding sequences. Lanes: M, molecular size markers (100 bp ladder); U, uninfected cell RNA; 4, 4 h infected cell RNA; 48, 48 h infected cell RNA; Pr, riboprobe. At both 4 and 48 h post-infection, a fragment of 320 bases is protected (arrow), corresponding to utilization of the upstream polyadenylation signal. With a longer gel exposure, a second protected fragment of 700 bases is noted using RNA from 48 h post-infection, but not with RNA from 4 h post-infection, suggesting that the downstream polyadenylation signal is only utilized at later time points post-infection, when gB transcription is active.
ments identified using RNA harvested at 4 h post-infection should be specific for pol transcription, since the overlapping gB gene is not transcriptionally active at this time (Fig. 2). A 556 nucleotide XhoI–PstI fragment spanning the pol promoter region and the first 92 codons of the pol ORF was cloned into pGEM (pKTS 183; Fig. 3a) and used to generate riboprobes which were used to map the 5' end of the pol message. A predominant protected fragment of approximately 550 bases was noted at both time points post-infection (Fig. 3b), suggesting that pol transcription initiated within a few nucleotides downstream of the XhoI site. Because of the proximity of the XhoI site to the start site of transcription, the nuclease protection assays were also performed with riboprobes generated from a clone containing a 1·1 kb ClaI–PstI fragment which included sequences from further upstream within the pol promoter (pKTS 205; Fig. 3a). A protected fragment of approximately 550 bases was again noted, identical in size to the fragment protected by the XhoI–PstI-derived riboprobe (Fig. 3b). These results indicated that the start site of pol transcription mapped to a region just downstream of the XhoI site, approximately 275 nucleotides upstream of the translational initiation codon. To identify the 3' end of the pol transcript, nuclease protection assays were performed with riboprobes synthesized complementary to a 929 nucleotide region spanning two potential polyadenylation signal sequences noted by DNA sequencing (pKTS 134, Fig. 3a). Analyses were performed using RNA obtained 4 h post-infection (when only pol transcription is active) as well as 48 h post-infection, when both pol and gB are actively transcribed. As indicated in Fig. 3(c), a fragment of approximately 320 bases is protected from RNase digestion at both time points post-infection, indicating that transcription terminates within 20 nucleotides downstream of the first polyadenylation signal. A less abundant protected fragment of 700 bases (arrow) was also noted using RNA obtained 48 h post-infection (at a time when gB transcription is active), but not with RNA obtained at 4 h post-infection, suggesting that at late times following infection there is either a low level of read-through of the upstream polyadenylation signal, or that there is differential utilization of polyadenylation signals as a function of kinetic class of transcript. Similar observations have been made in an HCMV gene in which differential mRNA 3' end processing and polyadenylation occurred at late times compared to early time points post-infection (Goins & Stinski, 1986).

To more precisely map the start site of pol transcription, primer extension analyses were performed. 32P end-labelled oligonucleotides with 5' termini 75 and 53 bases downstream of the translational initiation codon primed synthesis of cDNA products of approximately 350 and 330 nucleotides, respectively (Fig. 4a). More precise identification of the transcriptional start site was made using a third primer with a 5' terminus 151 bases upstream of the pol initiation codon. Primer extension analysis performed in conjunction with DNA sequence analysis identified a G residue 35 nucleotides downstream of the sequence GATATAA as the pol transcriptional initiation site (Fig. 4b). This sequence is similar to the TATA sequence noted in the MCMV pol promoter, GATAAAAA (Elliott et al., 1991). Inspection of the sequences surrounding the GATATAA element identified a number of potential cis-regulatory motifs. Two sequences with homology to the sequence GCCAT recognized by the CTF/NF-1 group of cellular DNA binding proteins are present at positions -272 and -101 relative to the transcriptional start site (Santoro et al., 1988). Just downstream of the putative TATA element is the core consensus sequence for the transcription factor Sp1, GGGCGG (Mitchell & Tjian, 1989). A consensus Sp1 binding site is also present in the MCMV pol promoter (Elliott et al., 1991). The sequence GTGACCTGCA, 39 nucleotides upstream of the TATA box, has homology to the core binding sites for the family of transcription factors designated ATF/CREB. ATF binding sites are present in a number of HCMV promoters (Spector et al., 1990). Mutagenesis studies have suggested that the ATF element in the HCMV pol promoter regulates transcription by conferring responsiveness to viral immediate early proteins (Stenberg, 1993). In another related betaherpesvirus, human herpesvirus 6 (HHV-6), an ATF-CREB site appears to be the major element regulating pol gene transcription (Agulnick et al., 1994). Intriguingly, the sequence GACACTCGAGTGTC between −9 and +5 of the...
GPCMV pol start site forms a perfect 14 bp palindrome, suggesting a role in regulation of transcription. An 8 bp inverted repeat in the HCMV pol promoter, IR1, is necessary for the binding of cellular proteins involved in upregulation of pol transcription (Kerry et al., 1994). Interestingly, the HCMV pol IR1 inverted repeat overlaps an imperfect 11 bp palindrome, AGGCtccc-GCCT, although the role that this palindromic sequence plays in transcription factor binding remains to be characterized.

Analysis of the protein coding sequences of a variety of eukaryotic DNA polymerases has identified several highly evolutionarily conserved domains which appear to be critical in substrate recognition (Hwang et al., 1992). Cytomegaloviral polymerases share conservation of these domains, which have been designated regions I–VII. An additional region, designated region A, is conserved only in viral DNA polymerases which are sensitive to antiviral drugs, in particular herpesvirus polymerases, but is not found in DNA polymerase α (Gibbs et al., 1988). Analysis of the amino acid sequence of GPCMV DNA polymerase region A revealed two nonconservative substitutions compared to the homologous DNA polymerase sequences of other cytomegaloviruses: an arginine residue at amino acid 527, and a serine residue at position 551. Nonconservative region A substitutions may contribute to reduced susceptibility to nucleoside analogues, since a substitution of a basic amino acid (lysine) for an acidic amino acid (glutamic acid) at position 597 in region A of the herpes simplex virus (HSV) DNA polymerase protein appears to be sufficient to confer acyclovir resistance upon an HSV mutant (Gibbs et al., 1988). Alternatively, differences in the phosphorylation of ganciclovir may be
the explanation for the relative resistance of GPCMV to nucleoside antivirals. Ganciclovir resistance in clinical isolates of HCMV is predominantly due to mutations in the virus UL97 gene, the gene responsible for phosphorylation of ganciclovir to its active triphosphate form (Lurain et al., 1994; Field & Biron, 1994). The GPCMV genome encodes a UL97 homologue (D. Fox & M. R. Schleiss, unpublished data), although it is not known whether this gene product can utilize ganciclovir as a substrate. Experiments are currently in progress to assess the extent of ganciclovir phosphorylation in GPCMV-infected cells. Comparison of evolutionarily conserved motifs in cytomegaloviral polymerases should facilitate an understanding of those amino acids which are critical in substrate recognition, and may provide insights into the mechanisms by which viruses develop resistance to antiviral agents. The use of recombinant technologies to generate GPCMV mutants sensitive to drugs active against HCMV would provide a uniquely useful tool for the preclinical testing of potential human antiviral therapies.

Thanks to Debbie Fox for technical assistance and to Larry Stanberry for helpful comments and for review of the manuscript. This work was supported by NIH grants HD 28827-01 and AI 15101, and CHRF grant 31-353-639.

References


(Received 31 January 1995; Accepted 27 March 1995)