African swine fever virus thymidylate kinase gene: sequence and transcriptional mapping

R. J. Yáñez, J. M. Rodríguez, J. F. Rodríguez, M. L. Salas* and E. Viñuela

Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Facultad de Ciencias, Universidad Autónoma, Cantoblanco, 28049 Madrid, Spain

A putative thymidylate kinase gene of African swine fever virus has been identified at the left end of the SalI I' fragment of the virus genome. The gene, designated A240L, has the potential to encode a protein of 240 amino acids with an \( M_c \) of 27,754 and is transcribed early after infection. Primer extension analysis indicates that transcription is initiated a short distance from the first ATG codon of open reading frame A240L. The deduced amino acid sequence of this open reading frame shows significant similarity with the human, yeast and vaccinia virus thymidylate kinases, the degree of identity being 23.7, 25 and 23.5\%, respectively. The putative African swine fever virus thymidylate kinase sequence is essentially collinear with the other thymidylate kinase sequences, but contains a carboxy-terminal extension of 37 amino acids rich in glutamic and aspartic acids. The A240L protein conserves the ATP-binding and nucleotide/nucleoside-binding domains characteristic of thymidylate kinases.

African swine fever (ASF) virus, the causative agent of a severe and economically important disease of swine, is a large enveloped DNA virus with an icosahedral morphology (Viñuela, 1985, 1987). Its genome is formed by a double-stranded DNA molecule of about 170 kbp with hairpin loops and terminal inverted repeats (González et al., 1986; Sogo et al., 1984), similar to those of poxvirus DNA (Baroudy et al., 1982; Wittek & Moss, 1980). Like the poxviruses (Moss, 1990), ASF virus packages a complete enzymatic system for the synthesis of early RNA (Kuznar et al., 1980; Salas et al., 1981, 1986).

Although ASF virus was thought to multiply exclusively in the cytoplasm of infected cells, recent evidence indicates the occurrence of an initial nuclear stage of DNA replication, followed by the cytoplasmic stage (García-Beato et al., 1992a). For the cytoplasmic replication at least, the virus should encode enzymes that catalyse DNA synthesis and, indeed, genes for DNA topoisomerase II (García-Beato et al., 1992b), DNA ligase (Hammond et al., 1992; Yáñez & Viñuela, 1993) and DNA polymerase (J. M. Rodríguez, R. J. Yáñez, J. F. Rodríguez, M. L. Salas & E. Viñuela, unpublished) have been found in the virus genome. In addition, new DNA ligase (Hammond et al., 1992; Yáñez & Viñuela, 1993) and DNA polymerase (Polatnick & Hess, 1972; Moreno et al., 1978) activities are induced in ASF virus-infected cells. Like other large DNA viruses, ASF virus encodes enzymes involved in nucleotide metabolism, such as thymidine kinase (Blasco et al., 1990; Martín Hernández & Tabarés, 1991) and ribonucleotide reductase (Boursnell et al., 1991). Another enzyme required for the synthesis of DNA precursors is thymidylate kinase, catalysing the phosphorylation of dTMP to dTDP in both the de novo and salvage pathways of dTTP biosynthesis. The yeast thymidylate kinase is also required for DNA replication, even in permeabilized cells or in vitro systems where the dNTP precursors are provided (Hereford & Hartwell, 1971; Kuo & Campbell, 1982; Celniker & Campbell, 1982; Arendes et al., 1983), suggesting some kind of interaction of this enzyme with the DNA replication complex, as has been discussed by Jong et al. (1984) and Sclafani & Fangman (1984).

In herpesviruses, a single virus-encoded protein possesses both thymidine kinase and thymidylate kinase activities (Chen & Prusoff, 1978; Chen et al., 1979), whereas in vaccinia virus two different genes encode these enzymes (Hruby et al., 1983; Weir & Moss, 1983; Smith et al., 1989; Hughes et al., 1991). Here we show that ASF virus has, like vaccinia virus, a separate gene for thymidylate kinase. The nucleotide sequence and transcriptional mapping of this gene, as well as a comparison of its deduced amino acid sequence with other known thymidylate kinase sequences are presented.

The SalI I' (SI') restriction fragment of the Vero-adapted BA71V strain of ASF virus is located at about
Short communication

Fig. 1. Sequence of the ASF virus thymidylate kinase gene. (a) EcoRI map of ASF virus BA71V DNA and localization of the SI' restriction fragment and of the thymidylate kinase gene (A240L) within this fragment. (b) Partial DNA sequence of SI' fragment from nucleotides 273 to 1072, and the deduced amino acid sequence of A240L protein. The DNA sequence is shown in the 5' to 3' direction and from right to left according to the restriction map. An oligonucleotide complementary to the underlined sequence was used as a hybridization probe or as a primer in the Northern blot and primer extension analyses, respectively. The transcription start sites are indicated by arrows.

30-6 kbp from the left end of the virus genome (Fig. 1). Random sequencing (Bankier & Barrell, 1984) of this fragment cloned in pBR322 (Ley et al., 1984) was performed essentially as described by Rodriguez et al. (1992a), using the dideoxynucleotide chain termination method (Sanger et al., 1980) and T7 DNA polymerase. The nucleotide sequences were compiled using the programs of Staden (1987). An examination of the nucleotide sequence revealed an open reading frame (ORF) reading in the leftward direction and localized close to the left end of SI' (Fig. 1). This ORF, designated A240L (Rodriguez et al., 1992a), encodes a polypeptide of 240 amino acids with a predicted relative molecular mass of 27754. Searches in the National Biomedical Research Foundation, SWISSPROT, EMBL and GenBank databases with the programs FASTA and TFASTA (Pearson & Lipman, 1988) and using the amino acid sequence of ORF 240L as a query showed significant similarity between the protein and human yeast and vaccinia virus thymidylate kinases. The optimized FASTA scores were 168, 166 and 211, respectively. No similarity was found with the herpesvirus thymidine kinase/thymidylate kinase.

To determine whether the putative thymidylate kinase gene of ASF virus was expressed during the viral infection, early and late poly(A)+ RNA isolated from Vero cells infected with ASF virus, as well as the early-like poly(A)+ RNA synthesized in vitro by the virion-associated RNA polymerase (Salas et al., 1986), were analysed by Northern blot hybridization, using the oligonucleotide specific for ORF A240L as the probe (Fig. 1). To obtain the poly(A)+ RNA from cultured cells, Vero cells were mock-infected or infected with ASF virus (BA71V strain) at a multiplicity of 10 p.f.u./cell. Early RNA was obtained from cells infected for 4 h in the presence of cycloheximide (200 µg/ml) or for 18 h in the presence of cytosine arabinoside (40 µg/ml). Late RNA was isolated from cells infected for 18 h in the absence of inhibitors. Poly(A)+ RNA was prepared as described (Salas et al., 1986). Then, 4 µg of the poly(A)+ RNAs from cultured cells and 1 µg of the in vitro poly(A)+ RNA were fractionated on a formaldehyde–agarose gel, trans-
ferred to nitrocellulose and hybridized with the $^{32}$P-labelled probe as previously described (Rodriguez et al., 1992a). As shown in Fig. 2, ORF A240L was transcribed preferentially early in infection. A single RNA band of a smaller size than the 18S RNA marker was identified. The early expression of this gene is consistent with the requirement for thymidylate kinase in the biosynthesis of DNA precursors.

The 5' end of the A240L RNA was mapped by primer extension analysis. For this, early (cycloheximide and cytosine arabinoside) and late RNA from infected cells or RNA from mock-infected cells was prepared as previously described (Rodriguez et al., 1992a). An oligonucleotide primer complementary to the underlined sequence in Fig. 1 was 5' end-labelled with $^{32}$P, annealed to total RNA and extended with avian myeloblastosis virus reverse transcriptase as previously described (Sambrook et al., 1989). The primer extension products were then electrophoresed in a 6% polyacrylamide sequencing gel. In agreement with the Northern blot analysis, the primer hybridized with early RNA but not with late RNA (Fig. 3). After extension with reverse transcriptase, two major bands were detected. The sizes of these bands (73 and 72 nucleotides) correspond to initiations of transcription at positions -7 and -6 relative to the first nucleotide of the translation initiation codon, respectively, as indicated by arrows in Fig. 1.

The transcriptional start sites for a number of early genes of ASF virus have been identified (Almazán et al., 1992). In all cases, transcription is initiated a short distance from the first ATG codon of the corresponding ORF. An examination of the sequences immediately upstream of the RNA initiation sites in the six genes analysed so far has revealed no obvious common sequence that could act as a transcriptional regulatory signal.

Recently, a stretch of seven or more thymidylate residues has been identified as the transcription termination site of ASF virus early and late RNAs (Almazán et al., 1992, 1993). A run of nine thymidylate residues is found 295 nucleotides downstream of the stop codon of ORF A240L. Termination at this site would produce an RNA of about 1025 nucleotides, similar in size to the transcript detected using the Northern blot.

As indicated before, the A240L amino acid sequence was found to be similar to the thymidylate kinase sequences described so far. A multiple alignment of ORF A240L and the human (Su & Sclafani, 1991), yeast (Jong et al., 1984) and vaccinia virus (Smith et al., 1989) thymidylate kinases is shown in Fig. 4. It can be seen that the A240L sequence is essentially collinear with the thymidylate kinase sequences. The ASF virus protein, however, contains a carboxy-terminal extension of 37 amino acids rich in glutamic and aspartic acids, not found in the other proteins. The similarity between the sequences in the multiple alignment was greater in the amino-terminal two-thirds of the proteins, where the putative ATP-binding and nucleotide/nucleoside-
binding domains are located [Smith et al. (1989) and references therein]. In the ASF virus protein, the sequence GINGVGKS, at positions 13 to 20 in the multiple alignment, conforms exactly with the consensus A motif found in ATP-binding proteins (GX2GXGKS/T) (Gorbalenya & Koonin, 1989), whereas in the other thymidylate kinase sequences the second glycine residue is replaced by an arginine or a lysine. The nucleotide/nucleoside-binding site between residues 97 and 109 is also highly conserved in the A240L protein. Additional regions containing identical amino acids in the four sequences compared might also be critical for enzyme function. In relation to this, the invariant glutamate residue at position 79 in the multiple alignment has been shown to be important for the activity of the yeast enzyme (Su & Sclafani, 1991).

Pairwise comparisons of the ASF virus thymidylate kinase with the cellular and vaccinia virus enzymes show a degree of identity ranging from 23.5 to 25%. In contrast, the thymidylate kinase from vaccinia virus is 38.2 and 41.2% identical to the yeast and human enzymes, respectively, and the yeast and human thymidylate kinases are 43% identical. Therefore, the cellular enzymes are more closely related to vaccinia virus thymidylate kinase than to ASF virus thymidylate kinase. A previous study has also indicated that the ASF virus thymidine kinase is relatively unrelated to poxvirus and cellular thymidine kinases (Blasco et al., 1990).

As discussed before (Blasco et al., 1990), a possible explanation for these findings would be that ASF virus acquired the thymidine kinase and thymidylate kinase genes long before the poxviruses. In keeping with this possibility, the results of a recent phylogenetic analysis carried out with topoisomerase II sequences have suggested that the ASF virus topoisomerase II gene could have been derived from an ancestral cellular gene before the divergence of protozoa, yeast and metazoa (García-Beato et al., 1992b).

The ASF virus thymidine kinase gene has been shown to be non-essential for virus replication in cultured cells (R. J. Yáñez, unpublished results), and this gene has been used as a site for the insertion of foreign genes into the virus genome (Rodriguez et al., 1992b). Despite its key role in dTTP biosynthesis, it seems likely that the
thymidylate kinase gene will also be non-essential, since the homologous gene of vaccinia virus can be inactivated without impairing the ability of the virus to grow in cultured cells (Hughes et al., 1991). In that case, the thymidylate kinase locus of the ASF virus genome could be used as an alternative or additional site for the insertion of foreign DNA.

We thank M. L. Nogal, L. Yuste and C. Enriquez for skilful technical assistance. This work was supported by grants from the Comisión Interministerial de Ciencia y Tecnología, the Comunidad Autónoma de Madrid, the Consejería de Agricultura de la Junta de Extremadura, and by an Institutional grant from Fundación Ramón Areces. R. J. Yáñez was a fellow from Ministerio de Educación y Ciencia.

References


(Received 25 January 1993; Accepted 9 March 1993)