An African swine fever virus gene with similarity to bacterial DNA binding proteins, bacterial integration host factors, and the *Bacillus* phage SPO1 transcription factor, TF1

J.G.Neilan, Z.Lu, G.F.Kutish, M.D.Sussman, P.C.Roberts, T.Yozawa and D.L.Rock*
Plum Island Animal Disease Center, Agricultural Research Service, USDA, PO Box 848, Greenport, NY 11944, USA

Received January 25, 1993; Accepted February 12, 1993 GenBank accession no. L10066

African swine fever virus (ASFV), the causative agent of African swine fever in domestic pigs, is a large icosahedral virus with a double-stranded DNA genome of 170–190 kb that shares many characteristics with poxviruses. Although ASFV and poxviruses are distinct morphologically, they both replicate in the cell cytoplasm, exhibit temporal regulation of gene expression, and have similar genome structures, which include terminal inverted repeats, terminal crosslinks, a central conserved region and variable regions at each end of the genome (1). ASFV is the sole member of an unnamed family of animal viruses.

LMW5-AR is an open reading frame (ORF) of 104 amino acids (9.5 kD), located 0.22 m.u. from the left-hand end of the ASFV genome within the *Bam*HI c restriction fragment of the Malawi Lil 20/1 isolate of ASFV (2). Fasta (3) searching of the Swissprot database revealed that LMW5-AR shared significant similarity to the family of histone-like proteins (4, 5), which include the prokaryotic histone-like DNA binding proteins (HU) (6), integration host factors (IHF) (7) and the *Bacillus* phage SPO1 transcription factor, TF1 (8). LMW5-AR has the highest level of similarity to the HU proteins encoded by the *Enterobacteriaceae* (28% amino acid identity/89 residues; 45% conservation, Fasta = 99; z = 9.82) with lesser similarity to the *Bacillus* phage SPO1 TF1 gene (25% amino acid identity/96 residues; 44% conservation, Fasta = 94; z = 8.58) and the IHF proteins (29% amino acid identity/72 residues; 47% conservation, Fasta = 75; z = 7.11). A search of the Prosite data base (release 10) indicates that the LMW5-AR amino acid sequence from Gly 55 to Thr 76 matches the signature sequence pattern for the family of histone-like proteins except for the two additional amino acids after Gly 55 (4, 5) (Figure 1). Also, the residues Phe 58 and Arg 69, which are important for DNA binding by HU proteins (9), are conserved in LMW5-AR.

LMW5-AR is transcribed in ASFV infected macrophage cell cultures at late time points in the infection cycle (data not shown). Northern blot analysis reveals multiple transcripts in this region. LMW5-AR transcripts are not detected when DNA replication and late gene expression are inhibited by cytosine arabinoside, thus indicating LMW5-AR is a late viral gene.

The function of LMW5-AR in ASFV replication is unknown. Because the gene is expressed late, coincident with viral DNA replication and progeny assembly, it may be a DNA binding protein involved in nucleoid formation of newly replicated viral genomes. Alternatively, it may function as a transcription factor involved in regulation of viral gene expression.

REFERENCES

1. Costa,J.V. (1990) In: Darai,G. (ed.), *Molecular Biology of Iridoviruses*. Kluwer Academic Publishers, pp. 247–270.
2. Dixon,L.K. (1988) *J. Gen. Virol.* **69**, 1683–1694.
3. Pearson,W.R. (1990) *Methods Enzymol.* **183**, 63–98.
4. Friedman,D.I. (1988) *Cell* **55**, 545–554.
5. Drlica,K. and Rouviere-Yaniv,J. (1 987) *Microbiol. Rev.* **51**, 301–319.
6. Kano,Y., Yoshino,S., Wada,M., Yokoyama,K., Nobuhara,M. and Iamamoto,F. (1985) *Mol. Gen. Genet.* **201**, 360–362.
7. Miller,H.I. (1984) *Cold Spring Harbor Symp. Quant. Biol.* **49**, 691–698.
8. Greene,J.R., Brennan,S.M., Andrew,D.J., Thompson,C.C., Richards,S.H., Heinrikson,R.L. and Geiduschek,E.P. (1984) *Proc. Nat. Acad. Sci. USA* **81**, 7031–7035.
9. Goshima,N., Kohno,K., Imamoto,F. and Kano,Y. (1990) *Gene* **96**, 141–145.

```
E. coli HU      1 MnKtQLidVIAekaELsKtqakaaLeStlaaItEsLKEgDaVqL... 44
E. coli IHF     3 LTKaEMseyLfdklGLsKrdakeLVelffeeIrrALeNgEQVkL... 45
BPSPO1 TF1      1 MnKtELikaIAqDTELtqvsVsKMLaSfEKIttEtVakgDkVqL... 44
ASFV LMW5-AR    9 ITKqELysLVAaDTQLnKalIeRIFtSqQKIIqNALKHnQEViIppq 55

E. coli HU     45 VgFgtFkVnhRAeRtGRNPqTGkeIkIaAa...nvpaFvsgKALkDaVk  90
E. coli IHF    46 sgFgnFdLRdKnqRpGRNPkTGEdIpItAR...RvVtFRPgqkLksrVE  92
BPSPO1 TF1     45 tgFlnIkpvARqARkGfNPqTqEaLEIaPs...vgVsVKPgesLkkaaE  90
ASFV LMW5-AR   56 IkFtvVtVKAKPARqGHNPaTGEpIQIkAKpehKaVkIRAlKPVhDmLN 104
```

Figure 1. Conservation of the amino acid sequence of LMW5-AR, *E.coli* histone-like DNA binding protein and integration host factor, and the *Bacillus* phage SPO1 transcription factor, TF1. Residues identical to LMW5-AR are shown in upper case bold letters, while conservative substitutions are depicted by upper case letters only. Residues matching the signature sequence for the family of histone-like DNA binding proteins are underlined. The numbering of LMW5-AR is as submitted to the database. The other proteins are numbered as documented in the Swiss Protein database. Accession numbers for *E.coli* HU, *E.coli* IHF and the SPO1 TF1 are P02342, P06984 and P04445 respectively.

* To whom correspondence should be addressed