Specific transcriptional and post-transcriptional regulation of the major immediate early ICP4 gene of GaHV-2 during the lytic, latent and reactivation phases

Perrrine Rasschaert, Isabelle Gennart, Imane Boumart, Ginette Dambrine, Benoit Muyldens, Denis Rasschaert and Sylvie Laurent

Abstract

Transcriptional and post-transcriptional mechanisms are involved in the switch between the lytic, latent and reactivation phases of the viral cycle in herpesviruses. During the productive phases, herpesvirus gene expression is characterized by a temporally regulated cascade of immediate early (IE), early (E) and late (L) genes. In alphaherpesviruses, the major product of the IE ICP4 gene is a transcriptional regulator that initiates the cascade of gene expression that is essential for viral replication. In this study, we redefine the infected cell protein 4 (ICP4) gene of the oncogenic Marek’s disease virus (MDV or gallid herpesvirus 2) as a 9438 nt gene ended with four alternative poly(A) signals and controlled by two alternative promoters containing essentially ubiquitous functional response elements (GC, TATA and CCAAT boxes). The distal promoter is associated with ICP4 gene expression during the lytic and the latent phases, whereas the proximal promoter is associated with the expression of this gene during the reactivation phase. Both promoters are regulated by DNA methylation during the viral cycle and are hypermethylated during latency. Transcript analyses showed ICP4 to consist of three exons and two introns, the alternative splicing of which is associated with five predicted nested ICP4ORFs. We show that the ICP4 gene is highly and specifically regulated by transcriptional and post-transcriptional mechanisms during the three phases of the GaHV-2 viral cycle, with a clear difference in expression between the lytic phase and reactivation from latency in our model.

INTRODUCTION

Marek’s disease virus 1 (MDV-1) or gallid herpesvirus 2 (GaHV-2) is an oncogenic alphaherpesvirus that causes fatal T-cell lymphoma in chickens. The viral cycle of GaHV-2, like those of all herpesviruses, is characterized by three phases: a lytic phase in macrophages, B-lymphocytes and epithelial cells, particularly those of the feather follicles; a latent phase in T lymphocytes; and a phase of reactivation from latently infected lymphocytes mimicking the lytic phase.

The replication of GaHV-2 viral particles during the lytic phase starts with the expression of the immediate early (IE) genes, which encode transcription regulators required for the expression of the early (E) and late (L) genes [1]. During the HSV-1 lytic cycle the expression of the IE genes of this temporally regulated cascade is enhanced by the fixation of the tegument protein VP16-induced complex (VIC) at IE promoters and then repressed by negative feedback from IE genes, partly supported by repression of transcription from IE promoters by ICP4, the major alphaherpesvirus IE gene [2–4]. During GaHV-2 latency associated with lymphomagenesis, all these genes are repressed, whereas a few other key genes are expressed, including the major viral oncogene meq, the UL47 major tegument protein and some viral genes.

Received 13 September 2017; Accepted 2 January 2018

Author affiliations: 1 Equipe Transcription et Lymphome Viro-Induit (TLVI), UMR 7261 CNRS, Université François Rabelais de Tours, UFR Sciences et Techniques, Parc de Grandmont, 37200 Tours, France; 2 Veterinary Integrated Research Unit, Faculty of Sciences, Namur Research Institute for Life Sciences (INARILIS), University of Namur (UNamur), 5000 Namur, Belgium; 3 INRA, Département de Santé Animale, Centre de Recherches de Tours, 37380 Nouzilly, France.

*Correspondence: Denis Rasschaert, denis.rasschaert@univ-tours.fr

Keywords: ICP4; herpesvirus; GaHV-2; transcriptional regulation; splicing; epigenetic.

Abbreviations: APA, alternative polyadenylation; BGSAs, bisulfite genomic sequencing analyses; BoHV-1, bovine herpesvirus 1; CEF, chicken embryo fibroblasts; CMV, cytomegalovirus; CRE, c-amp response element; E, early; EHV-1, equine herpesvirus 1; FHV-1, feline herpesvirus 1; FF, feather follicle; GaHV-2, gallid herpesvirus 2; HHV-8, human herpesvirus 8; HSV-1, herpes simplex virus 1; ICP27, infected cell protein 27; ICP4, infected cell protein 4; IE, immediate early; Inr, initiator element; IRs, internal repeat short; L, late; LAT, latency-associated transcript; IncRNA, long non-coding RNA; MCMV, murine cytomegalovirus; MDV-1, Marek’s disease virus 1; miRE, microRNA responsive element; miRNA, microRNA; PBL, peripheral blood lymphocytes; RE, response element; SS, splice site; TRs, terminal repeat short; TSS, transcription start sites; UL47, unique long 47; VIC, VP16-induced complex; VP16, viral protein 16; vTR, viral telomerase RNA.

Two supplementary tables and four supplementary figures are available with the online version of this article.
encoding non-coding RNAs, including microRNAs, the vTR (viral telomerase RNA) and the long non-coding RNA latency-associated transcript (LAT), which is complementary to ICP4 in the two short internal and terminal repeat regions (IR3 and TR3) [5–10]. During the GaHV-2 viral cycle, these antisense ICP4 and LAT genes display opposite patterns of expression that are typical of the lytic and latent phases, respectively [11]. The switch between the productive phases and latency operates through epigenetic, transcriptional and post-transcriptional modifications of both cellular and viral genes [12]. As the products of the ICP4 genes of alphaherpesviruses are required for the activation of IE, E and L genes, the expression of ICP4 genes must be fine-tuned at the transcriptional and post-transcriptional levels during the viral cycle. The marked differences in the regulation of expression between the ICP4 genes of different alphaherpesviruses have made it impossible to identify a common mechanism. Moreover, at the molecular level, epigenetic modifications to ICP4 genes and the modulation of their expression by RNA interference are common to all alphaherpesviruses, whereas the use of single or dual promoters and alternative splicing mechanisms seems to be virus-specific [2, 12, 13]. The proteins of the alphaherpesvirus ICP4 superfamily are phosphoproteins with an overall similarity of 55% and five conserved domains, on the basis of which a short ORF (4245 nt) was predicted for the GaHV-2 ICP4 [14]. However, sequencing of the whole GaHV-2 genome led to the prediction of a larger ORF of 6966 nt (ORF6966) and of a potential poly(A) signal downstream from this ORF [14, 15]. This prediction of a larger ORF6966 is supported by (i) the observed production of larger ICP4 mRNAs, up to 10 kb in length; (ii) the detection of the corresponding ICP4 proteins; and (iii) the identification of a dual enhancer region upstream from the ORF6966 [1, 14, 16, 17]. Two further studies showed that this large enhancer region was bounded by the transactivator MEQ/c-JUN heterodimers and displayed epigenetic alterations during latency, as a result of DNA and histone methylation [18, 19]. We have also shown that the LAT gene of GaHV-2 encodes a cluster of microRNAs, one of which, mdv1-miR-M7-5P, is strongly expressed during latency and is involved in RNA interference mechanisms contributing to the establishment and maintenance of latency through the targeting of both ICP27 and ICP4 IE mRNAs [9]. Finally, in addition to its conserved function in transcriptional regulation, the GaHV-2 ICP4 gene has also been implicated in the maintenance of transformation during latency [16, 20, 21], but its expression alone was not sufficient to induce viral reactivation in latently infected cells [22].

Following on from our former studies on the transcriptional regulation of GaHV-2 coding and non-coding genes (ICP27, vTR, LAT, ERL and clusters of miRNAs) throughout the viral cycle, we aimed to characterize the essential major transactivator ICP4 gene and its expression and regulation during the lytic, latent and reactivation phases of the GaHV-2 viral cycle. We showed, by RACE-PCR and transcript analyses, that the GaHV-2 ICP4 gene was 9438 nt long and consisted of three exons and two introns, ending in four poly(A) signals, and that it was controlled by two alternative promoters. Luciferase assays identified a distal and a proximal promoter and their functional regulatory elements. The characterization of alternative spliced transcripts suggested that there were five ICP4 ORFs. We observed differential transcriptional and post-transcriptional regulation of ICP4 gene expression in the lytic, latent and reactivation phases, involving the alternative use of distal and proximal promoters and DNA methylation.

RESULTS

The ICP4 transcript ends in four alternative poly(A) signals

We identified the 3′ end of the MDV ICP4 mRNA by performing 3′RACE-PCR on total RNA extracted from chicken embryo fibroblasts (CEF) infected with the GaHV-2 RB-1B virus strain, native MSB-1 lymphocytes corresponding to a latently infected cell line and MSB-1 cells treated with n-butyrate to induce the reactivation of the latent virus (MSB-1R). The 3′RACE-PCRs and nested PCRs were carried out with the 3′ and 3′nested reverse primers supplied in the GeneRACER kit (Invitrogen), and with the A231 and A477 ICP4 specific forward primers, which bind 240 and 39 bp upstream from the previously predicted poly(A) signal [14], respectively. An analysis of the sequences of 96 randomly selected clones for each cell type revealed the use of four different poly(A) signals defining four 3′UTRs with lengths of 36, 302, 745 and 835 nt, referred to as the small (S), medium (M), large (L) and extra-large (XL) 3′UTRs, respectively (Fig. 1). The sequence of the poly(A) signals of the 3′UTR-S, -L and -XL corresponded to the major consensus sequence, AAUAAA, whereas the sequence of the poly(A) signal of the 3′UTR-M corresponded to the minor consensus sequence, AUUAAA (Figs 1 and S1, available in the online online version of this article).

The principal 3′UTR, 3′UTR-XL, was found in about half (48–57%) of the ICP4 mRNAs (Fig. 1). The initially predicted poly(A) signal 3′UTR-S (18–22%) and the 3′UTR-L (18–28%) displayed a similar distribution in each cell type (Fig. 1). Finally, only the 3′UTR-M was associated with a differential use of poly(A) signals between the three biological samples. It was absent from latently infected MSB-1 cells, and accounted for a small proportion (6–8%) of the ICP4 mRNAs extracted from CEF RB-1B and MSB-1R cells (Fig. 1). Collectively, our data indicate that, except for 3′UTR-M, which forms a very small minority of 3′UTR sites and is specific to the active phases of viral replication, the 3′UTRs have similar distributions regardless of the phase of the viral cycle.
**ICP4 transcription is initiated by two alternative promoters**

We carried out 5'RACE-PCR on the RNAs initially extracted for 3'RACE-PCR analysis. We performed 5' and 5' nested RACE-PCRs with the 5' and 5' nested GeneRacer forward primers (Invitrogen) and with the ICP4-specific A876 and A654 reverse primers, which bind 175 and 131 nt, respectively, downstream from the predicted start codon of ICP4 (AUG) (Fig. 2a). Sequencing analysis on 48 5' nested RACE-PCRs performed on RNA extracted from CEF RB-1B, MSB-1 and MSB-1R cells, corresponding to the lytic, latency and reactivation phases of the viral cycle. The primers used for 3'RACE PCR are indicated by arrows. The annotated sequences of the ICP4 3'UTRs are provided in Fig. S1.

Splicing from TSSp (Tp) (Fig. 2a). The Td were alternatively spliced from a consensus constitutive 5' site (SS; GU, position 167020) to two consensus alternative 3' splice sites (SS; CAG, positions 168223 and 168424) (Figs 2a and S2). The first 3' SS, upstream from AUG1 (position 168399), was associated with splicing of intron 1 (1204 nt) in Td (intron 1 spliced distal transcripts) (Fig. 2a) and the second 3' SS, downstream from AUG1, was associated with alternative splicing of intron 1 (1405 nt) in Td (intron 1 spliced distal transcripts). Finally, Td contained the predicted 6966 nt ICP4 ORF (ORF6966), whereas Td contained a shorter predicted ORF of 6318 nt (ORF6318) starting from the second in-frame AUG (AUG2, position 169047) (Fig. 2a).

We then investigated which the Tp (produced from Pprox), Td (produced from Pdis) ICP4 transcripts were differentially expressed according to the replication phase of GaHV-2. We performed specific screening, by multiplex PCR, of hundreds of clones from the 5' nested RACE-PCR libraries for each cell type (CEF RB-1B, MSB-1 and MSB-1R cells), with the A90, M570 and A654 primers (Fig. 2d). Tp, Td (produced from Pdis) and Td (produced from Pdis) ICP4 transcripts were differentially expressed according to the replication phase of GaHV-2. We performed specific screening, by multiplex PCR, of hundreds of clones from the 5' nested RACE-PCR libraries for each cell type (CEF RB-1B, MSB-1 and MSB-1R cells), with the A90, M570 and A654 primers (Fig. 2d). Tp, Td (produced from Pdis) and Td (produced from Pdis) ICP4 transcripts were differentially expressed according to the replication phase of GaHV-2.
infection (as in CEF RB-1B cells), 25 % of ICP4 mRNAs corresponded to Tp transcripts from Pprox and 75 % were transcribed from Pdis (with 44 and 31 % of spliced transcripts Td_A and Td_B, respectively) (Fig. 2d). We observed a similar distribution during latency (MSB-1 cells) for the three types of transcripts Tp (35 %), Td_A (47 %) and Td_B (18 %). By contrast, during the reactivation phase (MSB-1R cells), 90 % of the ICP4 mRNAs were Tp transcripts, mostly associated with...
the TSSp1 Inr (80 %). The remaining 10 % of transcripts consisted solely of TdA spliced transcripts containing the long ORFp9665 (Fig. 2b). Thus, depending on the phase of the viral cycle of GaHV-2, two alternative promoters and alternative splicing mechanisms lead to the generation of three different transcripts potentially associated with two nested ICP4 ORFs.

**GC-boxes are strong response elements within the ICP4 proximal core promoter**

We studied the Pprox core promoter, by amplifying the 339, 501, 1007 and 1505 bp immediately upstream from the TSSp1, by inserting them into the pG3L-basic vector (firefly luciferase) to produce a nested set of four 5′-truncated constructs (Ppro×340, Ppro×500, Ppro×1000 and Ppro×1500; Fig. 3a). MSB-1 cells were cotransfected with the Pprox constructs, the pG3L-basic vector (negative control) or the pCMV-MLuc (positive control) together with the pRL-TK vector (Renilla luciferase), and the resulting promoter activity was measured in luciferase assays. The ratio of firefly/Renilla luciferase activity for each template was arbitrarily expressed relative to that for the longest construct (Ppro×1500).

Unlike the pG3L-basic control, Ppro×1500 displayed efficient promoter activity at levels that were 42 % that of the standard CMV promoter of pCMV-MLuc (data not shown). The three truncated Pprox constructs (Ppro×1000, Ppro×500 and Ppro×340) had levels of activity corresponding to 65, 88 and 91 % of Ppro×1500 activity, respectively (Fig. 3a). We therefore defined Pprox340 as the minimal functional proximal core promoter of ICP4.

We then used Genomatix software to identify four response elements (REs) in the Pprox340 core promoter sequence: two GC boxes with the Pprox constructs, the pG3L-basic vector (negative control) or the pCMV-MLuc (positive control) together with the pRL-TK vector (Renilla luciferase), and the resulting promoter activity was measured in luciferase assays. The ratio of firefly/Renilla luciferase activity for each template was arbitrarily expressed relative to that for the longest construct (Ppro×1500).

Unlike the pG3L-basic control, Ppro×1500 displayed efficient promoter activity at levels that were 42 % that of the standard CMV promoter of pCMV-MLuc (data not shown). The three truncated Pprox constructs (Ppro×1000, Ppro×500 and Ppro×340) had levels of activity corresponding to 65, 88 and 91 % of Ppro×1500 activity, respectively (Fig. 3a). We therefore defined Pprox340 as the minimal functional proximal core promoter of ICP4.

**DNA methylation patterns of the proximal and distal ICP4 core promoters**

We assessed the methylation patterns of the proximal and distal ICP4 core promoters by bisulfite genomic sequen-
cing analysis (BGSA) on viral genomic DNA extracted from MDV-infected cells in the three phases of the viral cycle. An analysis of the sequences of the PCR products indicated that the proximal and distal promoters had similar methylation patterns for each sample analysed (Fig. 5). The ICP4 promoters were unmethylated during the lytic phase (CEF RB-1B, <2 % of cytosines methylated), but hypermethylated (50–62 % of cytosines methylated) during viral latency (MSB-1 cells) (Fig. 5). We observed a decrease in the overall percentage methylation of ICP4 promoters during the reactivation phase (with 20–24 % of cytosines methylated in MSB-1R cells) relative to latency (Fig. 5).

Two regions seemed to be involved in the epigenetic regulation of Pdis500: the region from −387 to −117, which remained methylated in both the latency and reactivation phases (with 55 and 28 % of cytosines methylated, respectively), and the region from −109 to +10, which harboured five of the six effective REs and was hypomethylated in latency (56 % of cytosines methylated) but hypermethylated in the reactivation phase (8 % of cytosines methylated) (Fig. 5b).

We identified 17 CpG dinucleotides in the ICP4 Ppro×340 promoter, and 26 in the Pdis500 promoter. One of the CpG dinucleotides of Ppro×340 (position −61) did not seem to be regulated by DNA methylation, as it was almost entirely unmethylated in all replication phases (Fig. 5a). By contrast, epigenetic regulation was observed for four notable CpG dinucleotides hypermethylated in latency (MSB-1 cells), and hypomethylated or
unmethylated during the reactivation (MSB-1R cells) and lytic phases (CEF RB-1B cells) (Fig. 5). One of these CpG dinucleotides was enclosed in the GC box of Pprox340 (position −30) and the other three were in the CRE, GC₂ and CCAAT boxes of the −109 to +10 region of Pdis500 (position −56, −69 and −109). Given the involvement of these effective REs in ICP4 promoter activity (Figs 3b and 4b), the methylation of these REs may modulate ICP4 promoter activity during viral latency.

Alternative splicing of two introns in ICP4 mRNA defines five nested ICP4 ORFs

We then focused on ICP4 ORF transcription. We prevented the unwanted amplification of the ICP4 antisense LAT lncRNAs expressed in infected cells by first analysing the pattern of ICP4 transcripts in GaHV-2-free DT-40 cells (previously validated for GaHV-2 splicing studies [23]) transfected with a pcDNA vector containing the large ICP4 ORF (pcDNA-ICP4). We carried out six overlapping

Fig. 3. Characterization of the proximal ICP4 promoter. (a) Determination of the core proximal promoter of ICP4. Schematic representation of the 3’ nested promoter constructs in the pGL-3 Basic vector, with numbers indicating nucleotide positions relative to the RB-1B genome (GenBank accession no. EF523390.1). The six proximal TSSs are shown (arrows). The histogram represents the relative luciferase activity of each promoter construct measured in luciferase assays after the transfection of MSB-1 cells. The data are expressed relative to the luciferase activity of the longest construct (Ppro×1500, arbitrarily set to 1). Error bars indicate the SEM of three independent experiments performed in triplicate. (b) Identification of the response elements (REs) involved in the promoter activity of the ICP4 proximal core promoter construct Ppro×340. The Oct-1, ETS and two GC boxes identified by in silico analysis are shown in grey in Pprox340 and their locations relative to the TSSp are indicated. The mutated REs are shown in white in the corresponding constructs. The histogram shows the luciferase activity measured after three independent experiments performed in triplicate in MSB-1 cells, normalized relative to that of Ppro×340, set to 1. The error bars indicate the SEM of the three experiments. *, Student’s t-test P-value<0.05 and **, P-value<0.01 versus Ppro×1500 in (a) and Ppro×340 in (b).
RT-PCRs covering the entire ICP4 ORF on total RNA extracted from pcDNA-ICP4-transfected cells. The first RT-PCR, performed with the M570/A653 primers, amplified a single 315 nt product corresponding to the 5' end of ICP4 exon 2 (Fig. 6a, lane 1). The second RT-PCR, performed with the M570/M659 primers, amplified two products of 725 and 908 bp (Fig. 6a, lanes 2A and 2B). Sequencing analysis of these amplicons identified two splicing events relating to a newly identified intron (intron 2), involving the use of two alternative 5'SSs (GU, positions 168790 and 168973) and a single 3'SS (AAG, position 170492), resulting in the splicing out of introns 2A (1703 nt) and 2B (1520 nt), respectively. Each of the last four RT-PCRs amplified a single product, not corresponding to genomic amplification, as shown by the results for the RT negative control template (Fig. 6b), defining a unique 4872 nt exon 3 (position 170493–175364) at the 3' end of the ICP4 transcripts (Fig. 6a, lanes 3–6). Finally, the alternative splicing of introns 2A and 2B led to the prediction of two additional GaHV-2 ICP4 ORFs of 5082 and 5265 nt (ORF5082 and ORF5265, respectively), beginning at a third AUG (AUG', position 168580).

We also assessed the regulation of the alternative splicing of intron 2 in GaHV-2-infected cells (CEF RB-1B, MSB-1 and MSB-1R cells) by performing exon2–exon3 RT-PCR with
Fig. 5. DNA methylation patterns of the ICP4 core promoters Pprox340 and Pdis500 during the different phases of the viral cycle. Effective methylation, calculated as the rate of isolated cytosine conversion to thymine following bisulfite treatment, is shown in black, in the bar under each CpG site, according to the scale provided below. On the right, we indicate the global DNA methylation pattern for all sequences analysed (minimum 10) under 3 sets of conditions: lytic phase (CEF RB-1B); latency (MSB-1) and reactivation (MSB-1R).

(a) The numbers indicate the extremities of the ICP4 core proximal promoter Pprox340 sequence relative to the GaHV-2 RB-1B genome (GenBank accession no. EF523390.1). The response elements of Pprox340 are schematically represented. The TSSp1 (arrow) arbitrarily used as a reference for the precise location of the CpG dinucleotides (frame) is indicated. (b) The numbers indicate the extremities of the ICP4 core distal promoter Pdis500 sequence relative to the GaHV-2 RB-1B genome (GenBank accession no. EF523390.1). The response elements of Pdis500 are schematically represented. The TSSd2 (arrow) arbitrarily used as a reference for the precise location of the CpG dinucleotides (frame) is indicated.
the B070/B047 primers specifically designed to prevent amplification of the spliced antisense LAT lncRNA (Fig. 6b). Alternative splicing of intron 2 was confirmed in infected cells by the sequencing of all the amplicons, with different patterns being observed for different phases of the viral cycle. As observed with pcDNA-ICP4, two products of
266 and 449 bp corresponding to exon 2a/2b–exon 3 spliced transcripts were amplified during latency (MSB-1) (Fig. 6b). Furthermore, consistent with the results for the RT negative control template (Fig. 6b), a third 1969 bp product corresponding to unspliced exon 2–3 transcripts (intron 2US) was observed solely for the active phase of the viral cycle (CEF RB-1B, MSB-1R). Interestingly, intron 2a spliced transcripts appeared to be produced in smaller amounts than intron 2α spliced mRNAs, particularly during reactivation (MSB-1R), when intron 2α spliced transcripts were undetected (Fig. 6b). Finally, A90/B047 RT-PCR identified a new spliced transcript specific to the lytic phase, which joined exon 1 to exon 3 from the constitutive 5′ and 3′SS of intron 1 and intron 2 (intron 1C, 3473 nt), respectively, in which a fifth ICP4 ORF of 4812 nt (ORF4812) starting from AUG3 (position 170553) was predicted (Fig. 6c). Moreover, we confirmed the transcription of ICP4 in vivo in peripheral blood lymphocytes and feather follicles collected from RB-1B infected chickens (Fig. S3a) [24] and observed a kinetic transcriptional regulation of the Tp, TdA, Tdα, and TdC transcripts throughout the infection (Fig. S3b). Overall, we identified three exons and two introns, the alternative splicing of which resulted in the prediction of five nested ORFs for the GaHV-2 ICP4 gene.

**DISCUSSION**

Our results provide a new insight into GaHV-2 replication by characterizing, at the transcriptional and post-transcriptional levels, regulation of the transcription of the major transactivator ICP4 gene during the lytic, latent and reactivation phases of the viral cycle. We showed that the ICP4 gene is larger than previously predicted (6966 nt), with a length of 9438 nt (166762–176199) from TSSd1 to the termination signal of 3′UTR-XL (Figs 1 and 2), consistent with the uncharacterized 10 kb ICP4 mRNA reported in a previous study [21]. Its expression is controlled by a dual promoter located in the 2380 nts upstream from ATG1 (Fig. 2). The existence of such a dual promoter for the transcription of the GaHV-2 ICP4 gene, previously proposed on the basis of partial data [1, 17], has also been reported for the homologous ICP4 genes of EHV-1 and FHV-1 and for the GaHV-2 IE ICP27 gene [25–27]. The HSV-1 ICP4 promoter presents one cis-repressor motif recognized by ICP4 and two cis-activator TAATGARAT motifs recognized by the VIC [2]. No ICP4 repressor motif was identified in our two promoters, and subsequent studies of the self-regulatory nature of the GaHV-2 ICP4 protein will require its prior fine characterization. The lack of the TAATGARAT motif in the ICP4 promoter sequence and former data showing that VP16 is non-essential to GaHV-2 virus growth [28] suggest that activation of the cascade of genes during the viral cycle of GaHV-2 could operate through a different mechanism than HSV-1, which remains to be determined. The proximal and distal promoters had similar transcriptional activities, associated principally with two and three efficient GC boxes, respectively (Figs 3 and 4).

The ICP4 GaHV-2 gene consists of two introns and three exons. Its expression is differentially regulated during the viral cycle by alternative splicing, resulting in five predicted nested ORFs (ORF4812, ORF318, ORF1812, ORF5265 and ORF5082) starting from four different AUGs (AUG1, 2, 3 and AUG′, respectively) (Figs 2 and 6). Alternative splicing of the first intron was directly correlated with use of the distal promoter and associated with the production of TdA, TdB and TdC transcripts (Figs 2 and 6). As reported for the homologous ICP4 genes of BoHV, EHV and FHV transcribed from their distal promoters [25, 26, 30], the first intron (intron1A) of the Tdα transcript was spliced upstream from the first AUG (AUG1) of the large ICP4 ORF (ORF6966). By contrast, introns 1B and 1C of TdB and TdC, respectively, were spliced from two other 3′SSs located downstream from AUG1, directly prevented expression of the large ICP4. Similar splicing events, leading to the loss of the first AUG, and thus, to the expression of IE isoproteins with specific functions during the viral cycle, have been described for the homologous ICP4 of EHV [26, 32]. In addition, our transcript analyses revealed a differential regulation of ICP4 transcripts according to the phase of the viral cycle. Indeed, we observed all the alternative forms of intron 1 (1A, 1B and 1C) and intron 2 (2A, 2B and unspliced) during the lytic phase, whereas some of these forms were not detected during latency and reactivation (i.e. Tdc, unspliced intron 2US and intron 2B) (Figs 2 and 6). Although we cannot rule out the possibility that some of the unspliced intron
2 transcripts correspond to pre-mRNA amplification, three major observations provide strong support for the hypothesis of intron retention (Figs 2 and 6): (i) the lack of this amplification in MSB-1 cells but not in MSB-1R cells; (ii) the absence of amplification of unspliced intron 1 from the same potential pre-mRNA template in all the samples; and (iii) previous characterization of the large ICP4 protein necessarily harbouring the intron 2 sequence [17]. The promoter and splicing data, demonstrating specific transcriptional and post-transcriptional regulation of GaHV-2 ICP4 during the viral cycle, clearly suggest that the mechanisms governing ICP4 activation and initiation of the cascade of gene expression are different during the lytic phase and during reactivation from latency.

Interestingly, the third exon common to all five 3′ nested ORFs encodes all the homologous domains defining the ICP4 superfamily and sufficient for the conserved transactivator properties of GaHV-2 ICP4 [14, 20, 21]. All four iso-proteins of the early protein (E1) expressed following alternative splicing of the UL112/113 mRNA of MCMV were recently reported to be essential for the effective completion of the viral cycle of this herpesvirus [33]. Similarly, it is possible that some or all of the predicted isoforms corresponding to the five ICP4 ORFs differentially expressed during the different phases of the GaHV-2 viral cycle may be involved in the completion of this cycle. During the lytic phase, all five ORFs seem to be produced by the alternative splicing of introns 1 and 2, whereas only two and three of the five ORFs seem to be associated with reactivation (ORF<sub>6966</sub> and ORF<sub>5265</sub>) and latency (ORF<sub>6318</sub>, ORF<sub>5265</sub> and ORF<sub>5082</sub>), respectively (Figs 2 and 6). Finally, the role of each ORF remains to be investigated, particularly for ORF<sub>5265</sub>, which is detected in all phases of the cycle, the short ORF<sub>4912</sub> specific to the lytic phase and the large ORF<sub>6966</sub> undetectable during latency, a phase in which ICP4 has been shown to be involved in the maintenance of the GaHV-2 lymphomagenesis [16].

In addition to the predicted poly(A) signal of the 3′ UTR-S [14], three other effective poly(A) signals are found at the end of ICP4 mRNAs, the 3′ UTR-M, -L and -XL (Fig. 1). Contrary to our observations concerning the choice of proximal/distal promoter and the alternative splicing of GaHV-2 ICP4 introns, we observed no major variation in the distribution of the four 3′ UTRs according to the phase of the viral cycle. However, it is now widely accepted that the production of mRNA isoforms via alternative polyadenylation (APA) mechanisms plays a major role in post-transcriptional mRNA stability, nuclear export, localization and translation efficiency [34]. Depending on the inclusion or exclusion of potential miRNA responsive elements (miREs) in the different 3′ UTRs, APA may also lead to differential negative post-transcriptional regulation of the corresponding mRNA isoforms. Building on the findings of Bai et al., who identified numerous miREs targeted by viral miRNAs in the 3′ UTRs of polycistronic HHV-8 transcripts, leading to a fine-tuning of viral gene expression during the different phases of the viral cycle [35], we performed a bioinformatics analysis to search for viral miREs in the GaHV-2 ICP4 3′UTRs. In addition to the previously analysed mdv1-miR-M7-5P miRE [9], an miRE for mdv1-miR-M4-3P was identified in the 3′UTR-L and -XL (Fig. S1). This element seemed to be functional in luciferase assays performed in MSB-1 cells (data not shown), suggesting that, like mdv1-miR-M7-5P, mdv1-miR-M4-3P may also downregulate ICP4 mRNAs ending in the 3′UTR-L and -XL, corresponding to 70% of all ICP4 mRNAs (Fig. 1). Moreover, in the light of RNA interference mechanisms and growing knowledge about IncRNA, our results indicate that the ICP4 gene is fully complementary to the LAT IncRNA [9], which may thus play a major role in the post-transcriptional regulation of ICP4 mRNAs during the viral cycle, particularly during latency, through cis-acting effects, as a natural antisense IncRNA, or through trans-acting effects, as already described for HHV-1 [36].

We show here that the GaHV-2 ICP4 gene is expressed throughout the viral cycle and that, as the major IE gene, it is highly regulated at the transcriptional and post-transcriptional levels, with specific regulation controlling the transition between the lytic, latent and reactivation phases. The fine-tuning of ICP4 expression involves both transcriptional regulation by epigenetic modification of the two alternative promoters, and post-transcriptional mechanisms, such as RNA interference, the use of four alternative poly(A) signals and alternative splicing of the two introns, resulting in five predicted nested ORFs. The lack of the TAATGARAT motif in the ICP4 promoter of GaHV-2 suggests a specific feature for the activation of its transcription, which could be regulated throughout the viral cycle by a differential recruitment of the Sp1, 4 factors at GC boxes, as previously observed for other herpesviruses and notably the lymphotrophic ones [37–41]. ICP4, ICP27, the MEQ oncoprotein and the viral non-coding RNAs are also likely involved in the control of the transcriptional and post-transcriptional regulation of ICP4 throughout the viral cycle [9, 19, 27]. The functional consequences of such a regulation are difficult to predict, as the homologous ICP4 proteins were shown to be involved in many general cellular processes (i.e. histone modifications, transcriptional regulation and protein trafficking), but the first consequence appears to be directly correlated to the differential expression of the five ICP4 ORFs during the viral cycle [2, 13]. The effective translation of the five ORFs and their putative specific features remain to be explored, together with the interplay between the viral and cellular processes involved in the fine-tuning of GaHV-2 ICP4 gene expression. Our results also indicate that the reactivation phase does not simply mimic the lytic phase, but is a distinct viral phase in its own right. This may reflect features specific to GaHV-2, including the integration of the viral genome during latency and lymphomagenesis. We will now perform further studies to investigate the specific regulation of the essential and major IE ICP4 of GaHV-2 throughout the entire viral cycle in more detail to improve our understanding of the alphaherpesvirus cycle.
METHODS

Cells

The MSB-1 cell line is derived from a spleen lymphoma induced by a virulent strain of GaHV-2 [42]. It was maintained in RPMI-1640 medium (Lonza) supplemented with 1 mM sodium pyruvate, 10% foetal bovine serum and 5% chicken serum. Viral reactivation from latency was induced by the treatment of MSB-1 cells (MSB-1R) in the exponential phase of growth with 3 mM n-butyrate (Sigma) for 48 h. Chicken embryo fibroblasts were prepared for propagation of the RB-1B virus strain (CEF RB-1B), as previously described [43]. The DT-40 B-cell line, which is derived from an avian leukosis virus-induced lymphoma, was cultured in a similar manner, in Dulbecco’s modified Eagle’s medium (Lonza) supplemented with serum.

Cell transfection

DT-40 lymphoblastoid cells were transfected with the pcDNA-ICP4 construct by electroporation with an Amaxa Biosystems Nucleofector machine (Lonza). Cell density was adjusted to 3 x 10^6 cells/100 ml of Nucleofector solution T (Lonza) containing 3 µg of plasmid. The cells were electroporated in a 4 mm cuvette subjected to specific programme B-023. The electroporated cells were recovered in 3 ml of complete medium and dispensed into a six-well plate. They were harvested 48 h after transfection for the extraction of RNA.

For luciferase reporter assays in MSB-1 cells, we transfected one million MSB-1 cells with 2 µg pRL-TK Renilla luciferase reporter construct and 60 ng of the control vector pcDNAMLuc [44] with an Amaxa Biosystems Nucleofector machine (Lonza), following the same protocol as for the electroporation of DT-40 cells but with the X-001 programme. The transfected MSB-1 cells were harvested 24 h after electroporation for the measurement of luciferase activity.

RACE-PCR and RT-PCR analyses

Total RNA was isolated from GaHV-2-infected cells (CEF RB-1B, MSB-1 and MSB-1R cells), with the Trizol extraction kit (Invitrogen) and treated with DNase (Promega), and RACE was carried out with the GeneRacer kit (Invitrogen), according to the manufacturer’s instructions and with the primers shown in Table S1b.

We reverse-transcribed 5 µg of RNA with a mixture of oligo (dT) and random primers (Eurogentec) and the Super-script-III reverse transcriptase (Invitrogen). The efficacy of the DNase treatments of RNA was systematically confirmed by PCR. A negative control for reverse transcription (RT-) was carried out systematically in the same conditions. The cDNAs generated were amplified by PCR with the Gotaq DNA polymerase (Promega) or with the Thermo Scientific Phusion High-Fidelity DNA polymerase (New England Biolabs) for ICP4 mRNA splicing analyses, both used according to the manufacturer’s instructions and with the primers listed in Table S1b.

All PCR products were inserted into pGEM-T Easy (Promega) and the sequence of the corresponding inserts was checked (GATC Biotech). All of the sequences were analysed by comparison with the viral RB-1B sequence (GenBank accession number RB-1B EF523390) with Geneious software (www.geneious.com) and BLAST.

Plasmid construction

The large ICP4 ORF6966 corresponding to the nucleotide 168399 to 175364 region of the GaHV-2 RB-1B sequence was amplified by PCR from the bacmid RB-1B [45] (kindly provided by Venugopal Nair, Institute for Animal Health, Compton, UK) with the Thermo Scientific Phusion HF DNA polymerase (New England Biolabs) and the A534/A535 primers (Table S1b). It was inserted into the XhoI-BamHI double-digested pcDNA3.1 expression vector to produce the pcDNA-ICP4 construct.

For the pGL3 luciferase reporter constructs, all the putative promoter sequences were amplified with the primer pairs described in Table S1c and inserted into the pGL3-Basic vector (Promega), between the KpnI and HindIII restriction sites. We generated the mutated constructs by inserting the PCR products amplified with mutated primers (see Table S1c) from the corresponding RE sequences into KpnI-HindIII double-digested pGL3. All of the plasmids were purified with the NucleoBond Xtra Midi kit (Macherey-Nagel), and the sequences of all inserts from each construct were systematically verified (GATC Biotech).

Dual luciferase reporter assays

The assays were carried out as previously described [43]. Luciferase activity was quantified with the Dual-Luciferase reporter system (Promega), according to the manufacturer’s protocol. For the standardization of inter- and intra-assay luciferase activity, we systematically used the control vector pcDNAMLuc, which carries the firefly luciferase gene under the control of the CMV promoter [44]. Firefly and Renilla luciferase activity was measured 24 h after transfection. Luminescence was measured with a luminometer (Tristar luminometer; Berthold Technologies). Three independent experiments were carried out in triplicate. Student’s t-test was used for statistical analysis.

DNA isolation, bisulfite (BS) treatment and PCR

DNA was isolated with the DNeasy Blood and Tissue kit (Qiagen). BS treatment was performed with the EZ DNA methylation-Gold kit (Zymo Research), which converts unmethylated cytosine to uracil residues, which are then converted into thymidine residues during PCR. Nested PCR was performed with the primers listed in Table S2, and Epiclone Hot-Start Taq DNA polymerase (New England Biolabs). Amplicons were inserted into the pGEM-T Easy vector (Promega) and the DNA insert was checked by sequencing.

Sequence analysis software

The MatInspector module of Genomatix software was used to search for response elements (REs) in promoter
sequences. Geneious or Chromas and A plasmid Editor (ApE) software was used to analyse the DNA sequences.

Funding information
I.G. is a ‘Fonds pour la formation à la recherche dans l’industrie et dans l’agriculture’ (FRIA) PhD fellow (Belgium). This work was supported by grants from the ‘Ligue contre le Cancer Grand Ouest’ (France), including the committees for départements ’16, 18, 29, 35, 36, 37 and 85, and a grant from the ‘Fonds Européen de Développement Régional’ (FEDER) www.europe-france.gouv.fr (ExMIR (no. 2835-34204), a grant from the ‘Région Centre-Val de Loire’ (www.region-centre-valdeloire.fr) MIRES (no. 200900038264) and a grant from the ‘Fonds de la recherche Scientifique’ (FNRS; grant FRFC J.0165.13, Methyllycine in Marek’s Disease Virus (Belgium)).

Acknowledgements
We thank Venugopal Nair (Institute for Animal Health, Compton, UK) for kindly providing the RB-1B bacmid. We also thank E. Cottin for technical support.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References
30. Schwyzmer M, Vlcek C, Menekse O, Fraelch C, Paces V. Promoter, spliced leader, and coding sequence for BICP4, the largest of the


### Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.