Journal of General Virology (1998), **79**, 2529–2532. Printed in Great Britain

SHORT COMMUNICATION

Identification of a novel multifunctional structural domain in the herpes simplex virus type 1 genome: implications for virus latency

J. P. Quinn,[1] R. A. McGregor,[1] C. E. Fiskerstrand,[1] C. Davey,[2] J. Allan[2] and R. G. Dalziel[1]

Departments of Veterinary Pathology[1] and Biochemistry[2], University of Edinburgh, Edinburgh EH9 1QH, UK

A domain, previously termed RE1, exists within the herpes simplex virus type 1 genome potentially influencing expression of immediate early genes and the latency associated transcripts. This domain consists of 10 tandem copies of a CT-rich sequence. We demonstrate that this domain binds multiple host-cell factors that may allow RE1 to act either as a transcriptional regulator and/or to affect nucleosomal and DNA structure in the latent genome.

Herpes simplex virus type 1 (HSV-1) can establish a latent infection that persists for the lifetime of the host. Latent virus genome exists in an extrachromosomal state and is complexed with host nucleosome protein (Mellerick & Fraser, 1987; Deshmane & Fraser, 1989). This chromosomal structure may have consequences for the establishment, maintenance of, and reactivation from latency. The area encompassing the latency associated transcripts (LATs) and the immediate early (IE) genes is implicated in reactivation and we have analysed this region of the HSV-1 genome to determine whether it contains domains which might regulate transcription over this large area. CpG motifs in eukaryotic genes are often found clustered 5′ of the transcription start site and can regulate both the transcriptional and methylated state of the loci (Cross & Bird, 1995). Our analysis revealed extensive CpG motifs with one notable exception being a domain lying 5′ of the LAT promoter. This domain represents the tandem repeat region previously termed reiteration set 1 (RE1) (Perry & McGeoch, 1988) (Fig. 1 A). The sequence of the element constituting this repeat is (CCCCTCTCCCCCCCTCT) × 10. We predicted that this unit was capable of binding to factors that can alter chromatin structure or act as transcription regulators. Factors potentially recognizing and binding to this motif are shown in Fig. 1 (B). In all cases the effects could act over a large distance and affect several genes near this locus. Although proteins expressed during viral infection could further regulate this region, we chose initially to determine whether neuronal host-cell proteins could interact in a sequence-specific manner with this element. Host proteins are likely to be involved, at least in part, in regulating maintenance or reactivation from latency.

Binding of factors to this domain was analysed by electrophoretic mobility shift analysis (EMSA). Nuclear extract was prepared (Quinn *et al.*, 1989) from either the brains of adult Wistar rats or HeLa cells which were grown in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% heat-inactivated calf serum. Each reaction using double-stranded oligonucleotide contained 0·5 ng ^{32}P 3′-end-labelled oligonucleotide probe generated using Klenow enzyme. This end-labelled probe was mixed with 20 µg protein extract which had been pre-incubated on ice with 1 µg poly(dI–dC):(dI–dC). Competitor oligonucleotide where appropriate was added to the extract at the same time as the probe. Each reaction using single-stranded oligonucleotide contained 0·5 ng ^{32}P 5′-end-labelled oligonucleotide probe generated using T4 kinase. All probes were purified over G25 spin columns. Reactions were incubated for 15 min at room temperature and separated on a 4·2% polyacrylamide (29:1) gel. As RE1 is a multimerized element we chose to address binding to one of the repeats. EMSA demonstrated that the motif would interact with three classes of transcription factor.

(1) Classical double-stranded DNA-binding proteins recognizing CT-rich elements (Fig. 2 A)

(2) Single-stranded DNA-binding proteins (Fig. 2 B, C)

(3) Nucleosomal phasing proteins related to BGP1 (Fig. 3 A, B)

Growth factors, including nerve growth factor (NGF), are known to modulate HSV-1 latency in model systems (Wilcox & Johnson, 1988; Wilcox *et al.*, 1990). dG-rich motifs are growth factor-regulated in the gastrin gene (Bachwick *et al.*, 1992) and in general dGC-rich elements can mediate the action of growth factors, e.g. NGFIA (Kendall *et al.*, 1994). The HSV-1 GC motifs within RE1 may therefore act as transcriptional regulators under appropriate conditions during virus infection and latency in sensory neurons. Enhancer elements in viruses are often found to be multimerized motifs which support higher levels of gene expression than the single element alone,

Author for correspondence: R. G. Dalziel.
Fax +44 131 650 6511. e-mail r.g.dalziel@ed.ac.uk

0001-5650 © 1998 SGM

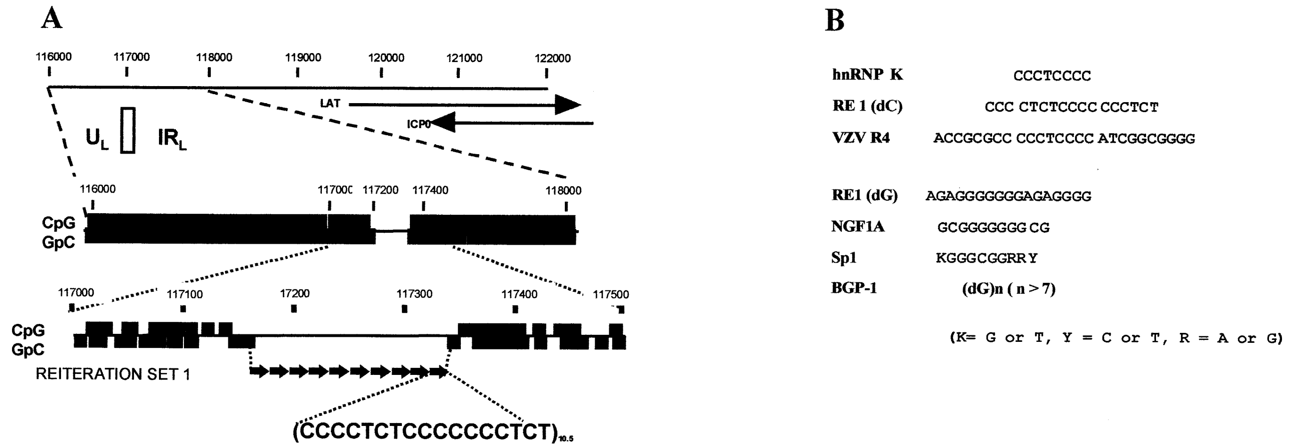

Fig. 1. (A) The occurrence of the dinucleotides CpG and GpC was mapped using the MAP program of GCG9. CpG is indicated by the presence of a vertical line above the horizontal line and GpC by a line below. The positions of the U_L/IR_L junction of HSV-1 and the location and orientation of the LAT and ICP0 transcripts are indicated. Short horizontal arrows indicate individual copies of RE1 and the sequence of a single copy of the RE1 repeats is shown. Numbering is from Perry & McGeoch (1988). (B) The consensus binding sites of hnRNP K, NGF1A, Sp1 and BGP-1 are aligned for maximal homology with the RE1 elements. A related element from VZV, termed R4, is also shown aligned with RE1.

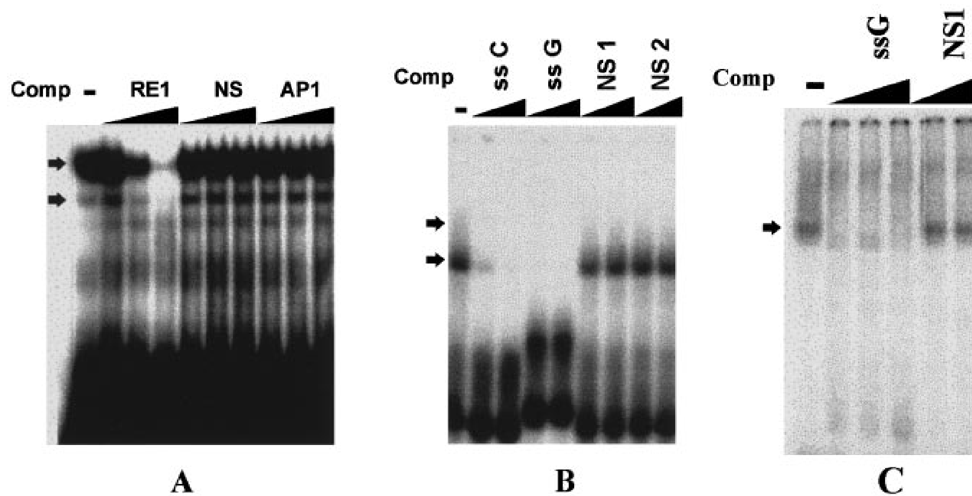

Fig. 2. (A) EMSA performed using rat brain extract and labelled double-stranded RE1 oligonucleotide: 10-, 50- and 100-fold excess competitor was added to the appropriate reactions. Competitors: RE1, homologous double-stranded oligonucleotide; NS, heterologous AT-rich double-stranded oligonucleotide; AP-1, AP-1-binding site from the gibbon ape leukaemia virus LTR (Quinn *et al.*, 1989). The arrows indicate the specific complexes formed. (B) EMSA performed using rat brain extract and the labelled dC strand of RE1: 10-, 50- and 100-fold excess competitor was added to the appropriate reactions. Competitors: ssC, homologous RE1 dC strand; ssG, complementary RE1 dG strand; NS1/NS2, single strands of heterologous AT-rich 'NS'. The arrows indicate the specific complexes formed. The complexes formed on the dC strand are competed by the dG strand as these strands form a double-stranded oligonucleotide in the reaction. This is observed in the slower migration of the unbound probe in those lanes that contain dG competitor. (C) EMSA performed using rat brain extract and labelled dG strand of RE1: 10-, 50- and 100-fold excess of homologous and 50- and 100-fold excess of heterologous competitor was added to the appropriate reactions. Competitors: ssG, homologous RE1 dG strand; NS1, single strand of heterologous AT-rich 'NS'. The arrows indicate the specific complexes formed.

often acting synergistically rather than additively (Herr & Clarke, 1986; Quinn *et al.*, 1989).

An enhancer function in a virus would not be unexpected; however, the two additional properties (2 and 3 above) of this domain may represent a novel mechanism involved in alphaherpesvirus latency. These additional properties would also change the topology over this region having subsequent consequences for the parameters regulating latency. Firstly, EMSA demonstrated that each of the single strands of a single RE1 repeat bound in a sequence-specific manner to nuclear proteins. Single-stranded DNA-binding proteins have been demonstrated to act as transcriptional regulators (DavisSmyth

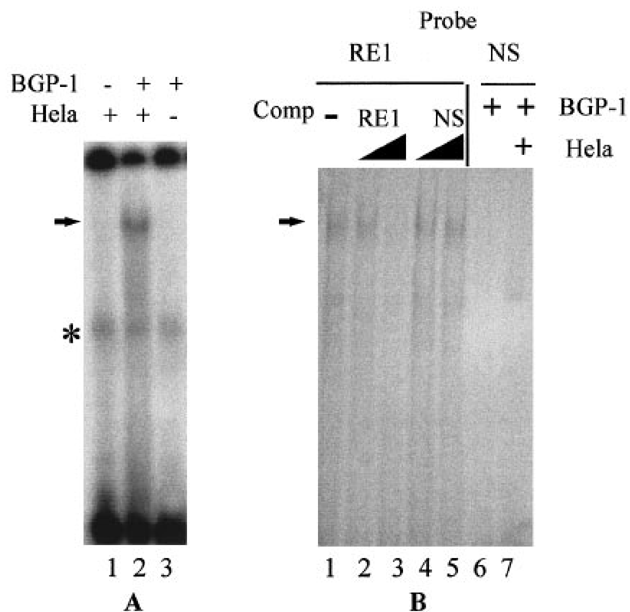

Fig. 3. (A) EMSA was carried out using purified BGP1 alone or complemented with HeLa cell extract and double-stranded RE1 oligonucleotide as probe. Lane 1, HeLa cell extract diluted such that addition of 1 μl to an EMSA resulted in the formation of no specific complexes. Lane 2, both cell extract and BGP1. Lane 3, purified BGP alone. The asterisk indicates the position of a non-specific Ku complex often observed in EMSA. The arrows indicate the specific complexes formed. (B) BGP1 and HeLa nuclear extract diluted as in (A) was mixed with either RE1 (lanes 1–5) or NS (lane and 7) probe.

et al., 1996; Duncan *et al.*, 1994; Michelotti *et al.*, 1996), e.g. the CT regulatory element in the c-*myc* gene (Michelotti *et al.*, 1995; Tomonaga & Levens, 1995, 1996). The sequence similarity between the c-*myc* site and the RE1 element suggested to us that they might bind related factors. This c-*myc* element will bind hnRNP K to the dC-rich strand and CNBP to the dG-rich strand. In a similar way the dC and dG strands of the RE1 element also bound specifically to factors in the extract (Fig. 2 B, C). We have also demonstrated that the c-*myc* element, not unexpectedly given the sequence homology, competes for the binding to the RE1 sites in each of the single-stranded oligonucleotides (data not shown). These proteins may represent a class of factor interacting with the RE1 domain. The hnRNP K protein can stimulate reporter gene activity via binding to its cognate motif; however, the function of the CNBP binding has not been determined. Similarly, binding of single-stranded binding proteins to RE1 could therefore act as transcriptional regulators analogous to the c-*myc* model. Alternatively, the binding of multiple single-stranded DNA-binding proteins could also modulate promoter structure by affecting topology of the DNA as outlined below.

The sequence of the CT element also suggested to us that it was potentially bound by nucleosome restructuring proteins, e.g. BGP-1, which could regulate interaction of transcription factors with eukaryotic promoters (Clark *et al.*, 1990; Lewis *et al.*, 1988). Remodelling of nucleosomal structure of viral genomes to allow transcription has been observed most elegantly in the mouse mammary tumour virus model (Archer *et al.*, 1992). To investigate BGP-1 interactions with RE1, BGP1 was affinity purified from adult chicken erythrocyte nuclear extracts as previously described (Clark *et al.*, 1990; Lewis *et al.*, 1988) and used in EMSA (Fig. 3). Purified BGP1 did not bind directly to RE1; however, complementing this protein with HeLa cell nuclear extract (at a level that does not itself generate any specific complex) demonstrated sequence-specific complex formation on the RE1 elements. We have previously observed such a precedent for complementation of transcription factor-binding activity using a purified fraction and extract (Quinn *et al.*, 1989). The binding of factors exemplified by hnRNP K and BGP1 to the DNA could have distinct and disparate consequences for chromatin/nucleosomal structure of the virus genome. Indeed, binding of double-stranded sequence-specific factors or BGP-related proteins might be expected to significantly antagonize the action of the single-stranded DNA-binding proteins and vice versa.

A recent publication has demonstrated that single-stranded regions are found *in vivo* to be associated with genes rescheduled for reactivation after mitosis (Michelotti *et al.*, 1997). Obviously, these single-stranded regions are therefore acting as genetic markers to control expression patterns. Although latency occurs within post-mitotic neurons such a genomic marker in the HSV-1 genome might allow for initiation of appropriate temporal gene expression. Alternatively, such a region may account for the expression of the LAT transcript observed within a sub-population of neurons containing latent HSV-1, by acting to keep the genome accessible to transcription factors. The potential complexity of this RE1 domain could be increased by the action of nucleosome restructuring by such as BGP1 and virus-encoded proteins.

Our data indicate that the RE1 domain may have an important and novel regulatory function. The location of this domain indicates that it could affect expression of both LAT and IE and transcripts. Alternatively, as regulatory elements can affect transcription over a very large genomic fragment, this domain could affect expression in other HSV-1 genomic locations. Consistent with this, single-stranded regions generated by the action of single-stranded DNA-binding proteins might facilitate the interaction of factors over areas that otherwise might be energetically unfavourable, by acting in a manner similar to a hinge within the promoter (Michelotti *et al.*, 1996, 1997). The number of copies of the repeat element, each capable of interacting with transcription factors, suggests that it could have a strong regulatory effect. Recently, the DR2 repetitive element within the HSV-1 genome has been shown to have regulatory properties (Martin & Weber, 1998). Interestingly, these authors suggested that DR2 might have distinct properties in a genomic structure compared to that present within a plasmid used in transient transfection analysis. It will be of interest to analyse cell-specific interactions with

this motif in dorsal root ganglia and the variation in binding in response to virus infection or stimuli that result in reactivation of HSV-1 latent genomes. Ultimately, the characterization of the function of this domain will require construction of recombination viruses deleted in both sets of the RE1 repeat and analysis *in vivo*. The potential for such domains as a general regulatory mechanism in alphaherpesvirus latency is indicated both by the HSV-1 DR2 data and sequence analysis which reveals similar motifs in the related varicella-zoster virus (VZV), in the reiterated element R4. The R4 domain is located between two IE genes that are transcribed in latently infected cells (Davison & Scott, 1986), perhaps suggesting a functionally similar genomic function in both viruses.

An MRC short-term project grant to J.P.Q. and R.G.D. and a BBSRC award to J.P.Q. supported this work.

References

Archer, T. K., Lefebvre, P., Wolford, R. G. & Hagar, G. L. (1992). Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science* **255**, 1573–1576.

Bachwick, D., Merchant, J. & Brand, S. (1992). Identification of a cis-regulatory element mediating somatostatin inhibition of epidermal growth factor-stimulated gastrin gene transcription. *Molecular Endocrinology* **6**, 1175–1184.

Clark, S., Lewis, C. & Felsenfeld, G. (1990). Properties of BGP1, a poly(dG) binding protein from chicken erythrocytes. *Nucleic Acids Research* **18**, 5119–5128.

Cross, S. H. & Bird, A. P. (1995). CpG islands and genes. *Current Opinion in Genetic Development* **5**, 309–314.

Davison, A. J. & Scott, J. E. (1986). The complete DNA sequence of varicella-zoster virus. *Journal of General Virology* **67**, 1759–1816.

DavisSmyth, T., Duncan, R. C., Zheng, T., Michelotti, G. & Levens, D. (1996). The far upstream element-binding proteins comprise an ancient family of single-strand DNA-binding transactivators. *Journal of Biological Chemistry* **271**, 31679–31687.

Deshmane, S. L. & Fraser, N. W (1989). During latency, herpes simplex virus type 1 DNA is associated with nucleosomes in a chromatin structure. *Journal of Virology* **63**, 943–947.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avigan, M. & Levens, D. (1994). A sequence-specific, single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes & Development* **8**, 465–480.

Herr, W. & Clarke, J. (1986). The SV40 enhancer is composed of multiple functional elements that can compensate for one another. *Cell* **45**, 461–470.

Kendall, G., Ensor, E., BrarRai, A., Winter, J. & Latchman, D. S. (1994). Nerve growth factor induces expression of immediate-early genes NGFI-A (Egr-1) and NGFI-B (nur 77) in adult rat dorsal root ganglion neurons. *Molecular Brain Research* **25**, 73–79.

Lewis, C. D., Clark, S. P., Felsenfeld, G. & Gould, H. (1988). An erythrocyte-specific protein that binds to the poly(dG) region of the chicken beta-globin gene promoter. *Genes & Development* **2**, 863–873.

Martin, D. W. & Weber, P. C. (1998). Role of the DR2 repeat array in the regulation of the ICP34.5 gene promoter of herpes simplex virus 1 during productive infection. *Journal of General Virology* **79**, 517–523.

Mellerick, D. M. & Fraser, N. W. (1987). Physical state of the latent herpes simplex virus genome in a mouse model system: evidence suggesting an episomal state. *Virology* **158**, 265–275.

Michelotti, E. F., Tomonaga, T., Krutzsch, H. & Levens, D. (1995). Cellular nucleic acid binding protein regulates the CT element of the human c-myc protooncogene. *Journal of Biological Chemistry* **270**, 9494–9499.

Michelotti, G. A., Michelotti, E. F., Pullner, A., Duncan, R. C., Eick, D. & Levens, D. (1996). Multiple single-stranded cis elements are associated with activated chromatin of the human c-myc gene in vivo. *Molecular and Cellular Biology* **16**, 2656–2669.

Michelotti, E. F., Sanford, S. & Levens, D. (1997). Marking of active genes on mitotic chromosomes. *Nature* **388**, 895–899.

Perry, L. J. & McGeoch, D. J. (1988). The DNA sequences of the long repeat region and adjoining parts of the long unique region in the genome of herpes simplex type 1. *Journal of General Virology* **69**, 2831–2846.

Quinn, J. P., Farina, A. R., Gardiner, K., Krutzsch, H. & Levens, D. (1989). Multiple components are required for sequence recognition of the AP1 site in the gibbon ape leukemia virus enhancer. *Molecular and Cellular Biology* **9**, 4713–4721.

Tomonaga, T. & Levens, D. (1995). Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *Journal of Biological Chemistry* **270**, 4875–4881.

Tomonaga, T. & Levens, D. (1996). Activating transcription from single stranded DNA. *Proceedings of the National Academy of Sciences, USA* **93**, 5830–5835.

Wilcox, C. L. & Johnson, E. M. (1988). Characterization of nerve growth factor-dependent herpes simplex virus latency in neurons in vitro. *Journal of Virology* **62**, 393–399.

Wilcox, C. L., Smith, R. L., Freed, C. R. & Johnson, E. M. (1990). Nerve growth factor-dependence of herpes simplex virus latency in peripheral sympathetic and sensory neurons *in vitro*. *Journal of Neuroscience* **10**, 1268–1275.

Received 27 April 1998; Accepted 2 July 1998