Updated Epstein–Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV

Orlando de Jesus,1 Paul R. Smith,1† Lindsay C. Spender,1 Claudio Elgueta Karstegi,1 Hans Helmut Niller,2 Dolly Huang3 and Paul J. Farrell1

1Ludwig Institute for Cancer Research, Faculty of Medicine, Imperial College, St Mary's Campus, Norfolk Place, London W2 1PG, UK
2Institut für Medizinische Mikrobiologie und Hygiene, Universität Regensburg, D-93053 Regensburg, Germany
3Department of Clinical Oncology, Sir Y.K. Pao Centre for Cancer, The Chinese University of Hong Kong, Shatin, Hong Kong Special Administrative Region

Two sequences required for activity of the Epstein–Barr virus BART RNA promoter in transfection assays have been identified by site-directed mutagenesis. One contains a consensus AP-1 site; the other has some similarity to Ets and Stat consensus binding sites. Candidate sequences were suggested by mapping a region of unmethylated DNA in EBV around the BART promoter followed by in vivo footprinting the promoter in the C666-1 nasopharyngeal carcinoma cell line, which expresses BART RNAs. The data are presented in the context of a revised EBV DNA sequence, known as EBV wt, that is proposed as a future standard sequence for EBV.

INTRODUCTION

Epstein–Barr virus (EBV) contains about 85 genes but only a few of these are expressed in the EBV-associated human cancers that occur in immunocompetent patients. Since it has become clear that many of the established immortalizing proteins of EBV are not expressed in human tumours, attention has turned to those viral genes that are expressed in cancers for interpretation of the oncogenic properties of EBV. The BART/CST (BamHI A rightward transcript/complementary strand transcript) RNAs were originally identified in nasopharyngeal carcinoma (NPC) samples (Hitt et al., 1989), although the transcripts have also been detected at low level in some Burkitt’s lymphoma and B lymphoblastoid cell lines (Brooks et al., 1993; Chen et al., 1992; Gilligan et al., 1990; Griffin & Xue, 1998; Karran et al., 1992; Raab-Traub et al., 1991; Zhang et al., 1993). There is also evidence for BART RNAs in EBV-positive gastric cancer (Sugiura et al., 1996), Hodgkin’s Disease (Zhang et al., 2001) and in normal EBV persistence (Chen et al., 1999; Gilligan et al., 1991; Kienzle et al., 1998), although the BART region can be deleted from the viral genome without any notable effect on B cell immortalization by EBV. The protein products from the BART RNAs have not yet been fully characterized but several potential products of the various spliced forms of BART RNA have been analysed (Smith et al., 2000) including RPMS1, A73 and BARF0 (RK-BARF0). Biochemical activities of these proteins have been identified that could be relevant to the role of the virus in cancer (Kusano & Raab-Traub, 2001; Smith et al., 2000; Zhang et al., 2001).

The splicing of the BART RNAs is complex (Sadler & Raab-Traub, 1995), with at least 16 different, partly overlapping exons identified already in cDNA. The main full-length cDNA isolated so far (Smith et al., 2000) was able to express the RPMS1 protein when transfected and there was evidence that such RNA constitutes a significant proportion of the BART RNA expressed in the C15 NPC xenograft tumour, which has relatively high expression of the BART RNAs. The transcription start was determined and a plasmid named SK containing EBV sequences from 442 nt upstream of the transcription start, through the first exon and some of the first intron was found to express the correctly initiated and spliced first exon of the BART RNA (Smith et al., 2000).

We now identify genomic sequences around the BART RNA first exon which are protected from DNA methylation in C15 NPC tumour cells, characterize sequences required for transcription from the promoter and demonstrate in vivo footprinting of those sites in a cell line derived from NPC that maintains episomal EBV. Description of exons of the BART RNAs has been complicated by the fact that the gene spans the region of EBV deleted in the B95-8 strain that was sequenced initially (Baer et al., 1984). A revised EBV...
sequence called EBV wt, renumbered to include the B95-8 deletion and various other corrections, is henceforth used to describe the virus and is proposed as a new standard reference sequence for EBV. The revised EBV wt sequence numbering is used to describe all EBV features in this paper.

**METHODS**

**Cell lines.** 293 and C666-1 cells (Cheung et al., 1999) were cultured in DMEM or RPMI with 10% foetal calf serum. Transfection into 293 cells using the calcium phosphate method and RPA analysis were as described previously (Spender et al., 2001; Wensing et al., 2001).

**Plasmid construction.** Plasmid SK (containing EBV 137908–138989 in pCAT) was made by cloning an SplI to KpnI EBV restriction fragment between the SplI and HinclI sites of pBluescript (Stratagene), the KpnI site having been made blunt with T4 DNA polymerase. The EBV fragment was excised using the HindIII and XbaI sites in the Bluescript polylinker and cloned between the HindIII and XbaI sites of pCAT-Basic (Promega). Plasmid BK (containing EBV 138173–138989 in pCAT) was made by cutting SK with HindIII and BglII, blunting the ends with T4 DNA polymerase and religation. Plasmid CK (containing EBV 138289–138989 in pCAT-Basic) was made by Pfu polymerase PCR from SK and cloned between the SplI and XbaI sites of pCAT-Basic. Plasmid SS (containing EBV 137908–138720 in pCAT) was made by cloning an SplI to SplI EBV restriction fragment between the SplI and HindII sites of pBluescript, the SplI site having been made blunt with T4 DNA polymerase. The EBV fragment was excised using the HindIII and XbaI sites in the Bluescript polylinker and cloned between the HindIII and XbaI sites of pCAT-Basic. Mutations were introduced into BK and CK using the Quikchange kit (Stratagene) as described previously (Spender et al., 2001).

**DMS interference footprinting and methylation analysis.** In vivo dimethyl sulphate (DMS) interference footprinting was performed as described previously (Niller et al., 2002). The C15 tumour was propagated in nude mice (Busson et al., 1988) and DNA was extracted by proteinase K digestion and phenol extraction. For DMS interference footprinting and methylation analysis, C15 tumour DNA was digested with either HpaII or MspI, electrophoresed on a 1 % agarose gel and Southern blots were hybridized with the probes indicated in the legend to Fig. 1, labelled by random priming.

**EBV DNA is unmethylated around exon I of the BART gene**

Several studies have shown that most of the EBV genome DNA is methylated in tumour cell lines that have a latent infection with EBV (e.g. Robertson et al., 1996; Salamon et al., 2001). Absence of DNA methylation in a region of the EBV genome can be an indicator of locations where transcription factors required for promoter activity may be bound during EBV latency. Comparison of restriction digestion by HpaII and MspI on Southern blots of C15 tumour EBV DNA indicated a region of mostly unmethylated DNA extending from about 138200 to 139200 on the EBV wt map (Fig. 1). For example, probes 4 and 5 in the hypomethylated region give mostly the same sized bands with the two enzymes whereas with probes 1 and 2 show no digestion with the methylation-sensitive HpaII enzyme. The region of hypomethylation extends from just upstream of the transcription start to a significant distance downstream of exon 1 (Fig. 1). We previously showed (Smith et al., 2000) that promoter activity could be observed in the plasmid SK (Fig. 1) in transfection assays.
Most of our previous investigation of EBV gene expression in NPC has used the C15 xenograft because there has been a lack of NPC cell lines that retain their EBV. The recently described C666-1 line (Cheung et al., 1999) is derived from an NPC, retains its EBV and has been shown to have a restricted latent pattern of EBV gene expression. The cells make EBNA-1 protein but not EBNA-2 or LMP1 (Cheung et al., 1999). Using similar RPA probes to those applied previously in C15 (Smith et al., 2000), BART RNA expression was also readily detected in C666-1 RNA. Correctly spliced exon I and the boundary between exons VIIA and VIIB were demonstrated (Fig. 2). In each case, RNA that was not spliced at the splice junction was also detected; this might reflect partly spliced nuclear RNA (total cell RNA was used for these RPA experiments) or may indicate a heterogeneity of splicing in the BART RNAs. The unspliced signal could not be derived from viral DNA contaminating the RNA because the 200 nt band in the exon I RPA corresponds to the length of correctly initiated RNA unspliced at the exon I 3’ end rather than the whole EBV content of the probe (379 nt), which would be protected by viral genomic DNA.

The C666-1 line is unusual in the sense that there has been great difficulty in obtaining an NPC cell line that retains its EBV, so we also checked that the EBV genome was in the normal episomal state and had not suffered major deletions. Southern blotting BamHI digests of C666-1 DNA (data not shown) revealed the normal BamHI fragments C, W, K and A, which are widely distributed along the genome, indicating no obvious major deletions. Gardella gel analysis (Gardella et al., 1984) showed the typical episomal EBV found also in a B95-8 lymphoblastoid cell line (LCL C) and in the Akata Burkitt’s lymphoma cell line (Fig. 2). The lymphoid lines have a small fraction of cells spontaneously in the productive cycle giving some linear EBV DNA but this was only present at a very low level in C666-1, consistent
with the latent cycle protein expression pattern reported previously (Cheung et al., 1999).

To identify likely binding sites for transcription factors within C666-1 EBV corresponding to the SK plasmid region, DMS interference in vivo footprinting was applied to the C666-1 cells. The interpretation of this type of data is sensitive to sequence variation relative to the prototype EBV-wt sequence so this region of C15, C666-1 and Akata EBV was first sequenced. A few sites of variation were detected, summarized in the legend to Fig. 3, but there was no variation from the EBV wt sequence in the sites identified by the footprinting. The footprinting showed several sites of either protection or enhanced cleavage relative to the equivalent naked plasmid DNA. These are shown in Fig. 3 and were named A (includes an AP1 consensus binding site TGGAGTCA), B (sequence similar to an Ets or Stat consensus site) and N (some similarity to an NF-κB site), the latter being downstream of exon 1. The B site showed protection, the NF-κB site had enhanced cleavage and the A site had both protection and enhanced cleavage. The positions of these sites relative to the transcription map are shown in Fig. 4(A). The footprinting was done in C666-1 cells but subsequent transfection assays for BART promoter activity (Fig. 4) were performed in 293 cells because of their higher transfection efficiency, so extracts of both 293 and C666-1 cells were tested for binding oligonucleotides containing the A, B or N site. Clear binding of the A and B sites was observed by EMSA (Fig. 3B) with both 293 and C666-1 extract and this was specific since it was competed by an excess of the same oligonucleotide but not by an oligonucleotide in which some nucleotides had been mutated (the same mutations as used below in functional assays of the promoter). A single major A site complex was observed with C666-1 extract but several complexes were seen with 293 extract (Fig. 3B). The most specific of these (arrowed), as determined by competitor oligonucleotides, migrated close to the position of the C666-1 complex. It was already well established, e.g. Kirch et al. (1999), that 293 cells contain AP1 activity which can bind to the same sequence that is present in site A. An
oligonucleotide containing the N site was also tested with extracts from 293 and C666-1 cells but no specific binding was observed in the EMSA (data not shown).

To determine more precisely the sequences required for BART promoter activity, deletions were made in the SK plasmid and site-directed mutations were made at the locations identified by the in vivo footprinting (Fig. 4A). Mutations were introduced into the A and B sites and the N site was deleted from the BK plasmid. The plasmids were transfected into 293 cells and resulting RNA was assayed by RPA for exon 1 of the BART RNAs (Fig. 4B). The results were normalized relative to an RPA assay for GAPDH (Fig. 4B) and the results quantified (Fig. 4C) from the phosphorimager data. The results show that truncation of the plasmid down to BK or CK gave about twice the amount of exon 1 RNA as plasmid SK. Mutation of either the A or B site within BK caused modest reductions in expression but mutation of both A and B sites substantially reduced expression (about 10-fold). Consistent with this, mutation of the B site reduced CK expression (the A site is not present in the CK plasmid). Deletion of the N site either by truncation in the SS plasmid or by localized deletion in BK caused only a small reduction in exon 1 expression.

DISCUSSION

The RNA mapping data shown here have further confirmed that the BART exon I starting transcription at 138350 is a significant point of initiation of BART family RNAs. The previously reported 5' end was confirmed in C666-1 cells (Fig. 2A). One surprising feature of the DNA methylation study of C15 EBV DNA reported here is that the unmethylated region of DNA extends significantly downstream of exon I. This suggests that proteins may be bound to this region during latent persistence of the virus in the C15 tumour cells, preventing DNA methylation. The downstream unmethylated region could represent components of the BART promoter, other promoters so far unmapped or other genetic functions within this region of DNA. Previous RNA mapping in B95-8 cells (Farrell, 1989) recorded poorly characterized leftward RNAs that might originate from this region and it remains unclear whether the A73 type of BART RNA (Smith et al., 2000) initiates at the normal BART exon I, so there are candidate RNAs that might come from a novel promoter in this region yet to be characterized. On the other hand, there is some evidence that could be consistent with downstream promoter elements in the BART promoter. To analyse the BART
promoter, we used RPA assays on the SK, BK and CK constructs because simple fusion of the upstream region (which would normally be expected to contain the promoter sequence) to a CAT reporter gave very little CAT activity (data not shown). Downstream promoter elements are one possible explanation of this. DMS interference footprinting in the C666-1 cell line showed enhanced cleavage at site N in the downstream region, indicating a distortion of the normal DNA structure at that point but no other protein binding was observed directly. These possibilities therefore remain to be resolved.

Although the cells grow relatively slowly, the C666-1 cell line seems to be a valuable system for studying EBV gene expression in epithelial cells. It appears to contain a relatively wild-type EBV, retains the EBV in culture and expresses the BART RNAs. DMS interference footprinting suggested Sites A and B upstream of exon I that might be involved in BART expression and mutation of both of these sites substantially reduced activity of the BART promoter in a transfection assay. Site A contains a perfect match to the AP-1 consensus site that has been shown previously to bind c-Jun/c-Fos and the mutation we introduced to site A is
known to prevent c-Jun/c-Fos from binding (Risse et al., 1989). A common specific EMSA band was obtained with both 293 and C666-1 cells so it is likely that this contributes to activity of the promoter in both cell lines. The factor that binds to the B site is less certain. Scanning the sequence with the TEMATRIX transcription factor binding site database (Heinemeyer et al., 1998) suggests imperfect matches to NRF-2 (93 %), c-Ets (87 %) and a STAT consensus, STATx (86 %). These are widely expressed factors with several family members and overlapping binding specificities; it is difficult to be certain which factors are the functional ones on the B site but we have demonstrated that there is a single major complex detected in EMSA analysis with this site that prevented activity of the promoter also prevented binding to the B site but we have demonstrated that there is a single major complex detected in EMSA analysis with this site using 293 and C666-1 cells and that the mutation of the site that prevented activity of the promoter also prevented complex formation. A factor containing NRF-2 is perhaps difficult to be certain which factors are the functional ones that prevented activity of the promoter.

NRF-2 has been purified from HeLa cells; Virbasius et al. (1992). NRF-2 was originally studied as part of GABP, a site in the rat cytochrome c oxidase gene and the binding site in the rat cytochrome c oxidase gene is almost identical to site B in the EBV BART promoter. Antibodies to NRF-2 are not available to test this directly. It has been proposed that Stat 3 is a major regulator of EBV latent cycle promoters in epithelial cells (Chen et al., 2001) based on the Qp and LMP1 promoters but we could find no evidence for Stat 3 binding to the B site or for binding of phosphorylated Stat 1 (data not shown).

These results are the first detailed analysis of sequences required for expression of the BART RNAs. They will provide an opportunity to identify cell factors that control expression of the BART RNAs and we have also shown that C666-1 will be a valuable system in which to investigate the BART genes since it contains an apparently normal episomal EBV genome and expresses the BART RNAs.

REFERENCES


