Isolation and sequencing of the Epstein-Barr virus BNLF-1 gene (LMP1) from a Chinese nasopharyngeal carcinoma

Li-Fu Hu,1† Eugene R. Zabarovsky,1† Fu Chen,1§ Shi-Long Cao,2 Ingemar Ernberg,1 George Klein1 and Gösta Winberg1*

1Department of Tumor Biology, Karolinska Institute, Box 60400, S 104 01 Stockholm, Sweden and
2Shanghai Medical University Tumour Hospital, 270 Dong An Road, Shanghai, People's Republic of China

The BamHI fragment containing the Epstein-Barr virus (EBV) LMP1 gene was cloned from a genomic library of the nude mouse-propagated Chinese nasopharyngeal carcinoma CAO. The sequence of the LMP1 gene and its promoter and enhancer was determined. The nucleotide sequence of the CAO isolate differed from those of the B95-8 and Raji isolates in the promoter/enhancer region; the amino acid sequence of the protein also differed. Structural differences in the protein were located mainly in the 20 N-terminal residues and the array of repeated amino acids in the C-terminal part of the protein, in which the CAO isolate displays a cluster of seven perfect repeats of 11 amino acids (aa). Three of these repeats have no counterpart in the other virus strains. This, together with two deletions of five and 10 aa in the C-terminal part, yields a protein of 404 aa, compared to 386 aa for B95-8 and Raji. The larger LMP1 protein was detected on immunoblots of tissue samples from the CAO nude mouse tumour, and was also present in EBV-negative B cell lines and immortalized keratinocytes transfected with the cloned gene. A XhoI restriction site in exon 1 of the B95-8 BNLF-1 gene was absent from the CAO EBV isolate, as well as from 36 of 37 Chinese NPC biopsies tested. In contrast, 17 of 19 NPC biopsies of African origin retained this XhoI site.

Introduction

Nasopharyngeal carcinoma (NPC), of all human malignancies, is associated most strongly with Epstein-Barr virus (EBV). Multiple episomal copies of the viral genome are found in all poorly differentiated or anaplastic NPCs, irrespective of ethnic origin (Zur Hausen et al., 1970; Wolff et al., 1973; Andersson-Anvret et al., 1977). In Burkitt's lymphoma (BL), the association is much weaker, with 97% of the tumours in highly endemic areas and only 20% of tumours in the sporadic form carrying EBV genomes (Rowe & Gregory, 1989). EBV genomes have not been found in squamous cell carcinomas or other tumour types in the nasopharyngeal region (Sixbey, 1989). This regular association cannot be due to chance. Most probably, the genetic machinery of EBV predisposes the epithelial precursor cell of NPC to malignant development, either alone or in conjunction with environmental or genetic cofactors. The EBV strains in areas of high NPC incidence, such as Southern China, may be functionally indistinguishable from strains isolated from B cell malignancies, in which case the cofactors are of greater importance to the understanding of the biogenesis of NPC. However, the possibility that EBV strains in areas of high NPC incidence have common functional and genetic differences which cause their different tropisms in relation to malignancy cannot be discounted.

Studies of EBV gene expression in NPC (Fähraeus et al., 1988; Young et al., 1988) reveal that, in addition to Epstein-Barr nuclear antigen (EBNA) 1, which is expressed in all EBV-carrying cells irrespective of phenotype, LMP1 is the only other transformation-associated EBV gene that is expressed in NPC tumour biopsies, and in a majority of cases. EBNAs 2 to 6 appear to be regulated in a B cell-specific fashion, and their expression has not been detected in NPC. As EBNA 1 has not been ascribed any transforming function, LMP1 is the most likely candidate for an EBV gene that participates in the transformation of the nasopharyngeal epithelium in NPC. This assumption is also supported by findings showing that LMP1 transforms established...
rodent fibroblasts in vitro (Baichwal & Sugden, 1988, 1989; Wang et al., 1985, 1988), alters the phenotype of immortalized human keratinocytes (Fähræus et al., 1990a) and inhibits their keratinous differentiation (Dawson et al., 1990).

To compare NPC isolates with B cell-derived isolates of EBV at the molecular and functional level, we isolated the BLNL-1 (LMP1) gene and its promoter region from a nude mouse-paissaged Chinese NPC, CAO (Cao et al., 1987; Fähræus et al., 1988), and compared its DNA sequence to those of the B95-8, Raji and C15 LMP1 genes (Baer et al., 1984; Hatfull et al., 1988; Gilligan et al., 1990). The B95-8 and Raji isolates show a very high degree of similarity in the BLNL-1 gene (97.5% similarity over the coding sequence, including the two introns), and therefore do not provide clues to the location of conserved and variable regions of the gene. The partial sequence of the LMP1 cDNA from the nude mouse-paissaged North African NPC, C15 (Busson et al., 1985, 1988), shows even greater similarity with B95-8 than Raji (Gilligan et al., 1990). Studies of EBV strains isolated from healthy Chinese carriers and NPC biopsies have revealed a number of alterations in different regions of the EBV genome, mainly involving repeated sequences (Lung et al., 1988, 1990; Zhang et al., 1989). Certain genome types appear to represent virus strains prevalent in Asia, particularly Southern China, such as the proposed type C strain which lacks a BamHI site between BamHI W1* and 11*, and a proposed substrain carrying, in addition, a deletion in BamHI F [strain A/C(f); Lung et al., 1990].

The experiments presented here demonstrate that CAO expresses a LMP1 protein larger than that of the B95-8 isolate. Sequence analysis of the BLNL-1 gene cloned from this tumour reveals a conserved overall structure for the CAO LMP1 gene, but with nucleotide changes producing many amino acid substitutions within the first 20 amino acids (aa) of the encoded protein, two deletions and the insertion of three additional 33 bp repeats in the C-terminal part of the protein. The additional repeats alter the size and the proposed structure of the 25K phosphorylated fragment which is cleaved off during processing of the LMP1 protein (Mann & Thorley-Lawson, 1987; Baichwal & Sugden, 1987; Moorthy & Thorley-Lawson, 1990). The enhancer and promoter sequence also contains several base substitutions, in particular in the LMP1 transcription control region between nucleotides -240 and -1, relative to the cap site. The absence of an Xhol site in the 5' exon of the CAO BLNL-1 (LMP1) gene is shown to provide a restriction fragment length polymorphism (RFLP) marker that distinguishes the BLNL-1 genes in Chinese EBV-positive NPC biopsies from those in NPC biopsies of African origin.

![Fig. 1. Map of the \(\lambda SK10\) vector. \(\lambda SK10\) is a replacement vector with a cloning capacity of 4.5 to 20 kb. It was constructed by inserting restriction site linkers into \(\lambda SK9\) (Zabarovsky et al., 1989) to allow the use of BamHI, EcoRI, Nael and Xhol for cloning. The vector can be selected by red-gum selection, and inserts may be subcloned in a plasmid by simple excision with SacI, Clal or BspMI, followed by religation at low concentration.](image)

### Methods

**Tumour material and cell lines.** CAO is an EBV-positive NPC derived from a 54-year-old male patient from the Shanghai area of China. It has subsequently been propagated as a xenograft in BALB/c nude mice (Cao et al., 1987; Fähræus et al., 1988). C15 is a North African NPC established in nude mice (a gift from T. Tursz; see Busson et al., 1988); C15/CBC is a lymphoblastoid cell line obtained by cocultivation of human cord blood lymphocytes with C15 cells (Ernberg et al., 1989). CBC-SEB-CHI-7 and CBC-SEB-CHI-20 are lymphoblastoid cell lines (LCs) established by infection of cord blood lymphocytes with EBV from a healthy Chinese donor and a Chinese NPC patient, respectively (Hu et al., 1991). C121 and C82 are NPC biopsies of North African origin (gifts from T. Tursz); SD and OM are East African NPC biopsies; N8, N11, N17 and N40 are NPC biopsies from Shanghai, China; N41, N45 and N48 are NPC biopsies from the Guangdong province in southern China (see Fig. 2b). B95-8 is a line of marmoset lymphocytes transformed with EBV type A from mononucleosis (Shope et al., 1973); Jijoye M13 is an EBV-positive BL cell line carrying an EBV type B strain (Hinuma et al., 1967); DG75 is an EBV-negative BL line (Ben-Bassat et al., 1977); BJB is an EBV-negative B cell lymphoma line (Menezes et al., 1975). Rhek-1 cells are human epidermal keratinocytes transformed by an adenovirus 12 (Ad12)-simian virus 40 (SV40) hybrid expressing SV40 large T and small t antigens, but not Ad12 E1A or E1B proteins (Rhim et al., 1985).

**DNA cloning.** DNA was extracted from the CAO tumour tissue by standard methods (Sambrook et al., 1989) and cleaved with BamHI for library construction. The phage \(\lambda\) vector SK10 (Fig. 1) was cleaved with EcoRI and BamHI to release and inactive the stuffer fragment. The short oligonucleotide fragments joining the BamHI sites on the vector arms with the EcoRI sites on the stuffer were removed by precipitating the DNA in 6.5% polyethylene glycol 8000, 0.85 m-NaCl (Lis, 1980). One microgram each of the BamHI-cleaved CAO and SK10 vector DNA was ligated at 200 \(\mu\)g/ml under standard conditions and packaged into phage \(\lambda\) using a packaging extract (Sambrook et al., 1989). The primary library contained about 10\(^{6}\) independent clones. The library was plated on NM621 (Whittaker et al., 1988) cells in \(22 \times 22\) cm dishes at a density of about 10\(^{5}\) plaques per plate, and replicated on Hybond-N filters (Amersham). After screening of the filters with the 1.9 kb Xhol fragment of B95-8, containing the coding sequence of the LMP1 gene (coordinates 167487 to 169423; Baier et al., 1984), 10 LMP1-positive clones were isolated and analysed. Clone CAO-5 was chosen for further analysis (see below).

The EBV insert in CAO-5 was subcloned by excision from the phage \(\lambda\) vector using Clal (Fig. 1). Since \(\lambda\) SK10 contains the pBR322 origin and the \(\beta\)-lactamase gene in the sequences flanking the insert, the Clal
Table 1. Location and sequence of primers

<table>
<thead>
<tr>
<th>Coordinate of 3' end</th>
<th>Primer</th>
<th>Nucleotide sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>168,373</td>
<td>5’ CTAGCGACTCTGCTGGAAAT 3’</td>
<td></td>
</tr>
<tr>
<td>168,592</td>
<td>5’ ACTCTGCTCTAAAACCTTAGC 3’</td>
<td></td>
</tr>
<tr>
<td>168,747</td>
<td>5’ TTATCATGACACGACACA 3’</td>
<td></td>
</tr>
<tr>
<td>169,037</td>
<td>5’ ATCTTTGGCTGCTTACTTGGTAAG 3’</td>
<td></td>
</tr>
<tr>
<td>169,456*</td>
<td>5’ CTTAGGGAGTGAACAGCCTTGAGA 3’</td>
<td></td>
</tr>
<tr>
<td>169,534</td>
<td>5’ TAGGCGGCCCTACATAACCCCTCT 3’</td>
<td></td>
</tr>
<tr>
<td>169,990</td>
<td>5’ AAATTCCTATATCCCCGTC 3’</td>
<td></td>
</tr>
<tr>
<td>Antisense</td>
<td></td>
<td></td>
</tr>
<tr>
<td>168,253</td>
<td>5’ AAAAGCAGCTGTTAGGAAGGGTGGA 3’</td>
<td></td>
</tr>
<tr>
<td>168,313</td>
<td>5’ CACCGTCTGTCTCATCGAAGGC 3’</td>
<td></td>
</tr>
<tr>
<td>168,392</td>
<td>5’ ATTTCCAGGAGTGCCTGAT 3’</td>
<td></td>
</tr>
<tr>
<td>168,669</td>
<td>5’ GTGCTGCGCTCTGGTGGTT 3’</td>
<td></td>
</tr>
<tr>
<td>168,495*</td>
<td>5’ GGCAGAGAATCTCCAAGATG 3’</td>
<td></td>
</tr>
<tr>
<td>169,103</td>
<td>5’ GTCACAATGGAGTGCCTTCGGCAGAACCC 3’</td>
<td></td>
</tr>
<tr>
<td>169,310</td>
<td>5’ GCTAGGAGCGAAGGAAGATG 3’</td>
<td></td>
</tr>
<tr>
<td>169,555</td>
<td>5’ AGAGGTTATATGAGGCGCCCTA 3’</td>
<td></td>
</tr>
<tr>
<td>169,697</td>
<td>5’ CACGGTGGTTAGGGTGCTGTGG 3’</td>
<td></td>
</tr>
<tr>
<td>170,028</td>
<td>5’ CAGAATGAGGGTGCCGATT 3’</td>
<td></td>
</tr>
<tr>
<td>170,068</td>
<td>5’ GTGCGGAGTGTCTGATGGA 3’</td>
<td></td>
</tr>
</tbody>
</table>

* Primers made from the B95-8 DNA sequence (Baer et al., 1984). All other primers were made from the sequence of EBV strain CAO.

The LMP1 gene and promoter/enhancer sequences, contained in the SacI-BamHI fragment corresponding to B95-8 coordinates 170,153 to 166,614, were subcloned (after adding a BamHI linker to the SacI site) into the BamHI site of the J124-A8 vector (a gift from L. Rymo), producing the construct J124-CAO-5. The J124-A8 vector is a pSV2-gpt-based eukaryotic expression vector (Mulligan & Berg, 1981) with a second SV40 origin and a BamHI cloning site placed between the end of the gpt gene and the SV40 poly(A) site (Fig. 2). The J132-G5 plasmid (a gift from L. Rymo) is isogenic with the J124-CAO-5 construct, but carries the B95-8 LMP1 gene.

DNA sequencing. Double-stranded plasmids were sequenced (Chen & Seeburg, 1985; Lim & Pene, 1988) using the dideoxynucleotide chain termination method (Sanger et al., 1977) adapted for the use of phage T7 DNA polymerase (Sequenase; USB). Sequencing primers were made from the sequence of EBV strain CAO. All other primers were synthesized on an Applied Biosystems 381A synthesizer, or purchased as primers made from the B95-8 DNA sequence (Baer et al., 1984; Seeburg, 1985; Lim & Pene, 1988) using the dideoxynucleotide chain termination method.

Table 1 shows the sequence and location of each primer used in the experiments. The sequence of the CAO and B95-8 LMP1 genes in viral episomes in the B95-8 cell line and the CAO nude mouse tumour (Fig. 3) was performed in a similar fashion, but cleavage products were separated on a 1.5% agarose gel and the filter was probed with the 1671 bp Smal fragment, containing the entire B95-8 LMP1 coding sequence (coordinates 169,591 to 167920). It should be noted that DNA fragments larger than about 2 kb cannot be reliably determined in this experiment owing to the high agarose concentration used to separate the fragments of small size, which were of interest.

Protein analysis by immunoblotting. Cultured cells were dissolved at a ratio of 10^7 cells/ml in RIPA buffer (150 mM-NaCl, 1% NP40, 0.5% deoxycholate, 0.1% SDS, 50 mM-Tris-Cl pH 7.5). Tumour material was ground to a powder in liquid nitrogen, solubilized in an appropriate volume of RIPA buffer and the proteins were separated on SDS-acrylamide gels (Laemmli, 1970). Electrophoretic transfer of the protein to nitrocellulose filters (Towbin et al., 1979) and immunoblotting with the anti-LMP1 monoclonal antibody S12 (Mann et al., 1985) were essentially as described (Fähraeus et al., 1988).

Transfection of DNA into human cells. Transient expression of the B95-8 and the CAO LMP1 genes from the expression plasmids J132-G5 and J124-CAO-5, respectively, was achieved by electroporation using the Gene Pulser apparatus (Bio-Rad). The conditions were optimized for each cell line using a chloramphenicol acetyltransferase reporter system. For DG75 cells, 250 V, 960 µF resulted in a pulse of 22 to 27 ms, and for BJAB cells, 300 V, 960 µF (22 to 27 ms) was used; for Rk-h1 cells, 250 V, 960 µF was used. Cells were harvested in log phase, concentrated to 6 x 10^6 cells/ml in Optimem (Gibco) supplemented with penicillin, streptomycin and glutamine, but without serum. Uncleaved plasmid DNA (10 µg) was added to 0.8 ml of cells and the cell suspension was left on ice for 5 to 10 min before electroporation. Immediately after the shock, the cell suspension was stirred briefly with a Pasteur pipette to even out pH gradient effects in the cuvette. The cells were then left on ice for 10 min before plating in 10 ml Iscove's modification of Dulbecco's MEM with 10% foetal calf serum. Cells were harvested for protein analysis 40 h post-transfection.

Results

XhoI restriction site polymorphism in the LMP1 gene of Chinese NPC tumours

Fig. 2 shows XhoI site polymorphism in a number of samples of African or Chinese origin. The presence of the 1.9 kb XhoI fragment (Fig. 3) characteristic of the B95-8 prototype strain was evident in all the African samples, but was replaced in the Chinese samples with larger fragments ranging in size between 6 and 11 to 12 kb (a).

Of a total of 19 NPC biopsies of North or East African origin and 37 Chinese NPC biopsies tested, Fig. 2 includes two NPC biopsies of North African origin (C82 and C121), one nude mouse-propagated North African NPC (C15) and two NPC biopsies of East African origin (SD and OM), as well as seven Chinese NPC biopsies (N8, N11, N17, N40, N41, N45 and N48) and one nude mouse-propagated NPC, CAO. The classification of these materials according to EBV (EBNA 2) serotype A or B and LMP expression is shown in Fig. 2(b). There was only one exception to the difference in XhoI polymorphism between African and Chinese samples shown in Fig. 2. Of the 37 Chinese biopsies, 32 were type...
A and five were type B. Three of the Chinese type B-carrying NPCs and three CBC lines established from healthy donors (two samples), or a NPC patient (one sample) were tested for \textit{XhoI} polymorphism in LMP1 and were found to lack the \textit{XhoI} site (Table 2).

Thus, EBV strains prevalent in China frequently carry a rearrangement affecting the sequence of the \textit{XhoI} sites flanking the 1.9 kb \textit{XhoI} fragment detected in B95-8 using an LMP1 probe. The mapping data indicate that this, in all probability, is the \textit{XhoI} site at the 5' end of the coding sequence of LMP1, since each tumour has a \textit{XhoI} fragment of unique size, indicating that the variable Nhet region at the right end of the viral genome is contained in the fragment (Fig. 3). Such \textit{in vivo} variation in the size of the Nhet repeat region has been observed previously as a distinctive characteristic of the monoclonal origin of individual NPC tumours (Raab-Traub & Flynn, 1986; Lung et al., 1990).

\textbf{Restriction mapping of the \textit{BamHI} Nhet fragment in the phage \textit{\lambda} clone CAO-5}

Fig. 3 shows a comparison of the restriction maps for the B95-8 (a) and CAO (b) \textit{BamHI} Nhet fragments using a number of restriction enzymes. The map shows that the 5' \textit{XhoI} site in the coding region of the LMP1 gene is absent from the CAO isolate, and also demonstrates the loss of the two internal \textit{NcoI} sites present in the LMP1 gene of the B95-8 isolate. The Nhet region (\textit{EcoRI} to \textit{SacII}) in the \textit{\lambda} SK10 CAO-5 clone shown in Fig. 3 was 1.4 kb in length. This corresponds to the size of the Nhet fragment in the original nude mouse tumour CAO, but is 0.7 kb shorter than that in the B95-8 prototype strain.

A comparison of 10 different \textit{BamHI} Nhet clones derived from the SK10 library of the CAO tumour also revealed a size variation which could not be accounted for by variation in the DNA from the tumour. From this, and our subsequent mapping data, we drew the conclusion that rearrangements had occurred in the Nhet region of the insert in the phage \textit{\lambda} vector, perhaps as a result of \textit{recA}-mediated recombination between reiterated sequences in Nhet. For instance, the Nhet fragment in another mapped clone, SK10-CAO-6, was reduced to 0.4 kb in length (Fig. 3). For further analysis, mapping with different restriction enzymes and expression studies, the phage \textit{\lambda} CAO-5 clone was chosen.

\textbf{Comparison between the DNA sequences of the CAO, B95-8, Raji and C15 LMP1 genes, and their promoter and enhancer sequences}

In Fig. 4, the DNA sequences for the LMP1 gene from the four EBV isolates have been aligned to show changes in the sequence. The 2038 bp CAO sequence is 95\% identical to the B95-8 and 94\% identical to the Raji LMP1 sequences. Divergent sequences appear to be clustered around the proposed binding site for NF-1 at nucleotides -240 to -230 from the transcription start, and from the putative Oct-1-binding site at -140 to -100 (Ghosh & Kieff, 1990). More extensive rearrangements are located in the 3' region of the coding sequence, where the imperfect system of repeats in the B95-8 and Raji strains has been replaced in the CAO isolate with a perfect tandem array of seven repeats of 33 bp. This has been accomplished partly by a deletion of 15 bp, reducing the size of the second repeat from 48 bp in B95-8 to 33 bp. Three of the repeats in the CAO LMP1 have no counterpart in the other isolates. The loss of the
Fig. 3. Comparison of the restriction maps of the BamHI Nhet fragments of the B95-8 and CAO isolates. The restriction map of the B95-8 BamHI Nhet fragment and the nucleotide numbers indicated on the map are from the published sequence (Baer et al., 1984). The corresponding map for the CAO isolate was derived by mapping the CAO-5 insert in J SK10 with single and double cleavages using the restriction enzymes shown. For the LMP1 gene (unfilled bars) and part of the EDL1 promoter/enhancer region, the restriction sites were verified by sequencing. The location of the EDL1 promoter is indicated with a flag. The CAO-5 clone, like the CAO tumour DNA, has a Nhet fragment of 1.4 kb, whereas clone CAO-6 has a fragment of 0.4 kb. The restriction map of the J124-A8 eukaryotic expression plasmid (from L. Rymo) is shown with the LMP1 gene cloned into the BamHI cloning site following the second SV40 early promoter.

Abbreviations for restriction enzymes: B, BamHI; E, EcoRI; G, BglII; H, HindIII; M, MluI; N, NcoI; S, SmaI; Sa, SacI; X, XhoI.

Table 2. XhoI polymorphism of the BRLF-1 coding region in NPC tumour biopsies according to geographic origin and EBV (EBNA 2) type

<table>
<thead>
<tr>
<th>Origin/EBNA 2 type</th>
<th>Number of biopsies</th>
<th>LMP1 expression*</th>
<th>1.9 kb XhoI fragment†</th>
<th>8 to 12 kb XhoI fragment†</th>
</tr>
</thead>
<tbody>
<tr>
<td>African/2A</td>
<td>19</td>
<td>ND†</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Chinese</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guangdong§/2A</td>
<td>10</td>
<td>6/9</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Guangdong§/2B</td>
<td>1</td>
<td>1</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Shanghai§/2A</td>
<td>22</td>
<td>11/17</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td>Shanghai§/2B</td>
<td>4</td>
<td>2/3</td>
<td>0</td>
<td>3/3</td>
</tr>
<tr>
<td>Cord blood cell line (NPC pat.)/2A</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Cord blood cell line (healthy donor)/2A</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

* Tested by immunoblotting.
† Number of samples expressing the fragment indicated.
‡ ND, Not determined.
§ Samples obtained from Zhaun San Medical College Tumour Hospital, Guangdong Province, People's Republic of China.
‖ Samples obtained from Shanghai Medical University Tumour Hospital, Shanghai, People's Republic of China. Two of three EBNA 2 B type-carrying NPC tumours tested expressed LMP1.

XhoI site (Fig. 3) is caused by a point mutation altering the recognition site from CTCGAG to CTCTAG. Both the internal NcoI sites in the LMP1 gene have been eliminated in the CAO isolate, one by a point mutation and the other by a 30 bp deletion beginning 122 bp 3' of the seven repeats. A new NcoI site is created in the methionine codon of the LMP1 gene. The existence of this site in CAO viral episomes in the nude mouse tumour cells was verified by restriction mapping and Southern blotting (not shown). Like the Raji isolate,
CAO lacks the initiation codon for the proposed late 28K gene product at the beginning of exon 3 (Hatfull et al., 1988). To verify that the structural alterations predicted from the DNA sequence of the cloned LMP1 gene (Fig. 4) reflect the structure of the gene carried on the EBV episomes in the CAO nude mouse tumour, we used membrane filter hybridization to compare a number of restriction cleavage sites predicted from the DNA sequence of the cloned LMP1 gene (Fig. 5). The creation of a new NcoI site in the methionine codon (coordinate 169475) simultaneously removes a Bsu36I site at coordinate 169479. A Bsu36I site is also present at position 168481 in B95-8, which falls within the deleted 15 bp segment in the second 33 bp repeat of CAO LMP1. Fig. 5 demonstrates the loss of the internal, 998 bp Bsu36I fragment of B95-8 and shows the appearance of a single fused fragment, larger than the two flanking fragments of the B95-8 LMP1 gene. Thus, both these Bsu36I sites were removed from the CAO viral DNA. Likewise, the internal, 495 bp NcoI fragment was not present in CAO DNA, nor are either of two fragments of 713 or 1208 bp, which would result from cleavage at the NcoI site mapped previously in the methionine codon and at one of the two internal sites. Therefore the predicted point mutation in the first NcoI site and the 30 bp deletion near the C terminus produce the expected polymorphisms in CAO tumour DNA. The 30 bp deletion is also predicted to remove a BglI site at position 168262 of the CAO sequence; another site at 169445 has been marked with + A on the C15 sequence line. Deleted nucleotides or sequence segments have been marked with dots. The 21 and 33 bp repeat segments are indicated by [21] or [33].

Fig. 4. Comparison of DNA sequences of EBV strains CAO, B95-8, Raji and C15. The start of each sequence is indicated by > and the end of the sequence by <. The locations of introns not existing in C15 cDNA have been indicated by <.....>. The positions of the EDL1 CAAT and TATA boxes, and the putative EDL1a TATA box have been overlined. The position of the ATG codons for the LMP1 gene at 169474, the putative lytic 'LMP' protein at 168936 and the new ATG in CAO at 168789 have the superscript Met. The nucleotides differing from the CAO sequence are indicated for the other three EBV strains. The extra A residue in the EDL1 TATA box sequence segments have been marked with dots. The 21 and 33 bp repeat segments are indicated by [21] or [33].
Fig. 5. Verification of structural features of the LMP1 gene in episomal EBV DNA from CAO nude mouse tumour cells. Paired samples represent the restriction fragments obtained after cleavage of B95-8 (B) and CAO (C) cell DNA with BglI (lanes 1), Bsu36I (lanes 2), NcoI (lanes 3), ApaI and BglII (lanes 4), and ApaI and StyI (lanes 5). The filter was probed with the 1671 bp SmaI fragment of B95-8, containing the coding sequence of LMP1. The BglI, Bsu36I and NcoI cleavages have been included to demonstrate loss of restriction sites due to point mutations or deletions predicted from the DNA sequence of the cloned CAO LMP1 gene. The double cleavages (ApaI and BglII, and ApaI and StyI) demonstrate an increase in the size of one fragment due to the increase in the repeat structure in the C-terminal part of the CAO LMP1-gene.

and the production, instead, of a single larger fragment is detected using the 1671 bp SmaI probe, of indicating that both the BglI sites are lost.

Measurements of the size of the repeat sequence are shown in Fig. 5, using a combination of either BglII (position 169037) and ApaI (relevant position 168314), or StyI (relevant position 168597) and ApaI. The shift in the size of fragments detected with both enzyme combinations between the B95-8 and CAO repeat segments corresponds closely to that predicted by the DNA sequences (BglII and ApaI, 723 to 807 bp; and StyI to ApaI, 283 to 367 bp for B95-8 and CAO respectively).

Comparison of the amino acid sequences of the LMP1 protein of B95-8, CAO and Raji

The overall size of the CAO LMP1 protein is 404 aa, compared to 386 aa for the other strains. Fig. 6 shows the alignment of the amino acid sequences for the LMP1 proteins of the three virus strains. If the gross rearrangements in the C-terminal part of the protein are not taken into account, the values for amino acid similarity can be calculated on the basis of net shared sequences, which amount to 371 aa. Therefore, the amino acid similarity between B95-8 and CAO is 92%, and between Raji and CAO 91%, whereas B95-8 is 95% similar to Raji. The most significant divergence in sequence is in the first

Fig. 6. Comparison of the amino acid sequences of EBV strains B95-8, CAO and Raji. Amino acid residues are represented in the single letter code. Only those residues of the sequences of the CAO and Raji strains of EBV which differ from that of the B95-8 strain are indicated. The main structural features of the LMP1 protein have been marked as subscripts: transmembrane segments are labelled m, segments exposed on the cell surface x; the proteolytic cleavage site for post-translational processing of the LMP1 protein (Moorthy & Thorley-Lawson, 1990) has been marked >>, and the seven, 11 or 16 aa repeat sequences are indicated by /7, /11 or /16; deletions are marked by an asterisk.
Expression of the CAO LMP1 gene in human lymphoid cell lines and immortalized keratinocytes

Expression vector clones of the B95-8 and CAO-5 LMP1 genes under control of the EDL1 (B95-8 coordinates 170153 to 166614) and the SV40 promoters (Fig. 3) were electroporated into DG75, BJAB and Rhek-1 cells as described above. After 40 h, cell lysates were prepared for immunoblotting to determine whether the expression clones were capable of directing the expression of a serologically reactive protein of the anticipated size. The immunoblots in Fig. 7 demonstrate that the correctly sized protein products were expressed in all the transfected cell lines. The J124-CAO-5 plasmid produced a larger protein (66K) than the isogenic construct J132-G5 expressing the B95-8 LMP1 gene (63K; Fig. 7). This agrees with the larger size predicted from the amino acid sequence, and is consistent with the size of the protein detected in the CAO tumour. Although the constructs were essentially identical, and the conditions of electroporation were standardized, the CAO LMP1 gene was consistently expressed at a higher level than the B95-8 construct.

Discussion

We have cloned and sequenced the BNLF-1 gene (LMP1) from a Chinese NPC to establish whether it shows major differences compared to previously sequenced, B cell-derived EBV strains. We have found that the LMP1 gene of the CAO tumour differs considerably from the corresponding gene in the B95-8, Raji and C15 isolates.

The XhoI polymorphism detected in the LMP1 gene is also present in the EBV genomes associated with 36 of 37 NPC biopsies from the Shanghai area and the Guangdong province in Southern China, but were not found in 17 of the 19 NPC biopsies of North and East African origin studied. This indicates that RFLP markers may be of use for studies of the epidemiology of NPC in relation to EBV strain variation. Taken together with our analysis of the EBNA 2 type of the biopsies (Table 2) it also indicates that, in all probability, the EBV genome present in the CAO tumour represents a substrain of EBV type A which is widespread in China, since practically all our Chinese biopsies show the same RFLP pattern with respect to the LMP1 XhoI site. Furthermore, the LCLs established from two healthy donors and one NPC patient from China (Table 2) were identical to the majority of the Chinese NPC biopsies with regard to the XhoI polymorphism. It is puzzling, however, that the three type B biopsies analysed also do not contain the LMP1 XhoI site.

The relevance of the LMP1 gene to NPC relies on the phenotypic effects observed in transfected human keratinocytes. Not only does LMP1 arrest the keratinous differentiation of epithelial cells in culture (Dawson et al., 1990), but transfectants show morphological changes...
perhaps related to an observed reduction in the synthesis of certain cytokeratins (Dawson et al., 1990; Fähræus et al., 1990a). Further evidence for an effect of the LMP1 gene on keratin expression is provided by the hyper-expression of keratin 6 detected in epidermis and tongue tissue of transgenic mice expressing the LMP1 gene (Wilson et al., 1990). LMP1 is also the only growth transformation-related EBV gene which is expressed, besides EBNA 1, in NPC biopsies (Fähræus et al., 1988; Young et al., 1988), and has been implicated as a transforming gene by its ability to transform rodent fibroblasts to anchorage-independent growth and tumorigenicity in nude mice (Baichwal & Sugden, 1988; Wang et al., 1985, 1988).

The LMP1 gene isolated from the Chinese NPC CAO shares basic structural features with the LMP1 genes of sequenced B95-8, Raji and C15 strains (Baer et al., 1984; Hatfull et al., 1988; Gilligan et al., 1990). It shows several structural changes that may have arisen as a result of random genetic drift, but which may also have functional significance. The latter is suggested by the base substitutions clustered in the promoter/enhancer part of the LMP1 cistron (Fig. 4). These may influence the binding of transcription factors (Ghosh & Kieff, 1990) or may affect the function of the cell-specific negative regulatory element mapped between nucleotides −106 and −54 (Fähræus et al., 1990b). Our preliminary studies of the methylation state of this promoter/enhancer region in Chinese NPC biopsies, using HpaII and MspI cleavage, also confirms the addition of new sites at nucleotides − 126 and − 517, and the loss of the B95-8 site at − 207, which indicates that regulation of the EDL1 promoter activity by methylation may be altered in the CAO strain.

The peculiar distribution of amino acid substitutions found in the short intracytoplasmic N terminus of the CAO LMP1 protein demonstrates that this domain is not conserved in all EBV strains (Fig. 6). Transformation studies with LMP deletion mutants suggest, however, that the N-terminal domain is important for the transforming activity of the B95-8 LMP1 gene (Baichwal & Sugden, 1989; Hammerschmidt et al., 1989).

The structure of the seven perfect direct repeats of 33 bp, which are preceded by a 21 bp sequence representing a subset of the 33 bp repeat unit, is remarkable. Size differences in the LMP1 protein have been noted previously in different cell lines derived from BL. Of particular interest is the observation that the BL line WeWak-1 expresses an LMP1 protein of significantly lower molecular size than that of B95-8, and that this protein appears to lack some or all of the antigenic determinants of the amino acid sequence HDPLPQDPDN (aa 276 to 285; Fig. 6), because a peptide antibody directed against this repeat unit fails to detect the WeWak-1 LMP1 protein (Rowe et al., 1986).

The CAO isolate lacks HDLP of this decapeptide. The 33 bp repeat region of LMP1 thus appears to be a target for rearrangement in vivo. Apart from the EBNA 1 protein of EBV, the glycine-alanine-rich repeat of which also varies in size between isolates, other examples of viral proteins with variable repeat regions include the glycoprotein C homologue of varicella-zoster virus (Kinchington et al., 1986), which has a variable copy number of a 14 aa repeat, and the 39K surface protein of fowlpox virus (Binns et al., 1990), in which the 12 aa repeat contains dominant immunogenic epitopes.

We believe that the extended repeat structure in the coding sequence of the CAO LMP1 gene, as well as the two C-terminal deletions and the polymorphic restriction sites found in the DNA sequence, faithfully reflect the structure of the original gene in the CAO tumour. First, a recA-mediated rearrangement would be expected to yield deletions like those observed in the Nhet repeats in our phage λ clones. In addition, a direct comparison of polymorphic restriction sites and RFLPs between EBV DNAs extracted from the B95-8 and CAO tumour cells (Fig. 5) confirm the existence in viral episomes of the original CAO nude mouse tumour of the structural alterations predicted from the DNA sequence of the cloned gene. Furthermore, despite the deletions and insertions found in the sequence of the CAO LMP1 gene, the open reading frame is preserved, and expression of the cloned and sequenced gene in two B cell lines, DG-75 and BJA-B, and in a human keratinocyte line, Rhek-1 (Fig. 7), produces an LMP1 protein with electrophoretic mobility identical to that detected in the CAO tumour material. Preliminary functional comparison of the CAO and the B95-8 LMP1 genes (L.-F. Hu, F. Chen, I. Ernberg, G. Klein & G. Winberg, unpublished results) shows that expression of either gene renders human Rhek-1 keratinocytes capable of anchorage-independent growth but demonstrates, in addition, that although the untransfected Rhek-1 cells, vector control clones and B95-8 LMP1-transfected Rhek-1 cells (nine clones tested) fail to produce tumours in SCID mice, nine of 24 clones of Rhek-1 cells transfected with the CAO LMP1 gene produce tumours, some of which grow invasively into surrounding tissue. These tumours express the CAO LMP1 gene and display human histocompatibility markers.

We wish to thank Mr Keith Andersson for excellent assistance with DNA sequencing. E.R.Z. and G.W. were recipients of stipends from The Cancer Research Institute, Concern Foundation, L.-F.H. is a recipient of a visiting scientist fellowship from the Swedish Cancer Society, and is also supported by the National Foundation for Natural Sciences and the National Oncogene Research Centre of China. This investigation was also supported by the Swedish Cancer Society, the Cancer Research Institute, Concern Foundation and by PHS grant 2ROI CA30264-10 awarded by the National Cancer Institute, DHHS.
References


(Received 30 April 1991; Accepted 17 June 1991)