The role of repetitive DNA sequences in the size variation of Epstein–Barr virus (EBV) nuclear antigens, and the identification of different EBV isolates using RFLP and PCR analysis

K. Falk, J. W. Gratama, M. Rowe, J. Z. Zou, F. Khanim, L. S. Young, M. A. P. Oosterveer and I. Ernberg

1 Microbiology and Tumorbiology Center, Karolinska Institute, S-171 77 Stockholm, Sweden, 2 Department of Clinical and Tumorimmunology, Daniel den Hoed Center, Rotterdam, The Netherlands and 3 Department of Cancer Studies, CRC Laboratories, University of Birmingham, UK

The six Epstein–Barr virus (EBV) nuclear antigen proteins (EBNA-1–6) show characteristic size variations between different virus isolates; this is a feature that has been used to identify the source of virus isolates in epidemiological studies (Ebnotyping). We have now studied the correlation between restriction fragment length polymorphisms (RFLPs) within exons coding for the EBNA3s and the molecular masses of the respective proteins. The B95-8 EBV strain was used as the prototype virus. The variation in apparent molecular mass of EBNA-1, -3 and -6 correlated positively with the size of RFLP coding for repeat sequences in these polypeptides. For EBNA-2, no correlation between apparent molecular mass and length of the repetitive sequences was found. The EBNA-4 protein showed virtually no variation in apparent molecular mass and RFLP size across the repeat sequence. Based on the strong correlation between apparent molecular mass and RFLP size for EBNA-6, we developed an EBNA-6 PCR assay that discriminated between different isolates of EBV. This assay offers the advantage of EBV characterization using uncultured material (e.g. throat washings, blood or biopsies), thus avoiding the selection against poorly transforming strains that occurs during establishment of lymphoblastoid cell lines required for Ebnotyping at the protein level.

Introduction

Initially two major subtypes of Epstein–Barr virus (EBV) were described, type A and type B (Sixbey et al., 1989; Zimber et al., 1986). They differ primarily in the coding sequence for the EBV nuclear antigen (EBNA)-2, but EBNA-3 (3a), -4 (3b) and -6 (3c) also show sequence divergence between the two types (Rowe et al., 1985, 1989). Both types have a world-wide distribution (Young et al., 1987; Zimber et al., 1986), and occasionally patients, particularly immunosuppressed persons, have been shown to carry both types (Kyaw et al., 1993; Sculley et al., 1990).

The molecular masses of EBNA-1, -2, -3 and -6 vary considerably between virus isolates. This variation has been employed by us and others to distinguish EBV isolates and variants from single carriers (Ebnotyping: Ernberg et al., 1989; Gratama et al., 1988; Sculley et al., 1987). The Ebnotyping method has been used to follow the spread of the virus between individuals and the transfer of virus with a transplant after bone marrow transplantation (Ernberg et al., 1989; Gratama et al., 1988, 1990, 1992). As a rule we detected a single Ebnotype in each virus carrier. However, in 15% of healthy donors one or sometimes multiple minority Ebnotypes were detected (Gratama et al., 1994). In families one predominant variant was found. Within a small geographical area, however, unrelated individuals show a wide distribution of variants (Ernberg et al., 1986; Gratama et al., 1990; Yao et al., 1991). In contrast, the coexistence of multiple Ebnotypes was frequently detected in recipients of allogenic bone marrow or organ transplantation (Gratama et al., 1994).

The Ebnotyping assay has been performed by immunoblot analysis of EBV proteins in lymphoblastoid cell lines (LCL) carrying different virus isolates. These LCLs could either be obtained by spontaneous outgrowth from peripheral blood lymphocytes or by in vitro transformation of cord blood lymphocytes with 'mouth wash' from the patient. LCLs express EBNA-1–6, latent membrane protein (LMP)-1, -2a (LMP-2a; terminal protein 1, TP 1) and -2b (LMP-2b; TP 2; Rowe et al., 1986, 1987, 1989; Rowe & Gregory, 1989; Zimber-Strobl
Fig. 1. (a) Schematic diagram of the EBNA-1 coding region in B95-8 prototype EBV. The BKRF1 reading frame and the Gly–Ala repeat are indicated. (b) Positions of the \textit{MspI} sites in the BKRF1 region, and the expected fragment size when the DNA is digested with \textit{MspI} and probed with E1 (coordinates 108159–108933). (c) Map of the EBNA-2 coding region. The filled boxes represent two different repeats within this region: $14 \times 9$ bp (coordinates 48678–48800) is a polyproline repeat and the $9 \times 6$ bp (coordinates 49525–49578) is a Gly–Arg repeat. (d) The two different probes used to analyse the EBNA-2 region are indicated: E2Y (coordinates 47778–48848) and E2H (coordinates 48848–49900). The expected fragment sizes when the DNA is digested with \textit{MspI} are shown. (e) Map of the EBNA-3, -4 and -6 coding regions of the prototype B95-8 EBV. Hatched boxes indicate the different reading frames and the filled boxes the repeats. EBNA-3 has been mapped to the reading frames BLRF3 and BERF1; EBNA-4 maps to BERF2A and BERF2B; EBNA-6 maps to BERF3 and BERF4. The coding exons for EBNA-3 contain one large repetitive sequence. EBNA-4 also contains one repetitive region, while EBNA-6 has two repeats. (f) Horizontal bars show the DNA probes which were used for the EBNA-3, -4 and -6 regions: \textit{BamHI} E (coordinates 92703–100613), E3 (coordinates 93162–95239), E4 (coordinates 97283–97693) and E6 (coordinates 99759–100613). The cleavage sites for \textit{PvuII}, \textit{KpnI}, \textit{EcoRI} and \textit{BgIII} are indicated. The numbers above the lines indicate the sizes of larger fragments from B95-8 DNA when digested with the different restriction enzymes.
All the EBNA5 apart from EBNA-5 are suitable for Enotyping. EBNA-5 varies due to size variation of the large internal repeat 1 (IR 1), but also, and in contrast to the other EBNAs, due to alternative splicing over the coding exons. EBNA-1 varies between 65–97 kDa, EBNA-2 between 75–105 kDa, EBNA-3 and -6 between 130–195 kDa, while EBNA-4 varies less, between 145–160 kDa (Dillner & Kallin, 1988). EBNA-4 has little value for Enotyping due to its limited variation and high molecular mass range, where immunoblot resolution is low.

All EBNA proteins contain repetitive amino acid sequences. EBNA-1 is encoded by the BKRFl exon in the BamHI K fragment which contains a repeat structure, coding for only glycines and alanines (Hennessy & Kieff, 1983; internal repeat 3, IR 3). EBNA-2 is encoded by the BYRF1 exon in the BamHI Y and H region (Hennessy & Kieff, 1985; Mueller-Lantzsch et al., 1985). This exon includes two repetitive elements; one polyproline repeat and one short repeat coding for only glycines and arginines. EBNA-3, -4 and -6 are derived from a family of related genes, each with two exons, within the BamHI E fragment. EBNA-3 has been mapped to the reading frames BLRF3 and BERF1, EBNA-4 maps to BERF2A and BERF2B and EBNA-6 maps to BERF3 and BERF4 (Ricksten et al., 1988; see Fig. 1a, c, e). The coding exons for EBNA-3 contain one large repetitive sequence. EBNA-4 also contains one repetitive region, while EBNA-6 has two repeats.

The repetitive sequences may be more prone to recombination events, and therefore most of the protein size variation may be ascribed to these regions. A positive correlation between polypeptide size and variation of the Gly–Ala repeat has previously been demonstrated for EBNA-1 (Hennessy et al., 1983).

We have established the correlation between apparent molecular mass and size of the repeat fragments in the coding exons for EBNA-1, -3 and -6, and the potential usefulness of restriction length polymorphism (RFLP) for Enotyping based on polymerase chain reaction (PCR).

**Methods**

**Cell lines and tissue culture.** Cell lines were propagated in RPMI 1640 (Gibco), supplemented with 10% heat-inactivated fetal calf serum (FCS), 200 μg streptomycin/ml and 60 μg penicillin/ml. They were kept in 5% CO2 with humidified air at 37 °C. For our purpose LCLs were established from the peripheral blood of bone marrow transplant patients and healthy donors by spontaneous outgrowth (Gratama et al., 1990) or by transformation of cord blood cells with EBV in mouth washes (Ernberg & Andersson, 1986).

B95-8 is a marmoset cell line carrying infectious mononucleosis (IM) derived EBV (Miller & Lipman, 1975). Jijoye is a Burkitt’s lymphoma cell line carrying an EBV type B virus strain (Adlindsay et al., 1985). K 562, Bjab and Ramos (Lozzio & Lozzio, 1975; Menezes et al., 1973; Klein et al., 1975) were used as EBV-negative controls in Southern blotting, immunoblotting and PCR.

**Immunoblotting.** Immunoblotting was performed as previously described (Ernberg et al., 1986; Gratama et al., 1988; Rowe & Gregory, 1989). The following primary antibodies were used. Serum from a patient with acute lymphoblastic leukaemia (JB) was used for the detection of EBNA-1, -2, -3, -4 and -6. For detection of EBNA-1 a human serum (JF) pre-absorbed with Ramos was used (Dillner & Kallin, 1988; Gratama et al., 1994). The mouse monoclonal antibody PE-2 (Young et al., 1987) was used for detection of EBNA-2. A serum from a patient with rheumatoid arthritis, RS22, containing antibodies to EBNA-3, EBNA-4 and EBNA-6 in high titres was processed by affinity purification. Monospecific antibodies were obtained by affinity chromatography using immobilized recombinant fusion proteins (EBNA-3, EBNA-4 and EBNA-6; Rowe et al., 1989). The serum MSt, derived from a healthy donor contained high-titred EBNA-6 antibodies.

**Southern blotting.** DNA was extracted according to standard procedures (Sambrook et al., 1989). For each μg of cellular/viral DNA 10 μl of restriction enzyme was added for endonuclease digestion. Southern blotting was performed as described previously (Falk & Ernberg, 1993). The digested DNA was separated in 0.8 or 1.5% agarose gels (BRl) in Tris-borate-EDTA buffer (TBE; 89 mM-Tris, 89 mM-boric acid, 2 mM-EDTA). Electrophoresis was carried out at 50 V (Sambrook et al., 1989). The probes were labelled with [32P]dCTP (Amersham) by random primer oligolabelling with the Klenow fragment of Escherichia coli DNA polymerase (Feinberg & Vogelstein, 1983). Hybridization with specific probes was performed at 42 °C for 16 to 24 h in 6 × SSPE, 100 μg herring sperm DNA, 50% formamide, 1% dextran sulphate and the radioactive probe. The filters were exposed to Fuji medical X-ray film for 12–96 h at −90 °C and then developed.

The following probes were used (see also Fig. 1): BamHI E fragment coordinates 92703–100613, an EBNA-1 MspI–MspI subfragment (E1) which covers the Gly–Ala repeat coding region (coordinates 108159–108933); for EBNA-2 a subfragment of the BamHI Y fragment (E2Y), covering the proline repeat (47778–48848), and another probe (E2H) (coordinates 48848–49090) in the BamHI H region covering the Gly–Arg repeat; for EBNA-3 a probe (E3) covering the ABCBA-CADD repeats (coordinates 93162–95239); for EBNA-4 a probe (E4) covering the 3 × 60 bp repetitive element (coordinates 97283–97693); and for EBNA-6 a probe (coordinates 99759–100613; E6), close to the 3’ end of the BERF4 exon, covering the 10 × 15 bp repeat, but not the 3 × 39 bp repeat in B95-8.

The PCR products (see below) were separated on 1.5% agarose gels and a standard protocol for hybridization was used. The hybridization was carried out at 50 °C using end-labelled oligonucleotide probes. Two different oligonucleotides were used for detection of amplified fragments. For type A and B a 20 nt (5’ CAACCCGGGTATGAGG-3') probe was used, coordinates 100599–100618. A type B-specific probe (5’ TCGGCTCAATGGGAGCCACA 3') was also used, corresponding to the same region in the AG 876 isolate.

**Amplification of DNA by PCR.** For PCR the DNA was prepared either according to Albert & Fenyö (1990) or the DNA extracted for Southern blotting was used (for isolates D46, D49 and D108). Ten μl (0.3 μg) of the DNA preparations was used as template for PCR. The reactions were performed in 100 μl containing 10 mM-Tris (pH 8.3), 50 mM-KCl, MgCl2 concentration according to the titrated optimum (1–8 mM); 0.2 μM of primers and 200 μM of each of the four nucleotides, dATP, dGTP, dCTP and dTTP. One μl of Taq polymerase (Cetus) was added per 50 μl of reaction mix. The samples were overlaid with 50 μl of mineral oil (Sigma) and then subjected to 35 cycles of denaturation (94 °C for 30 s; first cycle 5 min), annealing (52 °C for 60 s) and extension (72 °C for 90 s) in an automated DNA thermal cycler (Perkin
Table 1. **EBNA molecular masses (kDa) and DNA fragment size* (bp)** determined by immunoblotting and Southern blotting

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Mol. mass</th>
<th>Sizea</th>
<th>Mol. mass</th>
<th>Sizeb</th>
<th>Mol. mass</th>
<th>Sizec</th>
<th>Mol. mass</th>
<th>Sized</th>
<th>Mol. mass</th>
<th>Sizee</th>
<th>Mol. mass</th>
<th>Sizef</th>
</tr>
</thead>
<tbody>
<tr>
<td>B95-8</td>
<td>80</td>
<td>768</td>
<td>92</td>
<td>580</td>
<td>1160</td>
<td>155</td>
<td>2080</td>
<td>154</td>
<td>410</td>
<td>150</td>
<td>1600</td>
<td>720</td>
</tr>
<tr>
<td>D9</td>
<td>80</td>
<td>710</td>
<td>91</td>
<td>570</td>
<td>1120</td>
<td>163</td>
<td>2080</td>
<td>154</td>
<td>420</td>
<td>151</td>
<td>1600</td>
<td>800</td>
</tr>
<tr>
<td>D45</td>
<td>72</td>
<td>360</td>
<td>97</td>
<td>580</td>
<td>1370</td>
<td>158</td>
<td>2050</td>
<td>154</td>
<td>410</td>
<td>165</td>
<td>1770</td>
<td>970</td>
</tr>
<tr>
<td>D46</td>
<td>79</td>
<td>650</td>
<td>84</td>
<td>1860</td>
<td>146</td>
<td>1960</td>
<td>ND</td>
<td>ND</td>
<td>2000</td>
<td>ND</td>
<td>1850</td>
<td>860</td>
</tr>
<tr>
<td>D47</td>
<td>86</td>
<td>860</td>
<td>93</td>
<td>560</td>
<td>1160</td>
<td>148</td>
<td>2020</td>
<td>154</td>
<td>410</td>
<td>155</td>
<td>1680</td>
<td>880</td>
</tr>
<tr>
<td>D48</td>
<td>80</td>
<td>720</td>
<td>89</td>
<td>570</td>
<td>1140</td>
<td>145</td>
<td>1990</td>
<td>154</td>
<td>410</td>
<td>166</td>
<td>1740</td>
<td>900</td>
</tr>
<tr>
<td>D49</td>
<td>97</td>
<td>1120</td>
<td>92</td>
<td>570</td>
<td>1120</td>
<td>144</td>
<td>1990</td>
<td>154</td>
<td>410</td>
<td>167</td>
<td>1770</td>
<td>900</td>
</tr>
<tr>
<td>D50</td>
<td>84</td>
<td>525</td>
<td>88</td>
<td>560</td>
<td>1120</td>
<td>153</td>
<td>2030</td>
<td>154</td>
<td>410</td>
<td>146</td>
<td>1580</td>
<td>840</td>
</tr>
<tr>
<td>D108</td>
<td>88</td>
<td>920</td>
<td>89</td>
<td>565</td>
<td>1155</td>
<td>154</td>
<td>2030</td>
<td>154</td>
<td>410</td>
<td>137</td>
<td>1560</td>
<td>740</td>
</tr>
<tr>
<td>D109</td>
<td>85</td>
<td>875</td>
<td>91</td>
<td>570</td>
<td>1185</td>
<td>154</td>
<td>2030</td>
<td>154</td>
<td>410</td>
<td>166</td>
<td>1830</td>
<td>950</td>
</tr>
<tr>
<td>D110</td>
<td>84</td>
<td>720</td>
<td>92</td>
<td>550</td>
<td>1140</td>
<td>148</td>
<td>2020</td>
<td>154</td>
<td>410</td>
<td>180</td>
<td>1870</td>
<td>1110</td>
</tr>
<tr>
<td>D111</td>
<td>88</td>
<td>780</td>
<td>91</td>
<td>560</td>
<td>1160</td>
<td>158</td>
<td>2050</td>
<td>154</td>
<td>410</td>
<td>144</td>
<td>1680</td>
<td>800</td>
</tr>
<tr>
<td>D115</td>
<td>79</td>
<td>525</td>
<td>88</td>
<td>570</td>
<td>1140</td>
<td>157</td>
<td>2040</td>
<td>154</td>
<td>410</td>
<td>175</td>
<td>1850</td>
<td>1010</td>
</tr>
</tbody>
</table>

* a, The DNA was cleaved with MspI and probed with E1 (coordinates 108159–108933); b, the DNA was cleaved with MspI and probed with E2H (coordinates 48848–49990); c, the DNA was cleaved with MspI, and a subfragment of the BamHI Y fragment (E2Y) covering the proline repeat (47778–48848) was used as probe; d, the DNA was cleaved with EcoRI and probed with a subfragment of the BamHI E region (coordinates 93162–95239); e, the DNA was cleaved with PvuII and probed with E4 (coordinates 97283–97693); f, the DNA was cleaved with BglII and probed with a subfragment of the BamHI E region (coordinates 99759–100613); g, PCR over the n X 39 bp repeat in EBNA-6 size of resulting fragment in bp resulting from use of primers according to Methods.

ND, Not detectable; the EBNA-2 probe E2H did not detect the type B isolate, and the sera used in immunoblotting did not detect type B EBNA-4 or -6 protein.

The PCR products were purified using magnetic beads with covalently coupled streptavidin (Dynabeads-M-280-streptavidin from Dynal, Oslo, Norway). A neodymium–iron–boron permanent magnet (Dynan) was used to sediment the beads during the washing procedures. Beads (600 μg) were mixed with 80 μl of the PCR mixture and incubated for 15 min in a solution containing 10 mM-Tris-HCl pH 8.0, 1 mM-EDTA and 1 mM-NeNaCl. The immobilized double-stranded DNAs were washed with 50 μl TE buffer and then incubated with 0.1 M NaOH for 5 min. The supernatant was removed and the beads containing the immobilized single-stranded DNA were washed with 50 μl 0.1 M NaOH and then three times with TE buffer. After the last wash, the beads were dissolved in 10 μl of water. The sequencing reactions were performed with reagents from the Autoread T7 sequencing kit (Pharmacia) according to the manufacturer's protocol. We used 10 pmol of each of the fluorescent labelled primers. For the forward sequencing we usedKF 65-F (5' TATCGCAGAAGAAC-AACCC 3'); coordinates in the EBV genome 100515–100650) and for the reverse sequencing we used KF 68-F (5' TACATGGAATGTTGT-GACCTTGG 3'; coordinates in the EBV genome 108636–108658). The stopped sequence reactions were run on an ALF sequencer (Pharmacia LKB). The sequences obtained were compared using DNASIS software (Hitachi Software Engineering).

**Statistical analyses.** The Spearman rank correlation method was used throughout the study for the comparison of apparent molecular mass and RFLP fragment sizes.

**Fig. 2. (a)** Immunoblotting with polyvalent JB serum, which recognizes EBNA-1, -2, -3, -4 and -6. Extracts were prepared from 12 wild-type isolates obtained by spontaneous outgrowth or by transformation of cord blood cells with mouth wash. Ramos is an EBV-negative control cell line, and B95-8 was used as a positive control. Standard molecular mass markers were: myosin (200 kDa), β-galactosidase (116 kDa), phosphorylase (97 kDa) and albumin (66 kDa). See Table 1 for the estimated molecular masses of the EBNAs of the 12 isolates. **(b)** Schematic representation of the immunoblot in (a) showing the assignment of individual bands to specific EBNA proteins. The assignments were based on immunoblotting with monospecific reagents.
Ebnotyping registers RFLP in EBNA sequences

Fig. 2. For legend see opposite.
Results

Selection and characterization of EBV isolates by Ebnotyping

For this study we selected 13 LCLs based on the large variation in EBNA molecular masses. Twelve were type A isolates and one (D46) was a type B isolate.

The specific bands for all the EBNA as visible in Fig. 2(a) were identified by monospecific reagents (see also Fig. 4a; data not shown for EBNA-1, -2, -3 and -4), making the basis for the molecular mass estimations shown in Table 1. The procedure for rescuing these isolates in vitro results in clonality. However, multiple weak bands were detected in the immunoblots with monospecific reagents. In contrast, in the Southern blots only a single band was detected. This suggests that the isolates are indeed clonal, and that the additional bands in the immunoblots are non-specific or degradation products.

EBNA-1-4 molecular mass variation compared with variation of repeats within the coding exons

In the 13 isolates (including B95-8) we confirmed the positive correlation between the EBNA-1 apparent molecular mass in immunoblotting, and the size variation of the Gly-Ala repeat in Southern blotting, using the MspI–MspI cleaved probe (E1) covering this repeat (Fig. 1a, b; r = 0.86, P = 0.0002; Fig. 5a). We could detect some variation at the DNA level for EBNA-2. The correlation between the sizes of the DNA fragments and the protein molecular mass was not significant with the Gly-Ala repeat in Southern blotting, using the MspI–MspI cleaved probe (E1) covering this repeat (99759-100613, E6, Fig. 1 f) in a Southern blot. No or very little variation of this fragment was detected (data not shown). This excluded variation in the 10 × 15 bp fragment covering part of the large repeat within EBNA-3 (termed the ABCD repeat), which showed that this fragment has the same size in all the isolates (data not shown). From these data we concluded that there is no detectable variation in the first exon, BLRF3, or in the left half of BERF1 (Fig. 1e-f). The ABCD repeat consists of several imperfect copies of the sequences termed A, B, C and D, ordered ABCBCACADD in B95-8. The data can only be explained by further addition of coding DNA at the very 3' end of the BERF1 exon. To elucidate this further an EcoRI probe (93162–95239) covering most of and extending beyond the 3' end of BERF1 was used (Table 1; E3, Fig. 1 e-f). After digestion with EcoRI the resulting fragments in a Southern blot correlated with the variation of the EBNA-3 protein (r = 0.90, P = 0.0001; Fig. 5d; Table 1). We conclude that the variation of EBNA-3 seems to be due primarily to additional coding sequence at the 3' end of the ABCD repeat.

In 12 of the 13 isolates EBNA-4 molecular mass did not vary at all (Fig. 2, Table 1). EBNA-4 was not detected in the B-type isolate D46 with our monospecific antibody. A probe covering the 3 × 60 bp repeat within BERF2B (E4 probe; Fig. 1 e-f) revealed no variation between the isolates, except for D46 (Table 1). This was in line with the lack of variation in the protein molecular mass.

EBNA-6 molecular mass variation correlates positively with the size variation of the n × 39 repeat but not with the m × 15 repeat in BERF4

The 13 isolates varied considerably in the apparent molecular mass of EBNA-6, as shown by immunoblotting (Fig. 2, Table 1). EBNA-6 was identified with a human serum with high titres against EBNA-6 (Fig. 4a) and with a monospecific affinity purified antibody by immunoblotting (data not shown). In the KpnI and PvuII blots (Fig. 3), it was shown that the sizes of two fragments covering the 3'-half of the EBNA-6 exons (3185 bp/PvuII and 2634 bp/KpnI in B95-8) correlated positively with the size variation of the EBNA-6 polypeptide (Figs 1e-f, 3 and 4a). These fragments include the two repeat regions in EBNA-6, an m × 15 bp repeat and an n × 39 bp repeat. The number of repeats in the B95-8 prototype sequence are m = 10 and n = 3.

Subsequently, the DNA was cleaved with BgII and BamHI, and probed with a BgII–BamHI fragment (99759–100613, E6, Fig. 1f) in a Southern blot. No or very little variation of this fragment was detected (data not shown). This excluded variation in the 10 × 15 bp repeat. When the DNA was cleaved with BgIII only, and hybridized with the same probe, the resulting fragments showed a strong positive correlation with the EBNA-6 protein molecular mass (Figs 4a, b and 5e), suggesting that the variation was due to the n × 39 repeat (r = 0.91; P = 0.0001).
Fig. 3. (a) Southern blotting after cleavage with *Pvu*II and hybridization with the *Bam*HI E fragment. The positions of the molecular mass marker fragments are indicated. (b) Southern blotting after cleavage with *Kpn*I. This Southern blot was probed with the *Bam*HI E fragment. K562 was used as a negative control. The positions of the molecular mass marker fragments are indicated.
Fig. 4. (a) Immunoblot showing EBNA-6 size variation using a human serum with high titres to EBNA-6 followed by autoradiography. The arrows indicate the specific bands (note that D108 is not included; from a separate blot its EBNA-6 molecular mass was estimated to 137 kDa). (b) Southern blotting after cleavage with BglII (BglII cleaves in B95-8 at position 99759–101353) and probing with probe E6 (coordinates 99759–100613) covering the n x 39 bp repeat. The positions of the molecular mass marker fragments are indicated. (c) PCR (primers localized at coordinates 100395 and 101119 resulting in a 724 nt fragment in B95-8) covering the n x 39 bp repetitive element. PCR products hybridized with an oligonucleotide probe recognizing both type A and type B isolates (coordinates 100599–100618). The positions of the molecular mass marker fragments are indicated. (d) The same blot as in (c) stripped and hybridized with an oligonucleotide probe recognizing only type B isolates, according to Methods.
Ebnotyping registers RFLP in EBNA sequences

Fig. 5. Plot showing the relation between size (bp) of relevant DNA fragments and apparent molecular masses in immunoblots for (a) EBNA-1 \((r = 0.86; P = 0.0002)\), (b) EBNA-2 with probe E2H \((r = 0.18; P = 0.57)\), (c) EBNA-2 with probe E2Y \((r = 0.52; P = 0.08)\); the D46 type is not included in this calculation), (d) EBNA-3 \((r = 0.90; P = 0.0001)\); D46 is excluded from this calculation), (e) EBNA-6 with probe E6 \((r = 0.91; P = 0.0001)\) and (f) EBNA-6 with PCR amplification \((r = 0.88; P = 0.0002)\). Correlation coefficients were calculated according to the Spearman rank correlation method.

PCR covering the \(n \times 39\) repeat correlated with EBNA-6 protein size

The \(n \times 39\) repeat from each of the 13 isolates was amplified by PCR. The result demonstrated a strong correlation between the \(n \times 39\) repeat and the polypeptide size \((r = 0.88; P = 0.0002)\); figs 4a, c and 5f, Table 1). In some cases the sizes of the PCR products deviated from the Southern blot data; the correlation between the Southern blot fragments and the protein molecular mass was stronger than between the PCR products and the protein size (Fig. 5e, f).

By sequencing we confirmed that the \(n \times 39\) repeat varies between the isolates: for instance, D49 had eight 39-bp repeats in contrast to the prototype B95-8 with three 39-bp repeats (data not shown).

Type A and B isolates could be distinguished by hybridizing the PCR products with two different oligonucleotides, one recognizing both type A and type B isolates (Fig. 4c) and the other recognizing only type B isolates (Fig. 4d).

Discussion

Thirteen isolates were selected to cover the range of variation commonly seen (Gratama et al., 1994). In these, we found limited molecular mass variation of EBNA-4, while EBNA-1, -2, -3 and -6 variation was shown to be useful for distinguishing the isolates by Ebnotyping. The EBNA-1 and EBNA-6 polypeptide variation, but not that of EBNA-2, was shown to be predominantly dependent on size variation of repetitive parts of the coding regions. Although we detected RFLPs in the EBNA-2 coding region, this did not correlate with protein molecular mass. This may be due to secondary modifications of EBNA-2 by multiple phosphorylation resulting in several forms of the protein even within the same isolate (Petti et al., 1990). EBNA-3 molecular mass variation also correlated with size variation of the coding DNA within or just 3' of the ABCD repeat. This repetition has a more complex composition than those of the other EBNA and the repeats are not perfect copies of each other.

Variation in length of the repetitive amino acid sequences in the exons may serve as a more general explanation for the molecular mass variation of the EBNA. The variation may arise as a result of heterologous recombination within the repeats, particularly during productive virus infection, generating varying lengths of the sequences coding for the repeats. We have separately obtained further support for this hypothesis. Several isolates from a bone marrow transplant (BMT)
recipient differed only in the EBNA-1 molecular mass, due to different lengths of the Gly-Ala repeat. Irrelevant sequences outside the repetition in the EBNA-1 exon were shown to be almost identical between these isolates, differing only in a few bases (Gratama et al., 1992; Friis et al., 1995). Unrelated isolates showed larger sequence divergence in the same regions. This suggested that recombinations within the Gly-Ala repeat had occurred between endogenous viruses in the BMT patient.

On the other hand, our data show that the EBNA molecular mass variation is not exclusively due to the repeats, although these recombinations dominate. Other recombinations in the exons may influence the polypeptide size.

It has previously been shown (with limited material: five established EBV-positive cell lines) that there is a correlation between the length of the Gly-Ala repeat coding region in DNA and the wide molecular mass variation of EBNA-1 (Hennessy et al., 1983; Hennessy & Kieff, 1983). This is now definitely confirmed on a wider panel of wild-type isolates. The Gly-Ala repeat in IR 3 of our isolates showed a very strong correlation with EBNA-1 size variation. The DNA fragment encompassing the repeat region varied between 360 bp and 1120 bp, corresponding to an approximate variation of 108 to 336 amino acids, or approximately 23 kDa at the protein level. Little or no size variation can be ascribed to other coding regions or possible secondary modifications of the protein. Unfortunately, we have so far failed to establish PCR amplification covering the Gly-Ala repeat in IR 3. This is most likely due to technical reasons because of the high degree of repetitions and the high GC density of the region combined with a relatively long DNA segment containing the repeat.

We therefore focused on EBNA-6 which also showed large size variation and where most of the variation could be localized to a limited region. The coding region for EBNA-6 contains two repeat regions, a 10 × 15 bp repeat which was conserved in size between the isolates, and an n × 39 bp repeat coding for three repeats of the following 13 amino acids in B95-8: Pro-Ala/Pro-Pro-Gln-Ala—Pro—Tyr—Gln—Gly—Tyr—Gln—Glu—Pro. The size of this coding region varies between the isolates, and the variation correlates well with the variation of the protein. There is an approximate difference of 390 bp between the smallest and the largest isolate in this region as shown by PCR, which corresponds well to between three repeats in B95-8 and 13 copies in the largest, D110.

The Ebnotyping method has been useful for studies of the in vivo biology, host interaction and epidemiology of EBV. There have been two major limitations to the method. It is time-consuming to rescue isolates as proliferating cell lines, by in vitro transformation with mouth washes or by the more tedious spontaneous outgrowth. Yet more problematic may be the fact that the method involves a biological selection step, due to transformation in vitro before the isolates can be characterized. These limitations are overcome if isolates can be studied directly in in vitro explanted material, peripheral blood lymphocytes or saliva/mouth wash. PCR offers unique advantages in this respect. Therefore we established a PCR reaction over the varying n × 39 repeat in the EBNA-6 region. This allowed most of the isolates to be distinguished from each other. Characterization of the variation of only one of the EBNA1s means that not all the variants that can be identified by immunoblotting can be distinguished. Consequently, a combination of several PCRs over different variable regions would be optimal for identification of variants. Genotyping based on sequencing of EBV is an alternative, but it is too expensive and time-consuming for quick screening of large numbers of specimens (de Campos-Lima et al., 1993; Lin et al., 1993).

EBV type A and type B isolates could also be distinguished by the same PCR reaction, by subsequent hybridization of the PCR products with specific oligonucleotides. This was expected due to a wide sequence divergence between the two EBV types in this region (Rowe et al., 1989; Sample et al., 1990). Thus the EBNA-6 PCR allows identification of single isolates and of subtype by the same reaction, in contrast to earlier PCR used to distinguish between type A and B EBV (Sample et al., 1990; Kyaw et al., 1993).

Protein size variation and DNA RFLP have been used to classify and follow transmission of other herpesviruses. Pereira et al. (1976) identified seven variable polypeptides in herpes simplex virus, and attempted to classify variants in two groups based on this variation. The basis for the size variation was never established. Linneman et al. (1978) studied the transmission of herpes simplex virus type I in a day nursery by determining RFLP in the viral genome. Thus, polypeptide size variation is not unknown among herpesviruses, but has not been assigned to repetitive regions.

The relatively large size variations of the EBNA polypeptides between virus isolates is a remarkable phenomenon, considering that these are proteins proven or likely to be vital for establishment and maintenance of latent infection in the host. The variation must be localized to domains of the polypeptides, the size variation of which is compatible with functionality. The polypeptide size is normally stable over time in a single individual and in cell lines carrying a virus variant in vitro (Yao et al., 1991; our unpublished data). However, recombinants affecting the EBNA-1 region have been generated in vitro during the lytic cycle virus production in B95-8 (J. W. Gratama, unpublished results). This observation further supports the hypothesis that variants
arise by heterologous recombination during the lytic cycle.

This study was supported by the Swedish Cancer Society, the Swedish Medical Research Council and by the Cornell foundation. J. W. Gratama is a fellow of the Royal Dutch Academy of Arts and Sciences. M. Rowe was supported by the Wellcome Trust, UK. We are grateful for the excellent technical assistance of Anna Westman, Anita Aleman and Eva Ragnar.

References


growth phenotype reflect novel patterns of Epstein–Barr virus latent gene expression in Burkitt’s lymphoma cells. EMBO Journal 6, 2743–2751.


(Received 31 August 1994; Accepted 21 November 1994)