The EMBO Journal vol.8 no.9 pp.2639-2651, 1989

EBV gene expression in an NPC-related tumour

M.M.Hitt[1,2], M.J.Allday[1], T.Hara[1,3], L.Karran[1], M.D.Jones[1], P.Busson[4], T.Tursz[4], I.Ernberg[5] and B.E.Griffin

[1]Department of Virology, Royal Postgraduate Medical School, Du Cane Road, London W12 0NN, UK, [4]Laboratoire d'Immunobiologie des Tumeurs, 1156 Centre National de la Recherche Scientifique, Institut Gustave Roussy, 94805, Villejuif, Cedex, France and [5]Department of Tumor Biology, Karolinska Institute, Box 60400, 10401 Stockholm, Sweden

[2]Current address: Department of Biology, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S4K1, Canada

[3]Current address: Department of Food Science and Technology, Faculty of Agriculture, Kyushu University, Hakozaki, Fukuoka 812, Japan

Communicated by B.E.Griffin

A nasopharyngeal carcinoma tumour (designated C15) propagated in nude mice has been used to generate a large cDNA library that we have analysed for Epstein-Barr virus (EBV) gene expression. No gross alterations exist in viral DNA from C15 relative to other human isolates and the large deletion present in the B95-8 'prototype' viral strain established in marmoset cells is not found; C15 contains no linear virion DNA. In the cDNA library, of the six EBV nuclear antigens (EBNAs) expressed in latently infected B-lymphocytes, only clones for EBNA-1 are found. These data are confirmed by immunoblotting. Sequence analysis shows the EBNA-1 mRNA splicing pattern in the carcinoma to differ from that observed in B-lymphocytes. Further, contrary to observations with B-cell lines, most viral transcription in the tumour is localized onto the 'rightmost' region of the conventional EBV physical map. Transcripts identified corresponding to known genes include those for the latent membrane protein (LMP), the alkaline DNA exonuclease and probably the terminal protein; major transcripts are also derived from the *Bam*HI D fragment and the region deleted in B95-8 EBV DNA. Novel transcripts have also been identified that proceed in an anti-sense direction to genes encoding functions associated with replication, such as the viral DNA polymerase. They contain a large, hitherto unidentified, open reading frame in the viral genome that is complementary to the putative function known as BALF3 and a smaller open reading frame complementary to BALF5 (the DNA polymerase gene). From the present studies we can conclude that: (i) EBV transcription patterns in the epithelial cells vary markedly from those identified previously in B-cells, reflecting differential use of promoters or splicing patterns. (ii) Transcription is tightly regulated and restricted in the C15 tumour with many latent genes, notably EBNAs 2-6, being 'switched off'. (iii) A family of cytoplasmic RNAs are transcribed in an antisense direction to a number of existing open reading frames in the EBV genome. (iv) There are a number of mutations in C15 transcripts relative to the B95-8 genome, some of which could result in amino acid alterations in proteins.

Key words: complementary ('antisense') transcript/cDNA library/EBNAs/splicing/transcription patterns/transcription regulation

Introduction

The human herpes virus Epstein-Barr virus (EBV), a causative agent for self-limiting infectious mononucleosis, is associated with two well-characterized malignancies, Burkitt's lymphoma (BL) of B-cell origin and nasopharyngeal carcinoma (NPC), derived from poorly differentiated epithelial cells. The tumours themselves share few features in common, however, whether with regard to geographical prevalence, peak age incidence, response to treatment or chromosomal abnormalities (as reviewed by Simons and Shanmugaratnam, 1982; Lenoir *et al.*, 1985). Whereas translocation of the oncogene, c-*myc*, has been associated with BL, no cellular oncogene involvement with NPC has been identified. Previous studies aimed at understanding how a single virus might be associated with different pathologies have investigated the gross structure of the viral genomes seeking alterations that might prove relevant in this context. No consistent detectable major differences were observed, however, in EB viral genomes derived from Burkitt's lymphomas or nasopharyngeal carcinomas or even from the non-malignant diseases associated with EBV (Rymo *et al.*, 1979; Bornkamm *et al.*, 1984; Lung *et al.*, 1988). Serological studies have been equally unenlightening.

The purpose of the present investigation was to reassess this question by determining the transcription patterns in NPC tumours and comparing these with viral transcription data from transformed B-cells (as reviewed by Dillner and Kallin, 1988). Earlier studies to address this question have been hampered by the absence of NPC cell lines and difficulties of obtaining sufficient intact messenger RNA from NPC biopsies. Further, biopsy material poses problems in interpretation since the tumour mass is generally invaded with lymphocytes, including B-lymphocytes which carry EBV DNA (M.H.Ng, personal communication). The question could be addressed, however, once abundant tumour material with suitable properties, obtained from NPC transplantations in nude mice, became available (Busson *et al.*, 1987). One of these NPC-derived tumours, designated C15, appeared to harbour the virus in a tightly latent form in that, although it contained multiple copies of the EBV genome, it contained no linear viral DNA and produced no virions, as shown by lack of expression of early viral (EA) and capsid (VCA) antigens. Passage in nude mice also resulted in the loss of invading T- and B-cells (and

©IRL Press

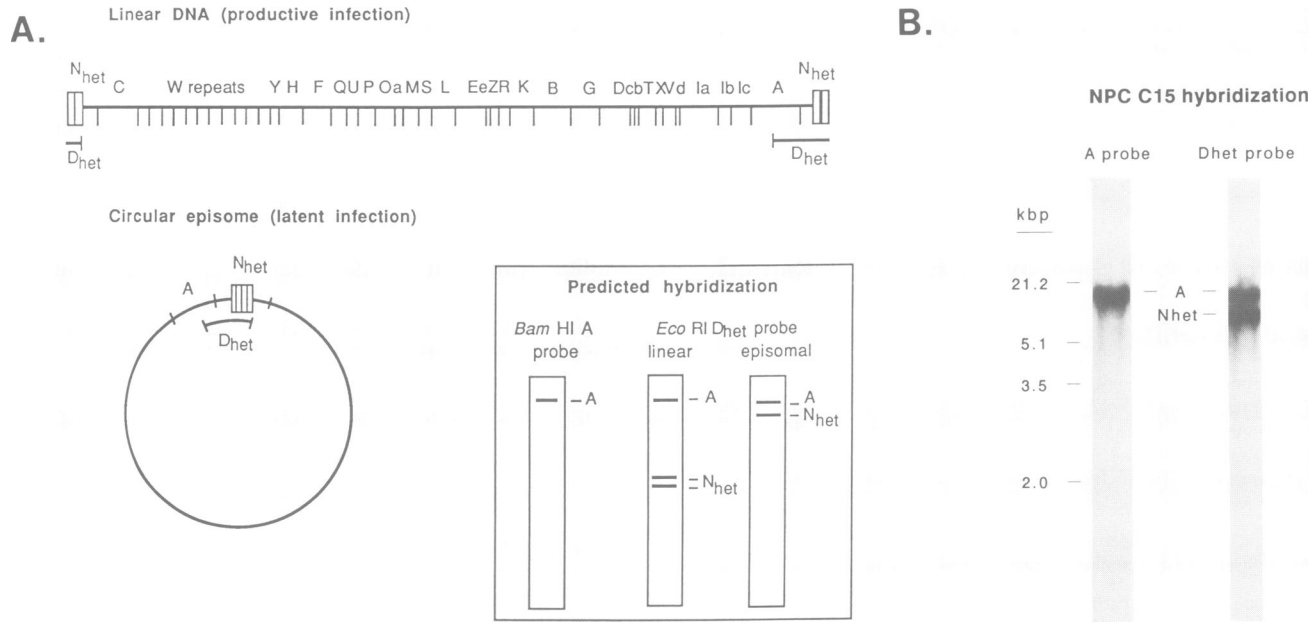


Fig. 1. (**A**) Schematic diagrams of linear and circular DNAs and predicted hybridization patterns using terminal probes. The *Bam*HI restriction map of EBV is shown. C15 DNA contains *Bam*HI restriction fragments, designated Ia, Ib and Ic, as indicated, that are truncated or absent in the B95-8 strain of EBV, but present in other EBV isolates. The location of the *Eco*RI Dhet fragment is shown on both the linear and circular maps of EBV and the predicted pattern of hybridization to *Eco*RI Dhet expected from fragments A and Nhet of either linear or fused episomal EBV DNAs cleaved with *Bam*HI given. (**B**) Hybridization data. Chromosomal DNA from the C15 tumour was cleaved with *Bam*HI and the resulting fragments, separated by electrophoresis, were probed with radiolabelled *Bam*HI A (as control) and *Eco*RI Dhet fragments, as indicated. Bands corresponding to episomal DNA, but not linear DNA, were detected.

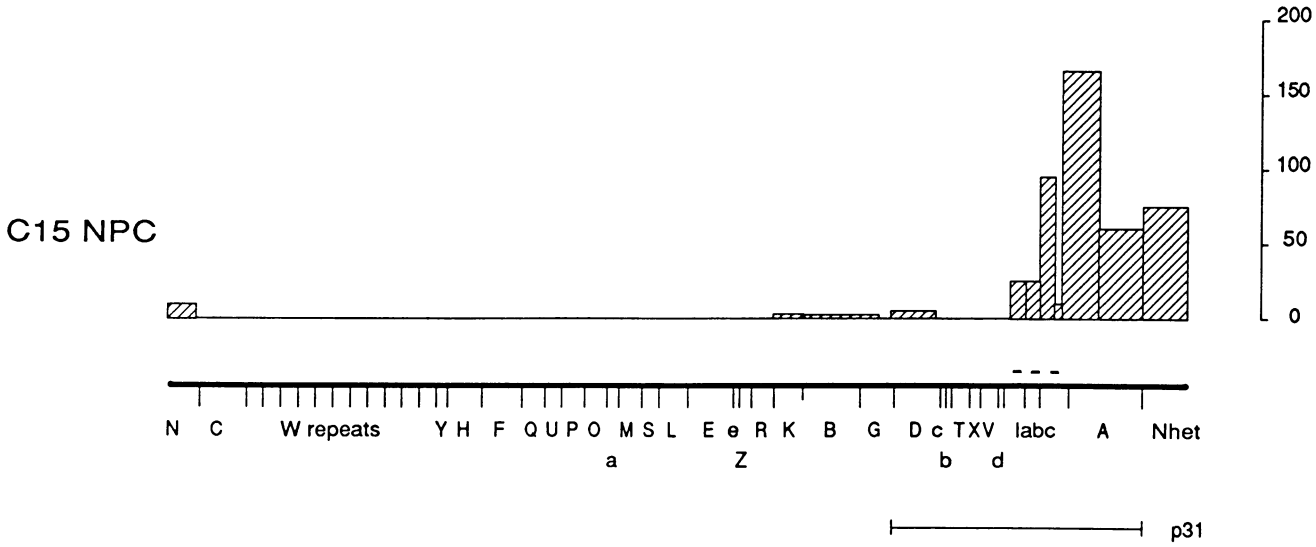


Fig. 2. Patterns of 'latent' gene expression in the C15 NPC tumour. A histogram (at top) shows the data obtained by hybridizing a comprehensive cDNA library (C15 NPC) generated in λgt10 (from mRNA isolated from the C15 tumour) with *Bam*HI restriction fragments of EBV DNA from B95-8 and Raji cells (Arrand *et al.*, 1980), as indicated in the text. The number of viral clones detected per 10^6 recombinants screened is shown on the ordinate, and their location on the EBV genome is shown relative to the *Bam*HI physical map (at bottom) which has been linearized to correspond with the sequence of B95-8 DNA (Baer *et al.*, 1984). The region corresponding to Ia−c, indicated by dashed lines (---), is deleted in the B95-8 genome. (cDNA clones which hybridized to *Bam*HI H are not included since they could not be localized in B95-8 DNA.) The location of the epithelial cell immortalizing fragment, designated p31 (Griffin and Karran, 1984), is indicated.

presumably any cells that might express viral lytic functions). Therefore, C15 appeared to be an ideal model for assessing viral gene expression linked to strict latency and tumorigenesis in the epithelium.

A comprehensive cDNA library has been produced from C15 tumour material and its analysis is presented here and discussed. The EBV transcription pattern in these carcinoma cells has proved markedly different from those patterns previously reported using B-cells: Only one gene encoding an EBNA (EBNA-1) is transcribed. Moreover, in addition to transcripts that could be ascribed to known viral genes such as the latent membrane protein (LMP) and the alkaline DNA exonuclease (Baer *et al.*, 1984), major cDNAs not hitherto predicted from transcription maps (Farrell, 1987) of the B-cell-transforming lytic strain of EBV, B95-8, were obtained. Our analysis of the library also showed that tran-

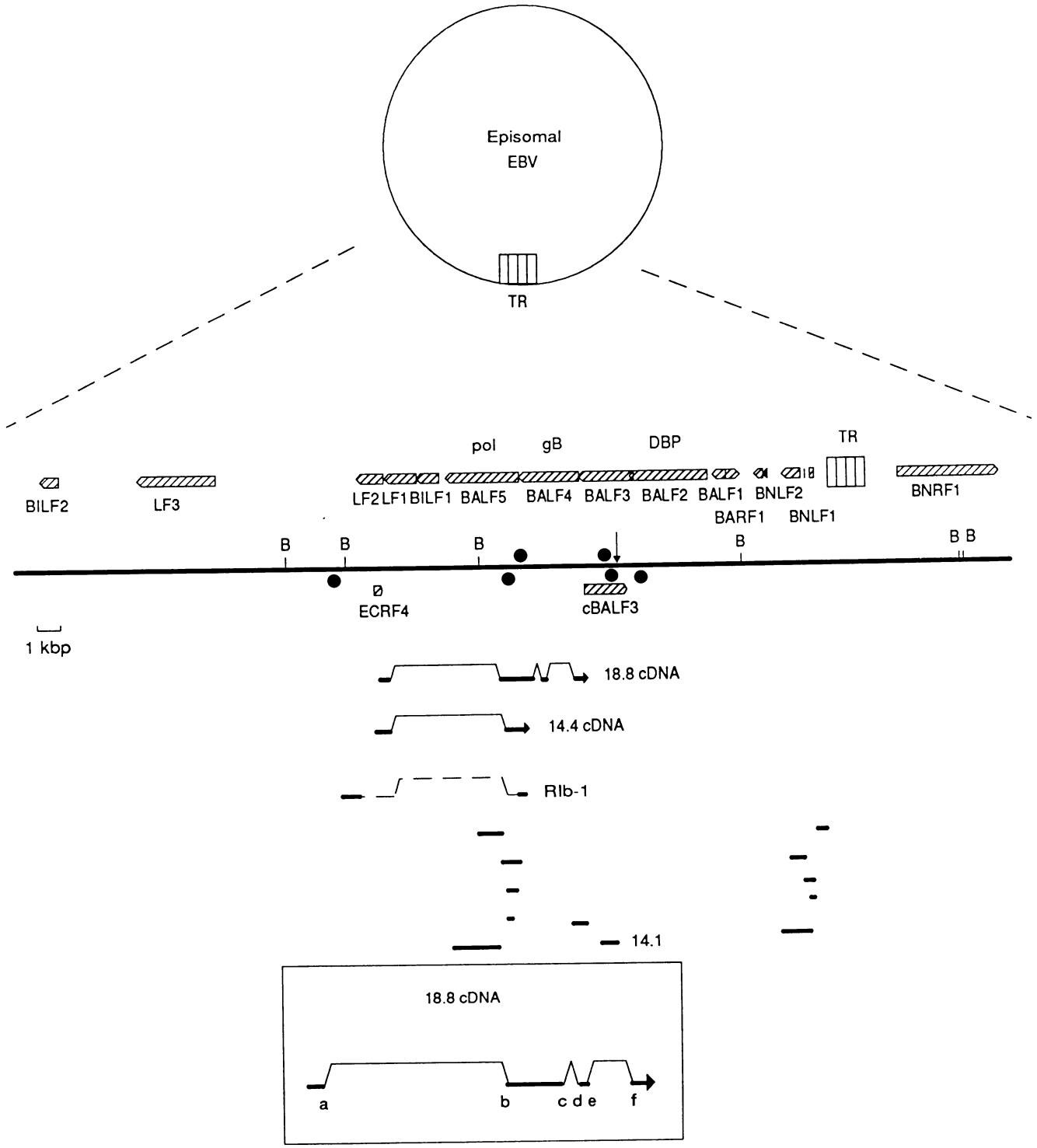


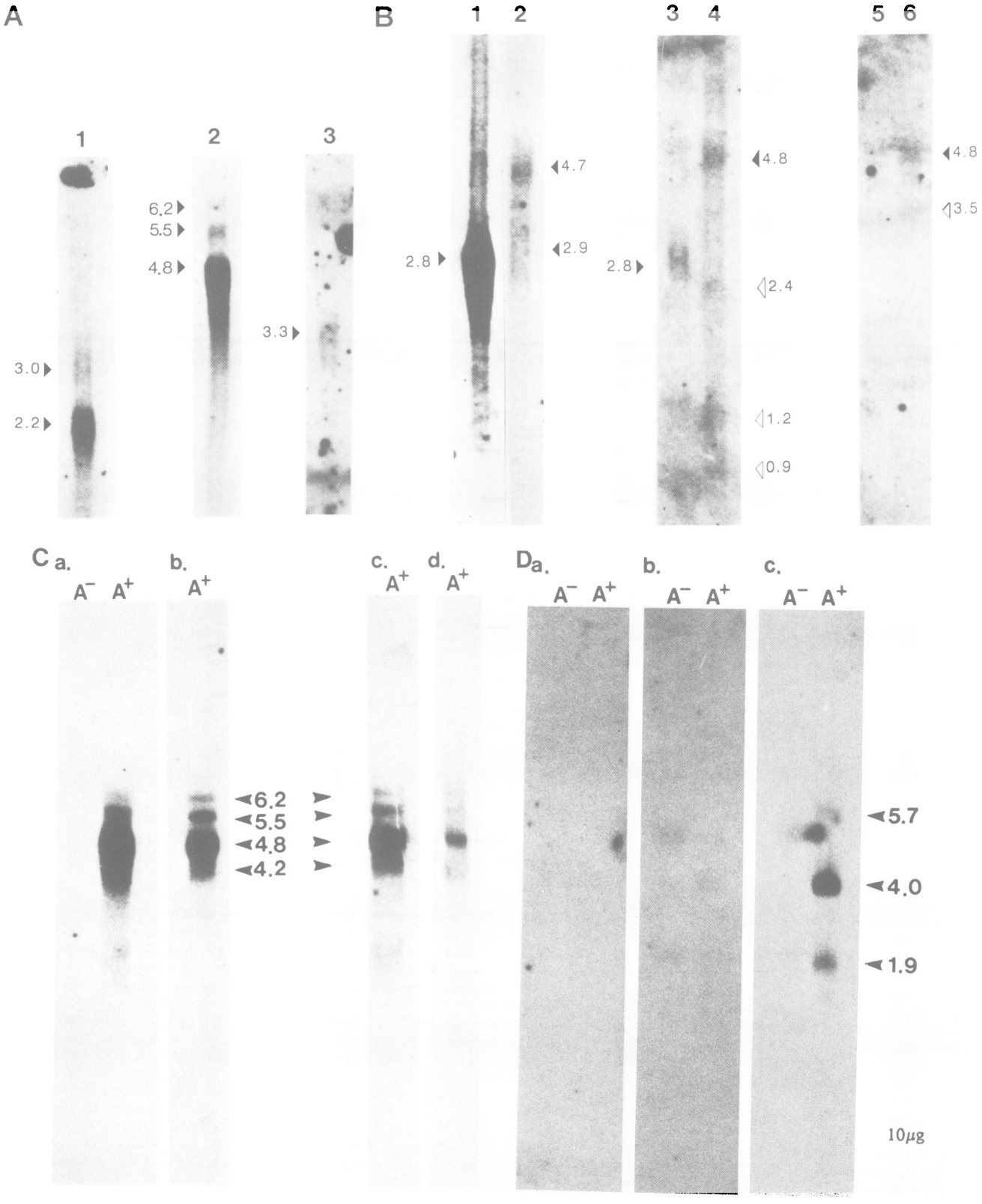
Fig. 3. Location of 15 cDNA clones, including 18.8, 14.4 and RIb-1, relative to open reading frames and functions identified on episomes of EBV by DNA sequence analysis. The terminal repeats (TR) are shown. Most open reading frames from the B95-8 viral stain are taken from data summarized by Farrell (1987); the 'complementary' open reading frame, EC-FR4, is from Bankier *et al.* (1983). Data on the three 'leftward' reading frames, LF1–LF3, present in the Raji genome, but absent in B95-8, are from B.G.Barrell (personal communication). pol = viral DNA polymerase, gB = glycoprotein B homologue, DBP = major DNA binding protein homologue. The splicing patterns of 18.8 and 14.4, taken from the cDNA sequences (T.Hara *et al.*, unpublished), show the transcript to proceed from left to right, as indicated; 18.8 and 14.4 have the same donor site in *Bam*HI Ic and acceptor sites in *Bam*HI A. (The acceptor site in the second intron of 18.8 cannot be completely correlated with the sequence of B95-8 DNA and appears to reflect limited sequence alterations in C15; alternatively, we cannot exclude the possibility of a change having occurred during cloning.) In these clones, the 5′ end of the cDNA is found in the LF2 open reading frame. Open reading frames in 18.8 >500 bp were not found. Only the 5′ and 3′ ends of RIb-1 have been determined by sequence analysis, but this clone has been further subjected to restriction enzyme analysis, allowing the structure as shown, which extends across the whole of the LF2 open reading frame, to be postulated. Northern blot analyses with oligonucleotide probes suggest that the AATAAA polyadenylation signal shown (↓) is used as a messenger termination signal in the 18.8 transcript; from left to right, the probes (●) used were 3852, 1811, 81, 3382, 3593 and 3594 (see Materials and methods). The unspliced cDNA clone designated 14.1 contained a long open reading frame, as indicated (cBALF3). **Inset**. Numbers at splice sites **b**–**f** in 18.8 are given relative to the *Bam*HI A restriction site (first G of GGATCC = 1): **b** = nucleotide 978; **c** = 2438; **e** = 2640; **f** = 4337. Site **a** is not found in B95-8, but is located at position −4974 in Raji Ic; side **d** cannot be precisely located with regard to the B95-8 sequence.

scripts from a large (9×10^6 dalton) sequence deleted in B95-8 (Raab-Traub *et al.*, 1980) are among those found in the C15 tumour.

Results

The C15 tumour contains no EBV virion DNA

The original NPC tumour from a nude mouse, designated C15, contained ~30 copies of the EB viral genome and expressed no EA or VCA (Busson *et al.*, 1988). Our cDNA library (see below) was made from (twice) poly (A)-selected mRNA isolated from a tumour at 16th passage. To ascertain whether virus reactivation had occurred on propagation of the tumour in the animal, chromosomal DNA was cleaved with *Bam*HI, the products separated by electrophoresis and analysed by Southern blotting, using the ^{32}P-labelled *Eco*RI Dhet fragment of EBV DNA as a probe (see Figure 1A). Two bands only were obtained (Figure 1B), as expected from



episomal EBV DNA, and not three or more as would be predicted if linear virus molecules were present (see figure and Raab-Traub and Flynn, 1986). Therefore, gene expression observed in the C15 tumour should be related to strictly latent forms of viral DNA.

Preparation and analysis of a cDNA library from the C15 tumour

A cDNA library consisting of $3-4 \times 10^6$ independent recombinant clones was constructed in λgt10 from mRNA (~ 5 μg from twice-selected polyadenylated RNA) isolated from the C15 tumour. Analysis of a cDNA library from the EBV-positive marmoset B-cell line, B95-8, had previously shown that low abundance transcripts (<5 copies/cell) encoding EBNA3 could be detected when 6×10^5 recombinants were screened (Allday *et al.*, 1987). This is consistent with the experience of others (Speck and Strominger 1985; Bodescot *et al.*, 1986; Sample *et al.*, 1986) who obtained similarly rare EBNA-associated clones from B-cell libraries consisting of $\sim 10^6$ recombinants. Therefore the C15 EBV cDNA library was deemed of sufficient size to represent low copy number transcripts in the tumour and was probed in detail, initially by screening a minimum of 10^6 independent clones with a mixture of EBV DNA comprised of *Bam*HI K, C and W, or E (which might be expected to encode EBNA genes) and Z fragments and a further 10^6 with *Bam*HI A and *Eco*RI Dhet fragments (which might encode other latent functions). The remaining clones were amplified and analysis carried out on a total of 10^7 clones from this amplified library, using individual EBV DNA fragments from both B95-8 and Raji genomic libraries (Arrand *et al.*, 1981). About 0.05% of the clones in each screening were found to be EBV specific: positively hybridizing clones were identified with B95-8 *Bam*HI fragments A, B, D, G, H, K and Ia–c (see Figure 1) and *Eco*RI fragments Dhet and I. Further, duplicate screenings were carried out with *Eco*RI Dhet and *Bam*HI A to distinguish clones that came exclusively from the *Bam*HI Nhet region as opposed to those that were derived from sequence in *Bam*HI A. No positively hybridizing clones were observed with any other fragment used, including *Bam*HI C, W, Y and E. The hybridization data obtained are shown schematically in Figure 2 and reflect a highly restricted pattern of transcription in the C15 tumour, with less than half of the genome being found to be transcriptionally active. Over 500 representative clones have to date been individually mapped onto the viral genome by restriction enzyme analysis and/or hybridization with fragments of EBV DNA. Greater than 90% of them hybridize to a single region at the 'right hand' of the genome, encompassed with the *Bam*HI Nhet, A and Ia–c fragments, the latter corresponding in part to a region deleted in the genome of B95-8. Smaller numbers of clones hybridized with *Bam*HI–H, K, B, G and D fragments. Other regions of the viral genome (including the highly repetitive *Bam*HI–W fragment) were not found during any of the screenings and therefore do not appear to be represented in the library. Further repeated hybridization screenings were carried out with regard to regions encompassing known EBNA transcripts, with probes used being derived from *Bam*HI fragments C, W, E, $e_{1,2,3}$ and K, as well as *Eco*RI fragments B and E. These experiments yielded further recombinants hybridizing to *Bam*HI K, corresponding to EBNA-1 transcripts, but none to the other EBNA-encoding regions. Reprobing the amplified library with *Bam*HI E, K and *Eco*R1 Dhet produced results essentially indistinguishable from those of the original screening. This suggests that amplification did not distort the transcription profile by under- or over-amplification of selected recombinants.

Table I. C15 transcription patterns

ORF[a]	Putative gene product[b]	cDNA abundance (see Figure 2)
BKRF1	EBNA-1	low
BGLF5	HSV-alkaline exonuclease homologue	low
LF1, LF2[c]	unknown	high
BILF1	unknown	low
BALF5	DNA polymerase[d]	medium-high
BALF4	HSV gB homologue[d]	medium-high
BALF3	unknown[d]	medium
BNLF1	latent membrane protein (LMP)	high

[a]Open reading frames (ORF) to which cDNA clones have been mapped by sequence analysis.
[b]As predicted by DNA sequence analysis and comparison with data from Baer *et al.* (1984) and Barrell (personal communication); only translation of EBNA-1 and LMP confirmed.
[c]Not present in B95-8 DNA.
[d]Probably not as transcribed but rather present as unspliced (but polyadenylated) version of 'anti-sense' transcripts (see Figure 3 and Results and Discussion).

DNA sequence analyses

More than 30 cDNA clones—containing sequences varying in size from a few hundred to a few thousand base pairs—

Fig. 4. Northern blot analyses of C15 transcripts. (**A**) Total poly(A)$^+$ RNA from the C15 tumour, 10 μg/track, separated on agarose gels and transferred to nitrocellulose (see Materials and methods) was probed with *Eco*RI I (**track 1**); *Bam*HI A (**track 2**); and *Eco*RI Dhet (**track 3**). See Figure 1 for map locations of fragments. The mRNAs detected are indicated (►) and their approximate sizes (based on comparison with ribosomal RNAs) given. Track 3, on longer exposure, appears to contain a number of discrete bands (see Table II); the major one at 3.3 kb is indicated. *Eco*RI-I and Dhet fragments were used to distinguish between transcripts from the 'left' (from *Bam*HI fragments Nhet and C), and 'right' (*Bam*HI Nhet and A) hand regions of the viral genome (see Figure 2) respectively. (**B**) Total poly(A)$^+$ RNA from the C15 tumour (**tracks 2, 4, 6**) is compared with that from the NAD-C15-STO-B lymphoblastoid line (**tracks 1, 3, 5**). The probes used were *Bam*HI fragments Ia (**tracks 1, 2**), Ib (**tracks 3, 4**) and Ic (**tracks 5, 6**). Bands corresponding to transcripts that can definitely be identified are indicated (►) and those whose identification are more tentative indicated (◁). Differential transcription expression in the Ia–c region between the two cell types is apparent. Data from Northern blot analyses of the C15 tumour are summarized in Table II. (**C**) Comparison of cytoplasmic poly(A)$^+$ RNA from the C15 tumour with total poly(A)$^+$ RNA from C15, NAD-C15-STO-B and MABA using the oligonucleotide probe, 1811. **Panel a** contains both poly(A)$^+$ and poly(A)$^-$ total cellular RNA from C15, and **panel b** contains poly(A)$^+$ C15 cytoplasmic RNA. **Panels c** and **d** contain total cellular poly(A)$^+$ RNAs from NAD-C15-STO-B and MABA respectively. The sizes of major hybridizing transcripts are indicated. (The 4.2-kb band may not be a separate species.) A similar pattern was obtained when the oligonucleotide probe 3593 (but not 3594) was used (data not shown). Each track contained 5 μg of RNA. (**D**) Comparison of RNA from C15, NAD-C15-STO-B and MABA using the oligonucleotide probe, 81. **Panels a–c** contain, respectively, poly(A)$^-$ and poly(A)$^+$ RNA from C15 and the lymphoblastoid lines NAD-C15 STO-B and MABA. Strong hybridization of bands (sizes are indicated) are observed with the virus producing line MABA, and weak hybridization of a 4.0-kb band in NAD-C15-STO-B. No hybridization is observed with C15. Each track contained 10 μg of RNA.

have to date been subcloned into M13 and subjected to DNA sequence analysis by the Sanger dideoxy method (Sanger *et al.*, 1977). This has allowed their precise location on the EBV genome to be determined and, with clones containing large segments of DNA, their identity with regard to viral functions and/or open reading frames (Baer *et al.*, 1984) to be assigned. Most of them, as predicted, could be localized onto the region of the genome encompassing *Bam*HI Ia to Nhet, as shown in Figure 3. The rest could be assigned to EBNA-1 (in *Bam*HI K), to alkaline DNA exonuclease (in *Bam*HI B-G) or to a transcript in *Bam*HI D. Only one of them, which hybridized to *Bam*HI H, could not be localized onto the genome of B95-8 EBV, suggesting either that it was non-viral in origin or, alternatively, that C15 and B95-8 DNAs differ markedly in this region. Table I summarizes the deductions allowed by sequence analysis of cDNA clones.

Table II. Size of EBV transcripts in the C15 tumour[a]

Probe	Transcript (kb)
*Eco*RI-I	3.0
	2.2
*Eco*RI Dhet	2.9–3.7
*Bam*HI-A	6.2
	5.5
	4.8
*Bam*HI-Ia (Raji)	4.6–5.0
	2.9
*Bam*HI-Ib (Raji)	4.8
	2.4[b]
	1.2[b]
	0.9[b]
*Bam*HI-Ic (Raji)	4.8
	3.5[b]
18-8 cDNA	6.2
	5.5
	4.8
	(4.2)
*Bam*HI-W, M, G, D	ND
*Bam*HI-K/*Hin*dIII-C[c]	ND

[a]As determined by Northern blot analysis.
[b]Low copy number transcripts.
[c]Probe for EBNA-1.
ND, none detected.

The analysis of clones related to EBNA-1 showed that its message was expressed from a spliced transcript that joined *Bam*HI fragments Q, U and K. The splice that joins U to K (from nucleotides 67 652 to 107 945; Baer *et al.*, 1984) was the same as that reported elsewhere for B-cells (Speck and Strominger, 1985; Bodescot and Perricaudet, 1986; Sample *et al.*, 1986) but the *Bam*HI Q to U splice is novel. The splice site sequence is:

Q↓U
5′ GTGCGCTACCGGATGGCGGGTAAC 3′

with sequence from Q corresponding to position 62 461 (in B95-8) spliced to nucleotide 67 481 in U. No clone that corresponds to the remainder of the 5′ end of the message has yet been isolated from the library. The 3′ end of the EBNA-1 transcript in C15 is the same as that identified in B-cells.

Abundant complementary ('antisense') transcript

A striking and unexpected finding was the discovery of a major set of transcripts, hybridizing to B95-8 *Bam*HI fragments I and A, which could not be correlated with the B95-8 genomic sequence (by restriction enzyme analysis of the cDNA clones) or with predicted transcription maps of EBV (Farrell, 1987). The sequence of one of them, the 1792-bp cDNA designated 18.8, was determined and found to have the structure shown (Figure 3). Based on the general consensus sequence rules for introns, 18.8 is transcribed from left to right on the linear map of EBV and has at least four exons and three introns (see inset, Figure 3 and figure legends for coordinates). Most of the long open reading frames and putative functions previously identified in this

A

```
ATG GCC GGA GCT CGT CGA CGG GCA AGG TGC[CAG CGT CAG CAG GAT GCG CCT ATA GCG CCC
 M   A   G   A   R   R   R   A   R   C [Q   R   Q   Q   D   A   P   I   A   P

GGC CTC CTC CCC TGT CGA CCA GAG GAC GCA GGA TAT CTG CAG GAT CAG GTC AGC CTC GTT
 G   L   L   P   C   R   P   E   D   A   G   Y   L   Q   D   Q   V   S   L   V

GGT GGC CGT GGG GAA GCC CTC CTC CCC CAG ACA CTC GAT ATC GAA GGC CAG GGC CTG GTA
 G   G   R   G   E   A   L   L   P   Q   T   L   D   I   E   G   Q   G   L   V

GGA GGG CCA GGA GCT GTC TTC ACG CCG GAC CGA GAG GTC GCC CAC CTC ACA GTC GTA CTC
 G   G   P   G   A   V   F   T   P   D   R   E   V   A   H   L   T   V   V   L

GAG CTC GGC GTA CGA GTC CCG GTG CTG GAG GCG GGG GAT GGC GCG GCG GCA GCT GTA CCA
 E   L   G   V   R   V   P   V   L   E   A   G   D   G   A   A   A   A   V   P

GCC AAA GGT GAC AAA GTC ATT GTC CAG GAC AAA GCG GCG CGT GGC ATC CAC GTT GGC CTC
 A   K   G   D   K   V   I   V   Q   D   K   A   A   R   G   I   H   V   G   L

AAA GAT CCG ACA CCG]TGC TTG TCT TGC AGC CAC GTG GCC ACG TGA
 K   D   P   T   P ]C   L   S   C   S   H   V   A   T   -
```

B

```
TGA GCC CCC GGG TAC GCT GTA GAA GCT GTT GAA GGA GGT CTC TAT CCA GTC GCT CGG CTC
 -   A   P   G   Y   A   V   E   A   V   E   G   G   L   Y   P   V   A   R   L

GAT GCC TGG CCA TAT CAG GGA AGT CAG GAA TGC CTT CTG GTG GGG CAG CGT ACC TGC GGC
 D   A   W   P   Y   Q   G   S   Q   E   C   L   L   V   G   Q   R   T   C   G

GTC ACA GCA GCG AGC CAG GGC CAC GTT GCT GGG TGG GGG AAA GAG CCC GCT CTC CTC CGC
 V   T   A   A   S   Q   G   H   V   A   G   W   G   K   E   P   A   L   L   R

CAG GGG CCC CGT GAT GAA GGT GTA CAG GCT GTG CGT CAG CGC GTG CAG GTG CTC CGA GCT
 Q   G   P   R   D   E   G   V   Q   A   V   R   Q   R   V   Q   V   L   R   A

CAG GGT CTG GGT AAA CAG GTG TGT TTT GAT GTA CTT GGA [ATT CTC AAA GGC GGC ACC CTC
 Q   G   L   G   K   Q   V   C   F   D   V   L   G [ I   L   K   G   G   T   L

GCC GGC GCG CCT GTC CTC CCA GGG ACC CGA GAC GAA GGC CCG TCT GTA GAG GAA GTG GTT
 A   G   A   P   V   L   P   G   T   R   D   E   G   P   S   V   E   E   V   V

GCG CAT GCG GGC CAG CTC CCA GTA GAC CAC GTC CCC CCA GAC GCG CAG GCA CAG GGT CTC
 A   H   A   G   Q   L   P   V   D   H   V   P   P   D   A   Q   A   Q   G   L

GGT CAG GGT CTC GCT CTG TTG CGC CAG GCA GGA CTG CAG CTT GGC CAG ACC CTC GGT GGC
 G   Q   G   L   A   L   L   R   Q   A   G   L   Q   L   G   Q   T   L   G   G

CAC CTG GCG CAG GTA CTG CTC CTT GCG CTT GAG CGC GTC CGA GAG GGC GCC GGA CGG GCC
 H   L   A   Q   V   L   L   L   A   L   E   R   V   R   E   G   A   G   R   A

GGG CTC TCG TGC CCC AGC CGG CCG GGG CAC CTC CGG GCT CTC CCG GGA CGC CTC CTC CTC
 G   L   S   C   P   S   R   P   G   H   L   R   A   L   P   G   R   L   L   L

GCC TCG GCC CAA CCG CTG CAT GGC TCG GTT GAG CCG CGT GTA CAG CTC GTT CCT CTT TTG
 A   S   A   Q   P   L   H   G   S   V   E   P   R   V   Q   L   V   P   L   L

CAG GAT GGC CCG GTA CTG GGG GTG CGC CGT GAA GGC GGC GGC GCA GTC CGC CTT CAG CGC
 Q   D   G   P   V   L   G   V   R   R   E   G   G   G   A   V   R   L   Q   R

CTC CAC CGC GTC GCC CGA GGA GCT GTA GAC CCC GCC GCA GAA GAG CCG CTC CGT GGC CCC
 L   H   R   V   A   R   G   A   V   D   P   A   A   E   E   P   L   R   G   P

GGG AGC CAC GGC GTC AAA CAG GTG AGT CAG CCT TGC CCC CGC CAG CGC CTC CTC GCA GGC
 G   S   H   G   V   K   Q   V   S   Q   P   C   P   R   Q   R   L   L   A   G

CCC CCG CAC CAG GGC CAG GCG ACG CTC CCG GGC AAA CAG GGC AGA GAG GCG GGA ATG GCC
 P   P   H   Q   G   Q   A   T   L   P   G   K   Q   G   R   E   A   G   M   A

GCC ACC CTC CCC CTG CCC CGT TGC ACC GAT AGC ATG GCC GCC AGA GTT CCA ATA GAG GAG
 A   T   L   P   L   P   R   C   T   D   S   M   A   A   R   V   P   I   E   E

CTC CGA GAG CTC CGC CAC CTC CGG GGG CAC TGT CGA GAA GAC GTT GTA GGT GTC CAG CGC
 L   R   E   L   R   H   L   R   G   H   C   R   E   D   V   V   G   V   Q   R

TCT GGT CGC CCC CTC TGC ]CTC CGG CCG CCC CGG GCC CGG GAC CGC GCC CTC CTC TGG GCC
 S   G   R   P   L   C ] L   R   P   P   R   A   R   D   R   A   L   L   W   A

GCC CGG CCT CGC CTT CTC CTC AGC CTC CAA CAG GTG CCC GAG CCC CGC CTG GCG GAC TTC
 A   R   P   R   L   L   L   S   L   Q   Q   V   P   E   P   R   L   A   D   F

ATT CTC AAA CAG TCC CGA GAC CGG CTC CGG ATT CAC CGG CAC CGC CAG GTG GTT ACA GGA
 I   L   K   Q   S   R   D   R   L   R   I   H   R   H   R   Q   V   V   T   G

GAC GTG GGT CCC CTC TGC CGT GGA AGG GTT GCC GTG GTT GGG CAG AAC CAT CAG CTC GCC
 D   V   G   P   L   C   R   G   R   V   A   V   V   G   Q   N   H   Q   L   A

CAC ACA GCG CCA GCA GGG CAC AGA GGT GAT GTA GAG GCG CGG GTC TGG GAT GGG ACT TAC
 H   T   A   P   A   G   H   R   G   D   V   E   A   R   V   W   D   G   T   Y

GCC CCG AAA GCG GCC CAG CAG ATC CAG GGC CCG TTC CAG GCT CTC CAG CCC CAT GGT GTG
 A   P   K   A   A   Q   Q   I   Q   G   P   F   Q   A   L   Q   P   H   G   V

AGA CAT GCA ATA AAA CAC GCT ATT GAT TCT CTT CAT TAA
 R   H   A   I   K   H   A   I   D   S   L   H   -
```

Fig. 5. Putative polypeptides encoded within the '18.8 family' of mRNAs from C15. (**A**) The cLF1/BALF5 polypeptide. This is generated from exons 1 and 2 of 18.8. The splice site is indicated (↓) and the homology with EC-RF4 (Bankier *et al.*, 1983; see Figure 3) given between brackets. (**B**) The large open reading frame (cBALF3) and predicted polypeptide corresponds to positions 159 582–160 994 in B95-8 DNA (Baer *et al.*, 1984; Figure 3). The sequence within the C15 cDNA clone, 14.1, is indicated between brackets. Another related but slightly larger clone, 18.6, included a further 276 bp of cBALF3 at the 3′ end, but its 5′ end could not be located in B95-8 DNA (see Results).

region of EBV are transcribed in the opposite orientation, as indicated. The complete sequence of 14.4, a clone related to 18.8 and containing 1749 bp, was also determined and its structure, shown schematically in Figure 3, overlapped part of 18.8 but also contained some sequence derived from LF1. Sequence data on the ends of a third clone, RIb-1, confirm that the mRNA sequence extends at least into the *Bam*HI Ib fragment, as shown; restriction enzyme analyses (not given) confirm the relationship between RIb-1 and the other two clones. The 18.8-related cDNA clones contain partial sequence homology with three viral genes found in B95-8—i.e. BALF5 (the DNA polymerase), BALF4 (homologous to the HSV major glycoprotein, gB) and BALF3, of unknown function—as well as LF2 (from the region deleted in B95-8, also of unknown function). The sequences of seven other related but unspliced cloned cDNAs were determined (Figure 3). Two of them (as shown) lie within the large intron in 18.8 and two extend transcription in the 3′ direction (including 14.1, see below). The other three correspond to the largest 18.8 exon.

Northern blot analyses

To confirm the hybridization data, and establish the sizes and strand orientations of the various mRNAs, Northern blot analyses were carried out on polyadenylated RNA from the C15 tumour. The data from several of the analyses are shown in Figure 4 and the findings are summarized in Table II. The method employed initially (Figure 4A, B) used high specific activity-labelled genomic restriction enzyme fragments from B95-8 as probes, but proved to be too insensitive to allow detection of transcripts from the *Bam*HI-K (EBNA-1) region, setting a lower limit with regard to our ability to observe transcription directly by this technique. (The EBNA-1 transcript could be detected using a more sensitive procedure involving labelled riboprobes; P.R.Smith unpublished data). Thus, transcripts represented by only a few (1−4) clones in the cDNA library, where 2×10^6 clones were analysed, were not detected by Northern blot analysis. On the other hand, mRNAs from the major transcription regions were readily observed. The data derived from Northern blot analyses using as probes *Eco*RI I (track 1), *Bam*HI A (track 2) and *Eco*RI Dhet (track 3) (see Figure 4A) are consistent with the analysis of the content of the cDNA library (Figure 2) and regions containing easily detectable transcripts were represented as multiple recombinant clones in the library. When the 18.8 cDNA clone was used to probe a Northern blot of C15 mRNA, a pattern identical with that seen using the *Bam*HI A probe (Figure 4A, track 2) was obtained (see Table II). Transcripts of 4.8 kb, the same size as the major *Bam*HI A-positive transcript, hybridized to the *Bam*HI Ib and Ic fragments as well (Figure 4B, tracks 2, 4 and 6). The 4.8-kb (and possibly

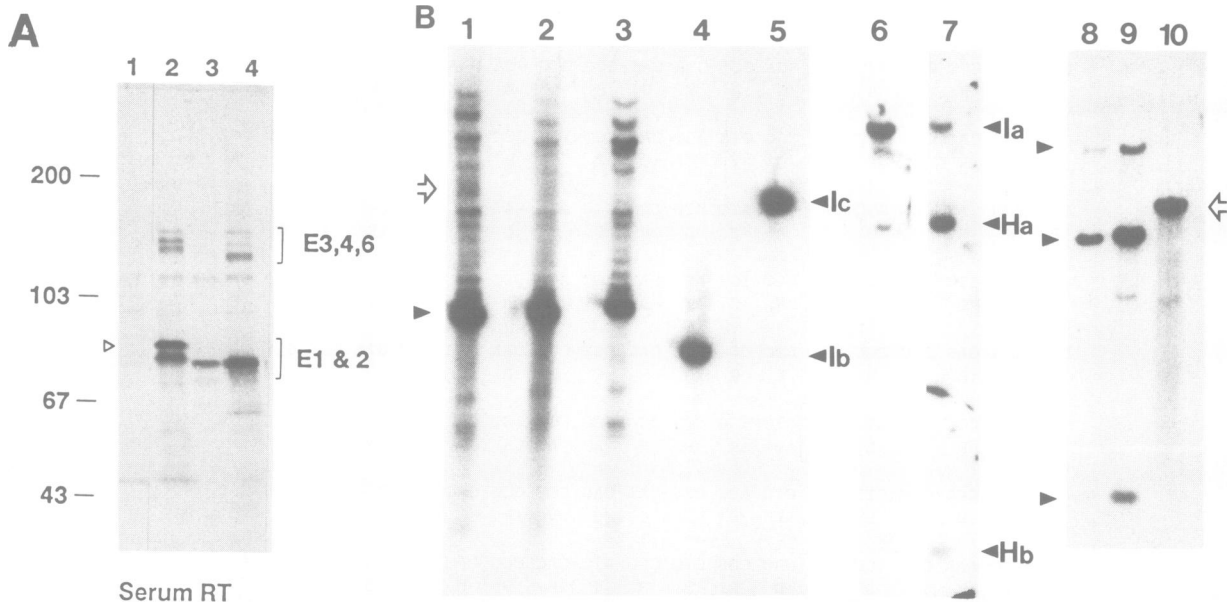


Fig. 6. Expression of 'latent' proteins. (**A**) Proteins extracted from: EBV-negative Ramos B-cells (**track 1**); 2C9 cells (**track 2**); C15 tumour (**track 3**); and NAD-C15-STO-B (**track 4**), separated on a 7.5% SDS−polyacrylamide gel, transferred to nitrocellulose and probed with serum RT (1/100) Allday *et al.*, 1988). The positions of the five EBNA species, E1, 2, 3, 4, 6, are indicated. In 2C9 cells EBNA1 (▷) has a mol. wt greater than EBNA2; in C15 and NAD-C15-STO-B, EBNAs 1 and 2 co-migrate. This has been confirmed using human serum which is specific for EBNA2 but not EBNA1 (not shown). (**B**) Southern blot analyses of *Bam*HI cleaved chromosomal DNAs using EB viral probes. The tracks contain, respectively; B95-8 DNA (**tracks 1** and **10**); Raji DNA (**track 2**); C15 tumour DNA (**tracks 3−7** and **9**); the lymphoblastoid C15 cell line, CBC-C15-STO-B (**track 8**). **Tracks 1−3** are probed with linear B95-8 virion DNA. The site of the large EBV repetitive fragment (*Bam*HI W, see Figure 1) is indicated (▶). The arrow (⇒) shows the site of the normal B95-8 *Bam*HI H and I fragments, which are apparently absent in the C15 tumour. Tracks 4−6 are probed with Raji *Bam*HI fragments Ib, Ic and Ia (for location, see Figure 1) respectively. The fragments detected in C15 DNA (indicated) are ~9.0, ~5.8 and ~2.8 kb in size, similar to fragment sizes observed in Raji cells. **Track 7**, probed with *Bam*HI H from EBV DNA, shows three bands, the larger of which is equal in size to *Bam*HI Ia (indicative of homology known to exist between *Bam*HI H and *Bam*HI I in Raji cells; Raab-Traub *et al.*, 1980). The other two, ~5.0 and ~1.0 kb in size (designated Ha and Hb), together comprising a fragment equal in size to the *Bam*HI H fragment in B95-8 DNA, are suggestive of an additional restriction enzyme site in the C15 tumour. This point is proved in tracks 8−10, where *Bam*HI digests of C15 tumour (**track 8**) and the C15-derived lymphoblastoid line (**track 9**), probed with *Bam*HI H, show two bands (in addition to the more weakly hybridizing Ia band) (▶) as compared with the single larger band (⇐) (6.0 kb) from B95-8 DNA (**track 10**).

the 5.5- and 6.2-kb) mRNA band identified in Figure 4, and that observed with 18.8, appear to be related and to represent a major transcript(s) proceeding 5′ → 3′ from *Bam*HI Ib—or possibly Ia since the Ia probe (see Table I) identifies a similarly sized transcript—into *Bam*HI A. When mRNAs from one of the lymphoblastoid lines (NAD-C15-STO-B) generated with virus 'rescued' from the C15 tumour were probed with *Bam*HI fragments Ia−c (tracks 1, 3 and 5) the major transcripts differed from those observed in C15 (tracks 2, 4 and 6).

In order to map more precisely the '18.8 family' of transcripts, and particularly to localize the ends of the large messages, a series of Northern blot analyses using strand-specific radiolabelled oligonucleotide probes was carried out. Six oligonucleotides were used (see Materials and methods and Figure 3). They were chosen in order to determine the strand specificity of the transcripts (oligos 1811 and 81) and in an attempt to localize the 5′ (oligo 3852) and 3′ ends (oligos 3382, 3593, 3594) of the transcripts. Moreover, cytoplasmic polyadenylated RNA (as opposed to total cellular polyadenylated RNA) was used to ascertain whether the mRNAs were separate species or products of incomplete splicing of a single message. The data, shown in Figure 4C and D, were consistent with the transcripts (4.8, 5.5 and 6.2 kb in size) all being derived from only one strand of the genomic DNA, that *complementary* to the strand encoding the previously identified open reading frames BALF3−BALF5 (see Figure 3). These data were confirmed using single-stranded recombinant M13 probes derived from 18.8 and 14.4 DNAs (not shown). The findings are consistent with DNA sequence results (see above) which suggested that the spliced transcripts (18.8, 14.4 and RIb-1) proceed in a 'rightward' direction on the viral genome. Further, it could be shown using a specific oligonucleotide probe (3593) that a sequence immediately preceding a polyadenylation signal on the strand complementary to BALF3 hybridized to the transcripts—whereas an oligonucleotide (3594) ~100 nucleotides 'downstream' did not—suggesting that the 3′ end of the messages terminate with a polyadenylation signal within the BALF3 region, as shown in Figure 3. An oligonucleotide (3382) from the other strand did not hybridize. A similar attempt to localize the 5′ end of the message at the nearest upstream TATA box (with 3852) was not successful and other approaches are in progress to map the promoter region.

observed when polyadenylated RNAs from NAD-C15-STO-B (or from the cell line, MABA) were probed for transcription in the *rightward* direction using the oligonucleotide 1811 (Figure 4A). Unlike the case with C15, in these cells transcripts having the opposite orientation were also observed (Figure 4B). It remains to be seen whether in these cases opposing transcripts are derived from the same cell or whether the results represent mixed cellular populations.

Polypeptides (projected) from the 18.8 family of 'antisense' transcripts

Once the strand specificity of the transcripts was known and their 3′ ends localized, two open reading frames that might correspond to translation products from this region in C15 could be identified. One of those was derived from the exons either side of the largest intron in the 18.8 (or 14.4 and RIb-1) cDNA. If translated, it would produce the polypeptide given in Figure 5A. This corresponds in part, as indicated, to a polypeptide from the EC-RF4 open reading frame identified previously by Bankier *et al.* (1983). The peptide from C15, designated c (complementary) LF1/BALF5 from its position on the genome (see Figure 3), would have different N and C termini from the hypothetical EC-RF protein, the former as a consequence of the splice and the latter a result of point mutations. (The EC-RF4 open reading frame would allow a polypeptide of 289 amino acids to be translated if no splicing occurred; the C15 cLF1/BALF5 open reading frame contains information for only 134 amino acids.)

The second open reading frame encompasses the AATAAA polyadenylation recognition site shown in Figure 3, which is the signal apparently used in the 'antisense' transcripts (see above). The sequence of this region and its putative translation product (designated cBALF3) is given in Figure 5B. The unspliced cDNA clone (14.1, see Figure 3) covers much of this region, as indicated. This open reading frame is bisected by an *Eco*RI site. It prescribes a polypeptide of 471 amino acids, but contains no obvious initiation codon. If translated it may (i) require a splice not yet identified to provide the AUG in the message; (ii) use an internal AUG codon or an alternative initiation codon; or (iii) have sequence alterations in the C15 genome (relative to B95-8 DNA) involving the 5′ region. Further sequence or other information is necessary to resolve this point. A computer search (M.Ginsburg, personal communication) has revealed no proteins from other sources that correspond to either cBALF3 or cLF1/BALF5.

Antigen expression in the C15 tumour

As a control for the transcription results and to assess whether any drift in expression occurred during tumour propagation, the C15 tumour, after being passaged further in the animal model, was subsequently examined for the presence of functions known to be associated with latent infection of B-cells, namely the six known nuclear antigens (EBNAs) and latent membrane protein, LMP, using a number of well-characterized human sera and monospecific antibodies. The findings reveal that only the *Bam*HI K-encoded EBNA-1 and LMP (not given) are expressed in the tumour. They are consistent with protein data reported elsewhere (Fähraeus *et al.*, 1988) and further suggest stability of antigen expression during limited propagation of the tumour *in vivo*. C15 virus has been rescued into continuously proliferating lymphoblastoid cell lines *in vitro* by long-term co-cultivation, using adult lymphocytes from an EBV-seronegative individual or cord blood lymphocytes (Fähraeus *et al.*, 1988). The data for EBNA gene expression are given in Figure 6A: whereas in C15 only EBNA-1 is detected (track 3), one of the C15-virus derived LCLs (NAD-C15-STO-B) (track 4), as well as a representative LCL (track 2), express EBNAs 1, 2, 3, 4 and 6. Thus, in the C15 tumour, only a sub-set of the B-cell latent genes appear to be expressed. On the other hand, no gross defects exist in the genome that prevent expression, as shown here and in the Southern blot analyses below.

Analysis of the EB viral genome in the C15 tumour

To investigate the nature of the viral DNA in C15 and ascertain whether alterations in transcription and translation between B cells and carcinoma cells might be accounted for

by changes in the viral genome, Southern blot analyses were carried out. The results (Figure 6B) show that DNA in the C15 tumour was not notably different from that found in most BL lines: using total linear virion DNA as a probe, the restriction enzyme pattern of C15 DNA (track 3) was found to differ from that obtained with chromosomal DNA from B95-8 cells (track 1) mainly in two regions, namely those containing the B95-8 *Bam*HI fragments H and I, as indicated, but to be similar to patterns seen in Raji cells (track 2). When recombinant cloned fragments Ib, Ic and Ia from Raji cells, related to B95-8 *Bam*HI I, were used individually as probes, they hybridized to bands with sizes of ~2.8, ~5.8 and ~9.0 kb respectively in C15 DNA (tracks 4–6), indicating that the EBV genome in C15 does not have the large deletion found in the marmoset B95-8 cells and in that regard is similar to viral DNA in human B-cell lines, such as Raji and Daudi (Raab-Traub *et al.*, 1980; our unpublished data). The apparent absence of the *Bam*HI H fragment in C15 was found by further analysis to be due to the presence of an additional *Bam*HI restriction enzyme site in C15 (track 7) relative to the corresponding fragment in B95-8 (track 10). Track 8 shows that this extra *Bam*HI site is retained in a C15-derived LCL line. A similar *Bam*HI pattern has been observed elsewhere in EBV isolates derived from mononucleosis patients (Lung *et al.*, 1988) and in the B-cell line MABA (Zimber *et al.*, 1986) derived from virus isolated from an NPC patient. (In the light of previous findings of clones apparently related to *Bam*HI H in the C15 cDNA library that could not be related by sequence analysis to B95-8 DNA, this region in NPC-derived materials deserves further study.) Other data (not shown) indicate that C15 DNA can be restricted to produce fragments that generally resemble their B95-8 counterparts. By hybridization analysis, the viral genome copy number in C15 appeared to be between 30 and 50.

Thus, studies of viral genomic DNA from the tumour support the notion that although there may be minor alterations in the viral genome, there are no gross alterations to explain the differences in gene expression found in the carcinoma cells relative to latent gene expression in B-cells.

Discussion

Malignant epithelial cells from EBV-associated NPCs cannot generally be successfully cultured *in vitro*—although there is a recent preliminary report of successful propagation (Chang *et al.*, 1988)—and primary tumour material for studies such as those reported here have been too scant and scarce to allow thorough characterization (see Raab-Traub *et al.*, 1985). In an attempt to overcome some of these problems, Tugwood *et al.* (1987) used NPC mRNAs as probes in Southern blot analyses of DNA derived from B95-8 genomic libraries. Surprisingly, they found no qualitative differences between the EBV hybridization patterns observed with tumour material and those obtained using probes derived from non-malignant nasopharyngeal biopsies. Their findings may reflect a sub-population of cells expressing lytic functions or the presence of interfering infiltrating lymphocytes in the tumour, but whatever the case, they emphasize the difficulties of analysing viral gene expression in NPCs specifically relating to tumorigenesis with available biopsy materials. Earlier studies (Klein *et al.*, 1974; Raab-Traub *et al.*, 1983) showed that NPCs could occasionally be propagated in nude mice and in such tumours, whereas infiltrating lymphocytes were lost, EBV gene information was retained. Transcription patterns obtained from some of these tumours were, however, complex and suggestive of 'abortive' viral infection (Raab-Traub *et al.*, 1983). Recently, Busson *et al.* (1987, 1988) were successful in propagating NPC tumour material that was negative with regard to expression of viral early and capsid antigens and therefore seemed highly suitable for studying gene expression of EBV in tightly latent epithelial cells. Before embarking on transcriptional analyses, we showed (Figure 1) that one of these tumours, designated C15, contained no linear viral DNA, a finding consistent with the absence of virions. Further, serological, genomic or immunoblot data suggested that the C15 tumour cells *in vivo* did not support a lytic infection. On the other hand, the integrity of the genome was intact (Figure 6) and prolonged culture *in vitro* in the presence of B-lymphocytes could produce lymphoblastoid lines (Fåhraeus *et al.*, 1988); the genomes in the latter resembled that found in the tumour (Figure 6B). Thus C15 provided a good source of material for studying EBV gene expression related to tumourigenesis.

cDNA libraries were therefore generated in λgt10 from polyadenylated C15 RNA and one library (described here) was found to contain $3-4 \times 10^6$ independent recombinant cDNAs—with sizes varying from a few hundred to a few thousand base pairs—of which ~0.05% were positive for EBV DNA. This library has been analysed in detail. The resulting data suggest a high degree of transcriptional regulation and restriction in the C15 epithelial tumour, the profile of which is shown schematically in Figure 2. This was derived by mapping a total of >500 individual clones onto the viral genome. Of those clones that have now been sequenced (>30), all except one (which hybridized to *Bam*HI H) could be correlated with either the EBV B95-8 (Baer *et al.*, 1984) or Raji (B.G.Barrell *et al.*, personal communication) genomes, making allowances for occasional base changes in C15 DNA.

The known viral genes localised in this manner include EBNA-1, the latent membrane protein (LMP), the alkaline exonuclease (DNase) and possibly the terminal protein (see Figure 2) (Laux *et al.*, 1988a,b)—cDNA clones that hybridize to both sides of the terminal repeats were found. Immunoblot data confirmed the expression of LMP and EBNA-1 (Figure 6A) in C15, a finding also observed by others (Fåhraeus *et al.*, 1988). It is of special interest that the EBV-specific DNase is transcriptionally expressed in the tumour in that antibodies to this protein have proved valuable as a marker for early detection of NPC in large-scale screening programmes among Chinese at risk of the disease (Yang *et al.*, 1987; Chen *et al.*, 1989). Other transcriptionally expressed regions could be correlated with BILF1 and the open reading frames LF1–LF3, present in Raji DNA but removed by a large deletion of B95-8; no function has been assigned to these genes. A puzzling finding was the apparent presence among the cloned cDNAs of transcripts for genes associated with viral DNA replication, such as the DNA polymerase (BALF5) and the EBV analogue of the HSV gB protein (BALF4). This point is reconsidered below. The overall transcription patterns identified in C15 are indicated schematically in Figure 2 and given in Table I.

A major transcript in the C15 tumour, exemplified by the three clones designated 18.8, 14.4 and RIb-1 (see Figure 3),

proved to be novel and could not be correlated with any previously described EBV transcripts (see Farrell, 1987). The complete DNA sequence analysis of clone 18.8 (1.8 kb in size, with coordinates given in Figure 3 inset and legend) suggested that it was derived from a large transcript that proceeded from *left* to *right* on the EBV genome, complementary to transcripts of genes encoding the viral DNA polymerase and the major glycoprotein, among others. Northern blot analyses (Figure 4A, see also Table II) further suggested the 18.8 clone to be related to a family of large mRNAs, 4.8 kb in size or greater. These findings, coupled with those that suggested that transcripts in the same region, but proceeding from *right* to *left* (i.e. BALF3−BALF5) existed in the C15 cDNA library, made it essential to identify the polarity of transcription over this region. To resolve this question, a number of oligonucleotides specific to each DNA strand were used to probe cytoplasmic polyadenylated C15 mRNAs on Northern blots. In addition, M13 single-stranded probes of 18.8 and 14.4 DNAs, cloned in both orientations, were used as probes. The data (see Figure 4C, D) showed: (i) that only those probes complementary to messenger RNAs which would proceed 5′ → 3′ in a *rightward* direction on the EBV physical map—i.e. antisense to the replicative functions, as well as BALF3 (of unknown function)—hybridized to the C15 mRNAs; (ii) polyadenylation must precede splicing in mRNA processing during C15 since unspliced cDNAs had the same polarity as the messages; (iii) three related transcripts (4.8, 5.5 and 6.2 kb) were identified (see also Figure 5A and Table II). Strand-specific oligonucleotide probes were also used to locate the 3′ ends of the transcripts. They showed that the AATAAA hexanucleotide nearest the 3′ end of the 18.8 clone—i.e. at position 161 013 on the genome (Baer *et al.*, 1984)—is used as a polyadenylation signal for these mRNAs. (Similar studies show that the 'TATA box' proximal to the 5′ end of the 18.8 clone, in *Bam*HI Ib, is not used as a promoter signal for any of these messages although the oligonucleotide used did reveal a smaller mRNA, ~2.4 kb in size. The precise 5′ end of the 18.8 family has thus still to be identified).

EBV therefore appears to encode transcripts, not hitherto predicted, that could be similar to the recently described family of alternatively spliced herpes simplex virus transcripts (Croen *et al.*, 1987; Spivak and Fraser, 1987; Stevens *et al.*, 1987; Krause *et al.*, 1988; Wagner *et al.*, 1988) which were originally thought to be important in controlling latency of HSV in neurons but whose function is no longer clear (Steiner *et al.*, 1989). However, whereas the latter have been found in abundance only in the nucleus, our results are obtained on apparently high copy number cytoplasmic polyadenylated RNAs, suggesting at this stage the possible existence of a protein product(s).

As shown in Figure 5, two open reading frames have been identified within the C15 18.8 family of transcripts. One of these, designated cLF1/BALF5, is small and would encode a polypeptide related to, but not identical with, the putative EC-RF4 function previously identified (Bankier *et al.*, 1983). The other, designated cBALF3 (Figure 5B), is much larger. Computer searches have not yet identified other proteins related to translation products of either of these open reading frames.

The findings from the C15 library [which show a markedly different pattern of expression in the nude mouse propagated carcinoma compared with those reported for latently infected lymphoblastoid cells or some long-cultured Burkitt's lines, like Raji (see King *et al.*, 1980; Bodescot *et al.*, 1984, 1986, 1987; Fennewald *et al.*, 1984; Weigel and Miller, 1985; Speck and Strominger, 1985; Sample *et al.*, 1986; Speck *et al.*, 1986; Joab *et al.*, 1987; Pfitzner *et al.*, 1987; Wang *et al.*, 1987; Laux *et al.*, 1988; Zhang *et al.*, 1988)] together with *in vitro* immortalization studies [which show that a subfragment of the viral genome (designated p31) is fully competent to immortalize, but not transform, primary primate epithelial cells *in vitro* (Griffin and Karran, 1984; Griffin *et al.*, 1985; L.Karran *et al.*, submitted) (see Figure 2)] lend considerable support to the hypothesis that viral gene expression in an epithelial cell environment is controlled in a manner different from that observed in B-cells. The differences observed in transcription patterns, reinforced by the observations that EBNA gene expression is rigidly controlled in tumours associated with EBV, are suggestive of cellular control over viral gene expression. The data obtained on the EBNA-1 transcription in the C15 tumour are further supportive of this notion. Moreover, this gene transcript is spliced differently from all hitherto reported B-cell EBNA-1 transcripts (Speck and Strominger, 1985; Bodescot and Perricaudet, 1986; Sample *et al.*, 1986) and includes a sequence (from *Bam*HI Q) previously identified in EBV DNA as a binding site for EBNA-1 (the so-called Region III site; Rawlins *et al.*, 1985; Jones *et al.*, 1989). These authors postulated that an EBNA-1/region III complex might block the progression of RNA polymerase and reduce the levels of EBNA-1 mRNA. Our findings allow an alternative hypothesis to be proposed wherein EBNA-1 binds to viral RNA, thereby controlling its own translation, or that of other EBNA functions (by controlling splicing). Thus, alternative modes of control could exist in the epithelial cell. Latent membrane protein (LMP) expression also differs. Whereas LMP is apparently undetectable in recently established Burkitt's lymphomas (Rowe *et al.*, 1987), it is clearly present in C15 and it has been detected on immunoblots in over half the NPC biopsies analysed to date (Fåhraeus *et al.*, 1988; Young *et al.*, 1988). The finding of complementary 'antisense' transcripts to viral replicative functions, that could be involved in tumorgenicity and possibly in viral latency, suggests an alternative mode of control. The clones from the C15 library can now be used to answer a number of the questions raised in this study as well as to examine primary NPC biopsies for transcriptional expression related to that observed in the C15 tumour.

Materials and methods

NPC C15 tumour and lymphoblastoid cell lines generated by C15 virus

The C15 tumour was derived from a primary NPC biopsy identified as a poorly differentiated carcinoma, and was maintained by propagation in nude mice as described (Busson *et al.*, 1987). Histologically, it remained unchanged from the time taken to remove infiltrating non-malignant lymphocytes (passage 3) to later (20th) passage (Busson *et al.*, 1988). The NAD- and CBC-C15-STO lymphoblastoid cell lines were established by co-culturing normal human adult (from an EBV-seronegative donor) or cord blood lymphocytes respectively with the nude mouse passaged C15 tumour (I.Ernberg *et al.*, in preparation, Fåhraeus *et al.*, 1988). These cells were cultured in RPMI-1640 medium supplemented with 10% FCS, 2 mM glutamine, penicillin (100 U/ml), streptomycin (100 μg/ml) and fungizone (0.25 μg/ml), under a 10% CO_2 atmosphere.

Other cell lines

Ramos (an EBV-negative Burkitt's cell), B95-8 (the 'prototype' EBV expressing cell), MABA (a lymphoblastoid line generated with virus from

an NPC patient) and 2C9, a spontaneous LCL from an African child which expresses EBNAs 1–6 (K.Lam and M.J.Allday, unpublished), were all grown in the same manner as the lymphoblastoid lines (above).

RNA isolation and Northern blotting analysis

Total RNA was isolated from the C15 tumour at about 16th passage in nude mice by the guanidinium/caesium chloride method (Maniatis *et al.*, 1982). Poly(A)$^+$ RNA was selected by chromatography on two sequential oligo (dT)–cellulose (BRL) columns as described by the manufacturer. Approximately 100 μg poly(A)$^+$ RNA was obtained per gram of tissue.

For Northern blotting analysis, either 5 or 10 μg of poly(A)$^+$ RNA were loaded in each lane of a 1.2% agarose–2.2 M formaldehyde gel and subjected to electrophoresis. 18 S (2.1 kb) and 28 S (5 kb) RNAs of C15 poly(A)$^-$ RNA were used as markers for electrophoresis. RNA was blotted onto nitrocellulose filters (Schleicher and Schuell) by passive transfer. Markers were detected by staining the baked filter with 0.04% methylene blue in 0.5 M sodium acetate, pH 5.2.

To detect EBV-specific transcripts, baked filters were pre-hybridized for 12–20 h at 42°C in 50% formamide, 50 mM sodium phosphate, pH 6.8, 5 × SSC, 5 × Denhardt's solution, and 0.1 mg/ml denatured calf thymus DNA. Each filter strip was then hybridized for at least 20 h at 42°C with probe prepared by nick-translation of 0.5 μg gel-purified EBV genomic DNA fragment in the presence of ^{32}P-labelled dCTP (Amersham). After washing at high stringency (0.1 × SSC–0.1% SDS, 68°C) the Northern blots were subjected to autoradiography for up to 14 days using Kodak XAR film. NAD-C15-STO-B LCL RNAs were prepared and analysed in the same fashion.

Zeta-probe blotting membranes (Bio-Rad) and either the dextran sulphate (for 5′ end labelled oligodeoxyribonucleotides) or formaldehyde (for labelled M13 probes) hybridization protocols were used, according to the supplier's instructions, to probe Northern blots with ^{32}P-labelled single-stranded probes. In the case of the oligonucleotides, these were washed first at temperatures 5°C below the T_m of the oligomer, and ultimately at the T_m (Strauss *et al.*, 1986). In some cases, filters were reanalyzed by hybridization with different probes, after removing radioactivity by stripping in boiling water containing 0.1% SDS. The oligonucleotide probes used were:

5′-A G G G T T A T A G A A G A G T C C C C C A G A (designated 81);
5′-T C A C G C G C C G T T C C A T T A T G G (designated 1811);
5′-G C C C C A T G G T G T G A G A C A T G C (designated 3382);
5′-A T G T C T C A C A C C A T G G G (designated 3593);
5′-G C T A G A G A A C T C G A G A G (designated 3594);
5′-G A T C C C T A C T A C T G C A G C A T (designated 3852).

The polarities of 81 and 3382 are 5′ → 3′, *left* to *right* and 1811, 3593, 3594 and 3852 are 5′ → 3′ *right* to *left*. Single-stranded 18.8 and 14.4 probes (see Results) were generated by cloning in M13; their orientations were confirmed by sequence analysis.

cDNA library construction and analysis

cDNA was synthesized from 10 μg C15 poly(A)$^+$ RNA template and copied as described by Watson and Jackson (1985). To generate *Eco*RI ends, oligonucleotide adaptors of the sequence

5′-A A T T C C T C G A G A G T T G C -3′
3′- G G A G C T C T C A A C G -5′

were ligated to the blunt-ended cDNA, and excess adaptors removed by chromatography on Sephacryl S-200 (Sigma). Ligation into the *Eco*RI site of the λgt10 vector (Huynh *et al.*, 1985), followed by *in vitro* packaging, yielded a library of 3–4 × 10^6 independent recombinants.

One-third of the preparation was amplified, and this amplified library was screened for hybridization to a series of DNA fragments corresponding to the entire EBV genome. At each screening, the *Escherichia coli* strain C600 Hfl was infected with 1–2 × 10^6 recombinant phage and plated on a total of 20 plastic dishes (150 mm diameter). After 8–15 h at 37°C, duplicate lifts (Schleicher and Schuell nitrocellulose) were taken from each plate, the recombinant DNA fixed to the filters, and baked as described (Allday and Jones, 1987). The filters were prehybridized batchwise for 2–5 h at 68°C in 0.1% BSA, 0.1% polyvinyl pyrrolidine (PVP), 0.1% Ficoll 40, 0.1% SDS, 7 × SSC, 50 mM sodium phosphate, 1.5 mM sodium pyrophosphate. This was followed by hybridization for 12–18 h, at 68°C, in the same buffer supplemented with 0.3 mg/ml denatured calf thymus DNA and labelled probe prepared by nick translation of 2 μg gel-purified EBV DNA fragments with [α-^{32}P]dCTP (Amersham). The filters were washed at high stringency, then subjected to autoradiography on Kodak XAR film for 4–7 days.

Four-millimetre plugs of EBV-positive clones were excised and replated at low density on 90-mm dishes. Lifts were taken from these plates and probed as described above. Single EBV-positive plaques were picked directly from these plates. DNA isolated from the recombinant phage was restricted with *Eco*RI and subcloned into *Eco*RI-digested Bluescribe cloning vector (Stratagene). The inserts of selected Bluescribe clones were subcloned into M13mp18 or M13mp93 and sequenced by the dideoxy sequencing method (Sanger *et al.*, 1977). The size of the EBV positive inserts varied from 200 to 3000 bp.

C15 DNA preparation and Southern blot analysis

DNA was isolated from the same tissue used for the preparation of RNA (see above). Following pelleting of RNA in 5.7 M CsCl, DNA was isolated from the gradient, precipitated with EtOH, resuspended in 2 ml 10 mM Tris–1 mM EDTA, pH 7.5 (TE), and dialysed extensively in 20 mM Tris–HCl, pH 7.6, 10 mM NaCl, 5 mM EDTA. The DNA sample was then treated for 3 h at 50°C with 100 μg proteinase K/ml in 0.2 M EDTA–0.5% Sarkosyl. After phenol extraction, the sample was dialysed extensively in 50 mM Tris, pH 8, 10 mM EDTA, 10 mM NaCl, then digested with RNaseA at 0.1 mg/ml. The DNA was phenol extracted once more, dialysed in TE–0.5 M NaCl, ethanol precipitated and resuspended in 1 ml TE. About 2 mg DNA was obtained per gram tissue.

For Southern blot analysis, DNA was cleaved with *Bam*HI, fragments were separated by electrophoresis on 0.8% agarose gels, transferred to nitrocellulose, and analysed with ^{32}P-labelled nick translated probes by standard procedures.

Protein extraction and Western blotting analysis

Fresh C15 tumour was disaggregated by fine cutting with a scalpel and subsequently needle-teasing in PBS + 0.02% EDTA. The cell suspension was washed in PBS and the proteins solubilized in SDS sample buffer by Dounce homogenization, sonication and boiling. Proteins from washed lymphoblastoid cells or the EBV-negative BL line, Ramos, were solubilized by sonication and boiling in SDS sample buffer. In all cases samples from ~2 × 10^6 cells were loaded for electrophoresis. SDS–PAGE and Western immunoblotting with human antisera were performed as described previously (Allday *et al.*, 1988).

Sera and antibodies

EBV-positive serum, RT, which has a specificity for EBNAs 1, 2, 3, 4 and 6 was from a normal healthy donor.

Acknowledgements

We thank Dr D.Kundu for help with screening of the library, and Dr B.G.Barrell for unpublished data on DNA sequence in Raji cells. We are grateful to Ms M.Ginsburg for carrying out computer searches and Dr I.Goldsmith for gifts of oligonucleotides. M.M.H. was funded by the US Public Health Service National Research Award no. 1-F32-CA07952 from the National Cancer Institute. We gratefully acknowledge the Cancer Research Campaign who supported most of this work.

References

Allday,M.J. and Jones,M.D. (1987) *Nucleic Acids Res.*, **15**, 10592.
Allday,M.J., Jones,M.D. and Griffin,B.E. (1987) In Levine,P.H., Ablashi,D.V., Nonoyama,M., Pearson,G.R. and Glaser,R. (eds), *Epstein–Barr Virus and Human Disease*. Humana Press, Clifton, NJ, pp. 175–177.
Allday,M.J., Crawford,D.H. and Griffin,B.E. (1988) *Nucleic Acids Res.*, **16**, 4353–4367.
Arrand,J.R., Rymo,L., Walsh,J.E., Björck,E., Lindahl,T. and Griffin,B.E. (1981) *Nucleic Acids Res.*, **9**, 2999–3014.
Baer,R., Bankier,A.T., Biggen,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G. (1984) *Nature*, **310**, 207–211.
Bankier,A.T., Deininger,P.L., Farrell,P.J. and Barrell,B.G. (1983) *Mol. Biol. Med.*, **1**, 21–45.
Bodescot,M. and Perricaudet,M. (1986) *Nucleic Acids Res.*, **14**, 7103–7114.
Bodescot,M. and Perricaudet,M. (1987) *Nucleic Acids Res.*, **15**, 5887.
Bodescot,M., Chambraud,B., Farrell,P. and Perricaudet,M. (1984) *EMBO J.*, **3**, 1913–1917.
Bodescot,M., Brison,O. and Perricaudet,M. (1986) *Nucleic Acids Res.*, **14**, 2611–2620.
Bodescot,M., Perricaudet,M. and Farrell,P.J. (1987) *J. Virol.*, **61**, 3424–3430.

Bornkamm,G., Knebel-Doeberitz,M.V. and Lenoir,G.M. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4930–4934.
Busson,P., Braham,K., Ganem,G., Thomas,F., Grausz,D., Lipinski,M., Wakasugi,H. and Tursz,T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6262–6266.
Busson,P., Ganem,G., Flores,P., Mugneret,F., Clausse,B., Caillou,B., Braham,K., Wakasugi,H., Lipinski,M. and Tursz,T. (1988) *Int. J. Cancer*, **42**, 599–606.
Chang,Y.-S., Lin,S.-Y., Lee,P.-F., Chung,H.-C. and Tsai,M.-S. (1988) *J. Exp. Clin. Cancer Res.*, **7** (supplement), 33.
Chen,J.-Y., Chen,C.-J., Liu,M.-Y., Cho,S.-M., Hsu,M.-M., Lynn,T.-C., Shieh,T., Tu,S.-M., Palmer Beasley,P., Hwang,L.-Y., Lee,H.-H., Kuo,S.-L. and Yang,C.-S. (1989) *J. Med. Virol.*, **27**, 269–273.
Croen,K.D., Ostrove,J.M., Dragovic,L.J., Smialek,J.E. and Straus,S.E. (1987) *New Engl. J. Med.*, **317**, 1427–1432.
Dillner,J. and Kallin,B. (1988) *Adv. Cancer Res.*, **50**, 95–158.
Fähraeus,R., Fu,H.L., Ernberg,I., Finke,J., Rowe,M., Klein,G., Falk,K., Nilsson,E., Yadav,M., Busson,P., Tursz,T. and Kallin,B. (1988) *Int. J. Cancer*, **42**, 329–338.
Farrell,P.J. (1987) In O'Brien,S.J. (ed.), *Genetic Maps 1987.* Cold Spring Harbor Laboratory, Cold Spring Harbor NY, pp. 99–107.
Fennewald,S., van Santen,V. and Kieff,E. (1984) *J. Virol.*, **51**, 411–419.
Griffin,B.E. and Karran,L. (1984) *Nature*, **309**, 78–82.
Griffin,B.E., Karran,L., King,D. and Chang,S.E. (1985) In Rigby,P.W.J. and Wilkie,N.M. (eds), *Viruses and Cancer.* Cambridge, Cambridge University Press, pp. 93–110.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) *DNA Cloning.* IRL Press, Oxford, Vol. I, pp. 49–78.
Jones,C.H., Hayward,S.D. and Rawlins,D.R. (1989) *J. Virol.*, **63**, 101–110.
Klein,G., Giovanella,B.C., Lindahl,T., Fialkow,P.J., Singh,S. and Stehlin,J.S. (1974) *Proc. Natl. Acad. Sci. USA*, **71**, 4737–4741.
King,W., Thomas-Powell,A.L., Raab-Traub,N., Hawke,M. and Kieff,E. (1980) *J. Virol.*, **36**, 506–518.
Krause,P.R., Croen,K.D., Straus,S.E. and Ostrove,J.M. (1988) *J. Virol.*, **62**, 4819–4823.
Joab,R., Rowe,D.T., Bodescot,M., Nicolas,J.-C., Farrell,P.J. and Perricaudet,M. (1987) *J. Virol.*, **61**, 3340–3344.
Laux,G., Perricaudet,M. and Farrell,P.J. (1988a) *EMBO J.*, **7**, 769–774.
Laux,G., Freese,K., Fischer,R., Polack,A., Kofler,E. and Bornkamm,G. (1988b) *Virology*, **162**, 503–507.
Lenoir,G., O'Conor,G. and Olweny,C.L.M. (eds), (1985) *IARC Scientific Publications*, no. 60. IARC, Lyon.
Lung,M.L., Chang,R.S. and Jones,J.H. (1988) *J. Virol.*, **62**, 3862–3866.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (eds) (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, p. 196.
Pfitzner,A.J., Strominger,J.L. and Speck,S.H. (1987) *J. Virol.*, **61**, 2943–2946.
Raab-Traub,N. and Flynn,K. (1986) *Cell*, **47**, 883–889.
Raab-Traub,N., Dambaugh,T. and Kieff,E. (1980) *Cell*, **22**, 257–267.
Raab-Traub,N., Hood,R., Yang,C.-S., Henry,B. and Pagano,J.S. (1983) *J. Virol.*, **48**, 580–590.
Raab-Traub,N., Huang,D., Yang,C.-S. and Pearson,G. (1985) In Levine,P.H., Ablashi,D.V., Pearson,G.R. and Kottaridis,S.D. (eds), *EBV and Associated Diseases.* Humana Press, Clifton, NJ, pp. 362–372.
Rawlins,D.R., Milman,G., Hayward,S.D. and Hayward,G.S. (1985) *Cell*, **42**, 859–868.
Rowe,D.T., Rowe,M., Evan,G.I., Wallace,L.E., Farrell,P.J. and Rickinson,A.B. (1986) *EMBO J.*, **5**, 2599–2607.
Rowe,M., Rowe,D.T., Gregory,C.D., Young,L.S., Farrell,P.J., Rupani,H. and Rickinson,A.B. (1987) *EMBO J.*, **6**, 2743–2751.
Rymo,L., Lindahl,T. and Adams,A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2794–2798.
Sample,J., Hummel,M., Braun,D., Birkenbach,M. and Kieff,E. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5096–5100.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5468.
Simons,M.J. and Shanmugaratnam,K. (1982) UICC Technical Report series 71, report 16. Geneva.
Speck,S. and Strominger,J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8305–8309.
Speck,S., Pfitzner,A. and Strominger,J. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9298–9302.
Spivak,J.G. and Fraser,N.W. (1987) *J. Virol.*, **61**, 3841–3847.
Steiner,I., Spivak,J.G., Lirette,R.P., Brown,S.M., MacLean,A.R., Suback-Sharpe,J.H. and Fraser,N.W. (1989) *EMBO J.*, **8**, 505–511.
Stevens,J.G., Wagner,E.K., Devi-Rao,G.B., Cook,M.L. and Feldman,L.T. (1987) *Science*, **235**, 1056–1059.
Strauss,M., Streuli,C. and Griffin,B.E. (1986) *Gene*, **49**, 331–340.
Tugwood,J.D., Lau,W.-H., Sai-ki,O., Tsao,S.-Y., Craig Martin,W.M., Shiu,W., Desgranges,C., Jones,P.H. and Arrand,J.R. (1987) *J. Gen. Virol.*, **68**, 1081–1091.
Wagner,E.K., Flanagan,W.M., Devi-Rao,G., Zhang,Y.-F., Hill,J.M., Anderson,K.P. and Stevens,J.G. (1988) *J. Virol.*, **62**, 4577–4585.
Wang,F., Petti,L., Braun,D., Seung,S. and Kieff,E. (1987) *J. Virol.*, **61**, 945–954.
Watson,C.J. and Jackson,J.F. (1985) In Glover,D.M. (ed.), *DNA Cloning.* IRL Press, Oxford, Vol. I, pp. 79–88.
Weigel,R. and Miller,G. (1985) *J. Virol.*, **54**, 501–508.
Yang,C.-S., Chen,J.Y., Liu,C.J., Hsu,M.M., Lynn,T.C., Shieh,T., Lee,H.H., Kuo,S.L. and Tu,S.M. (1987) In Levine,P.H., Ablashi,D.V., Nonoyama,M., Pearson,G.R. and Glaser,R. (eds), *Epstein–Barr Virus and Human Disease.* Humana Press, Clifton, NJ, pp. 67–69.
Young,L.S., Dawson,C.W., Clark,D., Rupani,H., Busson,P., Tursz,T., Johnson,A. and Rickinson,A.B. (1988) *J. Gen. Virol.*, **69**, 1051–1065.
Zhang,C.X., Decaussin,G., Daillie,J. and Ooka,T. (1988) *J. Virol.*, **62**, 1862–1869.
Zimber,U., Adldinger,H.K., Lenoir,G.M., Vuillaume,M., Knebel-Doeberitz,M.V., Freese,U.-K., Schneider,U. and Bornkamm,G. (1986) *Virology*, **154**, 56–66.

Received on December 12, 1988; revised on June 12, 1989