The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion similar to that observed in a non-transforming strain (P3HR-1) of the virus

M.D. Jones¹, Linda Foster², Teresa Sheedy and Beverly E. Griffin*

Imperial Cancer Research Fund, Lincoln's Inn Fields, London, WC2A 3PX, UK

¹Present address: Department of Chemistry, Imperial College of Science and Technology, London

²Present address: Department of Genetics, Stanford University, California, USA

*To whom reprint requests should be sent Communicated by B.E.Griffin

Epstein-Barr virus (EBV) DNA isolated from the frequently studied and unusual Burkitt's lymphoma cell line, Daudi, contains a 7.4-kb deletion, similar to (but larger than) that found in a non-transforming isolate of the virus, P3HR-1. A comparison of EBV sequence in Daudi cells with that from a comparable region in a wild-type, transforming strain of the virus (B95-8) indicates that at least two of the previously identified RNAs, a highly repetitive sequence, and other interesting coding or structural features should be absent in Daudi EBV DNA as a consequence of the deletion. The information removed by the deletion, as well as that which might be generated by juxtaposition of two regions of the genome that are not adjacent in most strains of the virus are discussed.

Key words: B-lymphoma/Daudi/deletion variant/EBV/ transformation

Introduction

The lymphotropic herpesvirus, Epstein-Barr virus (EBV), can be detected in latent form from early childhood onwards in most of the human population. Where infection has been delayed until adolescence, the virus can produce infectious mononucleosis (IM). Further, EBV is strongly associated with two malignancies, Burkitt's lymphoma (BL) and nasopharyngeal carcinoma (NPC). Many studies have been carried out to relate changes in the viral genome of EBV strains of different origins with the various diseases. To date, however, these attempts have proved largely disappointing. That is, regardless of the source, sufficient similarities have been observed among the different EBV strains to make it unlikely that there are disease-specific viral sub-types (Rymo et al., 1979; Bornkamm et al., 1980; Fischer et al., 1981; Heller et al., 1981). Nevertheless, considerable attention has been focussed on a sub-clone, P3HR-1 (Hinuma et al., 1967), of a BL line (Jijoye) that is the only known source of an EBV which lacks the ability to immortalise human immature B-lymphocytes, with the hope that such studies will lead to an understanding of the mechanism of cellular transformation by EBV (Miller et al., 1974; Bornkamm et al., 1980, 1982; Heller et al., 1981; Hayward et al., 1982; Heston et al., 1982; Hudewentz et al., 1982; King et al., 1982; Rabson et al., 1982; Jeang and Hayward, 1983; Stoerker et al., 1983). The conclusion reached is that a deletion in the EBV genome adjacent to the large internal repeat sequence of the virus is of importance for events that lead to initiation of cellular transformation. For example, Rabson et al. (1982) showed that the only modification detectable by restriction enzyme mapping which distinguished the non-transforming P3HR-1 strain of virus from its parental (Jijoye) transforming virus, was a deletion of $\sim 2.4 \times 10^6$ daltons of information from sequence residing within the BamHI-H and -Y fragments in the latter viral genome (for a map, see Figure 4A). The P3HR-1 viral strain is heterogeneous (Fresen et al., 1977; Heston et al., 1982), and this fact complicates the interpretation of all data obtained so far except those mentioned above (Rabson et al., 1982), in which cloned cell lines of P3HR-1 were used. In fact, studies by Gerber et al. (1982), using P3HR-1 cells from several laboratories, indicated the persistence of both transforming and non-transforming viruses. These authors suggested that, in general, virus from P3HR-1 cells which lacks transforming activity at the normal temperature (37°C) may regain it at reduced temperature (32°C), i.e., at least some of the EBV isolates from P3HR-1 may be temperature sensitive with respect to transformation. In spite of the reservations suggested by these data, studies on the P3HR-1 strain(s) of EBV provide a promising approach for analysing cellular transformation mediated by the action of EBV. The deletion present in one strain of this virus has recently been characterised by DNA sequence analysis (Jeang and Hayward, 1983).

We have been investigating the EBV of another BL line, Daudi, whose unusual properties have made it of broad general interest. Daudi continuous cell lines were established by Klein et al. (1968) from a young, male African with Burkitt's lymphoma. They were initially found to display a very strong surface accumulation of IgM molecules, without demonstrable cytoplasmic IgM. The cells carry the 8:14 chromosomal translocation typical of many BL lines (Manolov and Manolova, 1972; Zech et al., 1976; Lenoir et al., 1982) which results in some of the immunoglobulin V_H genes becoming associated with chromosome 8 (Erickson et al., 1982). They apparently have also lost the capacity to synthesise β 2-microglobulin, which makes them unique among lymphoid lines (Nilsson et al., 1973). Moreover, they have been of considerable interest in interferon studies in that they are particularly susceptible to the growth inhibitory effects of this agent (Adams et al., 1975; Gewert et al., 1981; Silverman et al., 1982). Tovey et al. (1982) report that although interferon can greatly enhance the expression of early viral antigens in Daudi cells, a similar effect was not observed in the same cells superinfected with P3HR-1.

Daudi cells contain many copies of EBV DNA (200-400 copies/cell; Adams, 1979) present mainly as episomal forms. They are naturally weak virus producer lines but can be induced to greater production with sodium butyrate (Luka *et al.*, 1979). These and other studies suggest that among the many valuable BL lines, the Daudi cell line may have particular advantages for investigating the molecular biology of EBV



Fig. 1. (A) Photograph of *Bam*HI-cleaved linear DNA from B95-8 virions (**lane b**) and episomal EBV DNA from Daudi cells (**lane c**), separated on a 0.8%/1.2% agarose gel by electrophoresis in the presence of ethidium bromide, and visualised under u.v. light. For size markers, the cosmid clone p65 (**lane a**; Griffin and Karran, *Nature*, in press), which contains *Bam*HI fragments A, C, I and W and the cosmid pHC79, was cleaved and included on the gel. Notably absent in the lane containing Daudi DNA is a band corresponding to *Bam*HI-H. Moreover, *Bam*HI-Y is also missing (see below). Novel bands among the Daudi fragments include (from top to bottom): mitochondrial DNA (Mit), a band derived from circularisation of 'ends' (N), a band that co-migrates with *Bam*HI-C (C') and a *Bam*HI I-related fragment (I'). The latter two bands are derived from a region within the *Eco*RI-C fragment present in most EBV isolates (Raab-Traub *et al.*, 1980; Fischer *et al.*, 1981), but not in the B95-8 strain. (B) *Bam*HI-cleaved Daudi DNA (**lane a**, similar to **lane c**, in A), hybridised against ³²P-labelled nick-translated DNA (Rige *et al.*, 1977) from cloned B95-8 *Bam*HI fragment H (**lane b**) and, following removal of most of this probe, against ³²P-labelled *Bam*HI fragment W (**lane c**). The arrowheads (at left) indicate sites at which normal (B95-8) fragments H and Y should appear and, (at right) the novel band (*Bam*HI-WH) formed by a deletion of 7.4-kb from Daudi DNA that fuses sequences from *Bam*HI-W to -H.

and its interaction with human B-lymphocytes.

Since the deletion found in the P3HR-1 strain of EBV has been assumed to be unique among viral isolates and to have resulted as an accidental consequence of propagation and cloning of the host cells in the laboratory (Miller *et al.*, 1974), we were surprised to find a similar, but larger, deletion in DNA derived from Daudi cells. This paper describes the deletion that exists in Daudi EBV DNA and the corresponding sequence found in the *Bam*HI-H fragment from 'wildtype' strain B95-8 virion DNA.

Results

Episomal (covalently closed circular, ccc) EBV DNA was

isolated from Daudi cells, cleaved with *Bam*HI, and the fragments resolved by electrophoresis on agarose gels. When the results were compared with data from supercoiled EBV DNA of Raji cells (not shown) or linear DNA of B95-8 virions (Griffin *et al.*, 1981), a number of differences were observed (Figure 1A). Some of the variations between B95-8 virion DNA and Daudi ccc DNA can be ascribed to the known deletion in the *Eco*RI-C fragment in the former (Raab-Traub *et al.*, 1980), and a new band which corresponds to the endjoined fragments in the latter. In these respects, EBV DNAs from Raji and Daudi cells were similar. Other discrepancies, such as a possible small decrease in size in *Bam*HI-P of Daudi EBV DNA resemble changes described for the P3HR-1 and

Jijoye strains of virus (Rabson et al., 1982). The most conspicuous and surprising difference observed was the apparent absence of the two fragments corresponding to BamHI-H and -Y in the Daudi isolate. Although a deletion in this region is known to exist in virion DNA from the non-transforming strain P3HR-1, it has not been observed so far in EBV DNA from other sources. To confirm this finding, Daudi ccc DNA was cleaved with BamHI and the resulting fragments separated by agarose gel electrophoresis, transferred to nitrocellulose, and probed with a [32P]nick-translated recombinant clone of B95-8 EBV DNA, BamHI-H (Arrand et al., 1981). Two major bands were observed (Figure 1B, lane b), one corresponding in approximate size with BamHI-B (or -C) and the other being marginally larger than BamHI-W, a fragment present in multiple copies comprising the large internal repeat found in all strains of EBV virion DNA (Rymo and Forsblom, 1978) as well as in episomal forms of the virus (see Griffin et al., 1981). To confirm that the smaller of the two fragments observed here was not simply a copy of the repetitive sequence, the nitrocellulose filter was washed to remove most of the hybridised material, then re-examined with a ³²Plabelled probe derived from a clone of B95-8 BamHI-W (Jones and Griffin, 1983). The results are shown in Figure 1B (lane c). Of the three bands now observed, only the smallest had initially failed to hybridise to B95-8 BamHI-H; it could be assigned as corresponding to the BamHI-W fragment. Since hybridisation failed to reveal the presence of either normal B95-8 BamHI-H or -Y, and since the fragment of ~3.5-kb hybridised to both BamHI-H and W (see Figure 1B, band labelled BamHI WH), it presumably represents a fusion of the normal fragments and a deletion over part of this region.

To confirm this, BamHI fragments of Daudi-derived EBV ccc DNA were cloned in the vector, pAT153 (Twigg and Sherratt, 1980) by standard procedures. Recombinant clones which hybridised to both B95-8 fragments BamHI-Y and H (note that BamHI-Y will also hybridise to BamHI-W since it contains part of the same repetitive sequence, Jones and Griffin, 1983) were selected, and one of them (designated 10G) was used to prepare DNA which was further cloned into single-stranded M13 vectors and sequenced by the dideoxy nucleotide method (Sanger et al., 1980). To compare the sequence from clone 10G with corresponding sequences from B95-8, most of BamHI-Y and all of BamHI-H from the latter virus strain were also sequenced. The primary structure of BamHI-W had been determined previously (Cheung and Kieff, 1982; Jones and Griffin, 1983). The sequence of BamHI-H is given in Figure 2. When the recombinant clone containing BamHI-H was partially digested with NotI, it was found to contain 12 complete copies of this repeat (data not shown) and these are included in the total DNA sequence of the fragment. The primary structure of Daudi clone 10G, designated Daudi BamHI-WH, is given in Figure 3. The data show that Daudi ccc DNA contains a deletion of \sim 7.4-kb in a region encompassed by BamHI fragments W, Y, and H in B95-8 virion DNA. The sequence data are analysed schematically in Figures 5 and 6. It is notable that all the sequence 'unique' to the BamHI-Y fragment in B95-8 virion DNA that is, not also found in BamHI-W - is missing in Daudi ccc DNA, as well as a presumptive leftward promoter, the long repetitive sequence characterised by the NotI repeat (Hayward *et al.*, 1982; Jones and Griffin, 1983; Jeang and Hayward, 1983), and a portion of the DS_L region from *Bam*HI-H. Overall, the deletion is somewhat larger than that which has recently been defined for the P3HR-1 strain of the virus (Jeang and Hayward, 1983).

Analysis of the primary DNA sequence of *Bam*HI-WH suggests that the deletion in the Daudi EBV genome removes a number of potential coding regions found in B95-8 DNA and deletes or rearranges some putative transcriptional control signals (Figures 5 and 6). A comparison of the DS_L sequence in *Bam*HI-H (Figure 2) with the DS_R sequence from *Eco*RI-C of Raji DNA (B.G.Barrell, personal communica tion) shows that although there are some minor sequence differences between the two homologous regions, the (leftward) putative promoter control signals are unaltered. However, the part of DS_L that is deleted in the Daudi DNA, although leaving the 'CAT' boxes intact, removes the potential 'TATA' box. From the sequence data (see Figure 2), it would appear that over a large part of the DS_L the DNA is non-coding.

Discussion

The main question to answer in considering the EBV DNA deletion in Daudi cells is whether it can be correlated with specific viral functions lost in the cells relative to properties of isolates, such as EBV from the marmoset B95-8 line, which retain the region. Alternatively, have unusual viral functions been created by the deletion?

The complete DNA sequences of B95-8 BamHI fragments W (Cheung and Kieff, 1982; Jones and Griffin, 1983) and H (see Figure 2), and the partial sequence of BamHI-Y (our unpublished data, and B.G.Barrell, personal communication), and the sequence of Daudi EBV BamHI WH fragment (see Figure 3), coupled with transcriptions and marker rescue experiments (cited below) allow us to address these questions. The data of van Santen et al. (1981) and Hummel and Kieff (1982) suggest that at least two mRNAs may be encoded in the region considered here. One, a major transcript 3.0-3.1 kb in size, appears to be derived from sequences present in BamHI fragments W, Y and H. The other, 2.5-kb in size, is derived solely from BamHI-H. Jeang and Hayward (1983) have defined the limits of the latter and shown that it includes the NotI repeat sequence within BamHI-H. No product has been directly associated with the 2.5-kb mRNA. [However, recent data of Glaser et al. (1983) indicate that the early antigen, EA-R, is encoded entirely within BamHI-H and this may be the 2.5-kb mRNA function.] When the primary DNA sequence is analysed, in a left to right direction, on a standard physical map (see Figure 4A), three areas with long open reading frames are observed. (i) The first encompasses information from BamHI fragments W, Y and H. This could encode >370 amino acids (~42 K) from BamHI-H alone, as shown in frame 2, Figure 5A. It may correspond to the 3.0 -3.1 kb mRNA mentioned above. It should be noted that although this region in BamHI-H is deleted in P3HR-1 cells, it also appears to be deleted in Jijoye cells, a line that produces transforming virus (King et al., 1982). Although no protein has yet been assigned to this area, the open reading frame in BamHI-H, if translated, would produce a polypeptide with a very high proline ($\sim 25\%$) content. Similarly the region in BamHI-Y into which it extends would yield a very 2510 2520 2530 2540 2550 2560 2570 2580 2590 2600 CESSENTICE: SECRETICES RESERVANCES SEGECACES STERETICES CECCECTES SETERESET SECCECES SEGECACES 10 20 30 40 50 60 70 80 90 100 Beatcebeta Beatateace Tebescater accectable actestate Baatesetat Tatattetaa Bacaatcace BestBacebe <u>Cetabe</u>cbat Cetatactec Abecestace tebescates cetacebaate Ctacebatta atatament cestrabese cecacebec SCCCCAAGGA CEREGAGGEC CECETEGGEC CACCEAGEGE ACCEGECGAC GEGGEGAGGE CEACCECECEA CEGGGGEGAC CEGTEGGEGAC CACCELCERC ANTICITIES BEARCORES TICCTERESE TICCERESE ACCESSED ACCESSED TICCTERESE ANTICITIES ANTICIDE ACCESSED BECCTTATTT TECCACAAAGE CCCACAAAACA SCCCCTCAGE CCAGETTEST CCAGECCACAT STCCCCCCCTC TACECCCBAC ASCACCCCACC ATTTESTCAC COBBAATAAA ACROTETICC SOGTETITET CROBBAGTCC SETCCAACCA SETCRETETA CARRESOGAE ATSCSBETE TCSTBEETEE TAAAACABTE 2710 2720 2730 2740 2750 2760 2770 2780 2790 2800 SEGETERECC CECTEGECAC CECTECECCE CCECCAGETC CTEGERCACC CEGERETICCT RECEXTCCES RESECCECCE RESECCECCE RESECCECCE RECERCESEC RECERCERCERCESEC RECERCES CTCTBTCAC ACCRAGECT ACCCCTCCAC AACLACICAT BATBCCACCA AGBCCTACCC CTCCTACCCC TCTBCCACCT BCAACCATA CBBTBCCACC BABACABTBT TBBCTCCCBA TBBBBABBB TTBBTBABTA CTACBBTBGT TCCBBATBGB BABBATBBGB AGACBBTBBA CETTBTATT BCCACCBTBC AMBRICIACE CRITCIACCA CICIBECACE CACACCACIA CICACERIAE TACAAABGEC TACEGAACII CAACCACAC CAICAACAAC AAGAACBACA TERRICISCIE COCCEPTICE RETEREMENT RECOCCECTE GECACCECTE CECCECCCC AGETOCIDES BLARCEREDE TICCTERES TICCEREDERE ACCESECEAL REFERENCE CACCECCCA CERENCEAL CETERERE RERECECE REFERENCE CETERECEC AAGEACCECE AGEACCECE 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000 GCC8880585 C5CC861865 TCC8C1865C CBC16CCC5 C1CC888166 6868156CCC CSC1866CAC CBC16CC6C5¹CCSCCA66TC C18656CAC C86CCCSCC5 SC88CCACCC A68C5ACCC5 SC8AC5868C 5A88CCCACC CCCCAC588 SC8ACCC515 SC8AC56C56 SC58CCA6 SACCC515 TITATANCAT ICCACCTATE CCATTACCCC CCICACAATT BCCACCACCA BCACABCCACG CACABCCACC ICCABGEGIC ATCAACBACCA ACAATTACA AAATATTATA ABGTBBATAC BETAATBBBB BBABTBITAA CBETBETBET CETCETBETC STETCBETB ABSTCCCCCA TAATTACTBE TISTTAATE CSENETICI SECNICOSE SEGLARCESE SCRECECES STREFICEC TERRICES DESCRECES BETREFE SECLECESTE BECCCESTE BELACESTE SCCLARGEA CCSCRAGECC CCCETCRECC ESCERESEC CACCEGESE ACCESEGES ACCESEGES CACCESTERESC CCCCETCRESC CCCETERESC 610 620 630 640 650 660 670 680 690 700 TCATCTACCC TCREGECCACCESTER CACCCATCEC RACCCCCCC AGAGECCEGE RACABASCAE BESEABOSEC ASTAGATEGE ABCCCCESTE STACCACCES TERSTACE CTESSEBBEC TTESSAGATT CTEASTTCCE STCCESECC CTETCTCETC CCCCTCCCCE AGGGECAGES SCAGGGECAS GECANGESC ANSTCCAGES ACAMECALES CAMECALES GECAGESC GACCAGESC ANSTCCAGES ASTCCAGES AS SOBSTRECC CSCTOBECK CSCTORECC TCCSCCAGET CTOBOBCACE CSEMETICT BECKTOCAC TELACTORA ATSCARGET BESECTET CCCCACCEGE SCRACCCETE SCRACSCESC SECENTCAE BACCCCSTCS SCCCCAASEA CCSCARGET ACETEBACCT TACSTCCCAC CCCCSCACCA 810 820 830 840 850 860 870 880 890 900 Teccedance andeccamete cerestette atcambana andecesso accesses andeccamete accesses tableada accesses tableada Acsence atcameter accesses tableada andecesso accesses tableada accesso accesses and accesso accesses accesses CCCCT69ACC CCAGCCCCSC CBATCCTCC CCCA68GCCH ACCOS6CTT6 CCT69TCT6 656CTCCTC 65605TCSCT 6CATCC5CCS 6TA685TC5 Segmactes batc695C2 bcta656Ac6 bsttcccccc atabase batccccca teseccaaac bsaccaaca ccccaa6ca csta656Ctc catcccaac ACACACEGEA ACCCCTAACE TITCACCAT ACATEMACCE BASTCCCATA ATASCCLASA BECTCCCATT CICTTCCCE ATEATISETA TCCTCCATC TSTBTBCCBT TEBEBATTEC AAMETBETTA TETACTTEEC CICAGESTAT TATCEBETCT CEMADESTA BABAASBEC TACTAACCAT ASBABETAE ANTEREGETS STCCRETTEC TOTECTERECE CRETACECT SATTRECES CTREMERCES GESTICCESS ACSCCCCTC COTECTOCA COCRETICE TTACCESCAC CARECEAACE AACEAACES ECATECESA COTACCECE SACTECECEAC COCRESCACE TECESSES AS ATAGACCCC CAGACTTAGA CBAAAGTTGG GATTACATTI TIGAGACAC AGAATCTCCT AGCTCAGATG AAGATTATBT GGABGBACCC AGTAAAGAC Taiciggege gittgaatci getitteaac citaatgitaa Agctettitg tetitagagga tesageteac tetatate getitga citeratata TCCCCCAGES CATECCCCRC TISCCTERIC CIBRACICA TCCERERAT CISCATCCRC TARTCORCC TERETERIS CRETCRCTE SCCCCACCCT ASSERTICC REACEBERCE ANCERACIAE SACCICARE ASSCCCTA CORRECT SCATCRC TARTCORCC SCARECES ACCERERAT ASSCCCCCCC 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 Seberacc eccessics creaticest ecactera becauses because because according ac CTORCECTE CATCEMBTAM AMACCETTRE CETECTARE AACCAATGTA TECEMANTAM ATGTTACTTE TITTBECTET AACCATTGAC ACCECTGTAC SAGEDBORAG BTAGBTCATI TITTBORAACS BBAGAGBTCB ITBGTTACAT AGBGTTTATT TACAATGAAG AAACCAAGAA ITGGTAACTG TECEBACAGT CCCCCATCSS CSSCCCASAC SACCASSCCA CSTSSACCTT CCSTCCCCC CCCCSTCACT CCCTCCCSC ACCASSACCC TSSSSCSCSS CTSACCSTCC 1210 1220 1230 1240 1250 1260 1270 1280 1290 1300 Techat Tamacangge Annangetti Asciaticca compacede cocangeare ectificana attretecti tectotore actitecca Angatagita attretecci tittecama tesatanget getere gegetectic csacesiti taaccaega acsagates tsaaaegete 3710 3720 3730 3740 3750 3760 3770 3780 3790 3800 Sestencent Scharger Thesester Sessenage Schercese Conservation Statement Carbonese Anderene Construction Sestence Econsesta Constructs Attended Conserve Conserve Schergera Statement Conserved Constructions CAACTIATAG CATGETAGGE ABCTEANCTE GGECEBTETT ACTGECEAGE CTACTECEA CTECEAGTEC ATGTICBEAC TECTATGEAT TITAAATATE GTACATICE TEGAGTTBAG ECGBGENABAA TBACGBGTEB BATGAGAGET BABGETEAGE TACAAGESTB AGSATACESTB AGSATACESTB CCCCACCCCC CACSCGASSE TCCGECCTEE GACCACGETC CETCCCTEGE GCGCSSTEGE CBAAGTACCC CCCCCTCCSE CSECETTCCT BCSECCCESEC 3910 3920 3930 3940 3950 3960 3970 3980 3990 4000 Sctebsabet Biscaccece Csascette Gacbacete Scbaccese Cesectere Tittitate Cittitate Getetete Taatacitta Cbacceteca Cacsterese Sctereare Thethere Cesteresece Seceraces Angamatas Babamates Ccabbarae Attatbarat CCCACTITIC CCCCARTCCC AACCCAAAAC CACACACACA CACTAGAATI STIAGTITAA ACASITIATI SATAGBIBGC TECTITIAG CIAATEBIB SGGISAAAAT SGGSICAGGS ITSGGITTG SIBISISTE ISTACTIAA CAATCAAATI TEICAAATAA CIATCCACE ACBAAAATES BATTAACACA 4010 4020 4030 4040 4050 4060 4070 4080 4090 4100 AGETITECTC AGGAGTERES ECTICITATI GETTAATICA BETETECAT ITTAECCCET TOBETITCAT TAABETETET CACCAGETEG ETGETACCTE ATTRETCTCS TIGCLAMAGE CATTETIAMS SECCERCAC SECARCISE SEAMACATA ACCECSECA TECCATEGES ASSTRAGES CATTETIACAT TANCENSAGE ANCENTITE SACAMCATTE CEGECESTES SESTISTACE CETTITETAT TERCESEESET ASSTRACECE TECCATETEE SACAMCATE TCCAAACGAG TCCTCACCCC CGAAGAATAA CCAATTAAGT CCACACAGTA AAATCGGGCA ACCCAAAGTA ATTCCACACA GTGGTCCACC CACCATGGAC 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200 GAGGTTATTC TATTEGGATA ACGAGAGGAG GAGGBGCTAG AGGTCCBCCG AGTTEGGBGT AGGCBGAGCC TCAGGAGGGT CCCCTCCATA GGGTTGAACC CTCCAATAAG ATAACCCTAT TGCTCTCCTC CTCCCCGATC TCCBGCCCCG AGCCCTCCGB AGTCCTCCCA GGGGAGGTAT CCCAACTEGG STABSTEAST ASTETAASAA SCATBSCEAA STABACAGET TACTITITAGA STSTABISTA CASSGCCEGE CSCAACASTS CCACCAACCC SEGSTCTEAG Catccactca tcacattctt cstaccbctt catctstcca atbaaaatct caccatcact stcccbsccc scattstcar betgettege ccccaact 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 ABEAGEGESEA EGATTEGECT CEBECCEGEAT ATACCTAETE BETBEAGECT ACAGETABET ATCCATAGEE TECCATATE CEBEAGETAT ECTAMBETCE CATTCCATSE SCARCASSEA CACTECACTA COSCCARSTC CISESSCARC CERETICCT SECECTCESE SESCARCESE SCENCESCE STERETCESC STANDSTACC CSTCSTCCCT STSACSTBAT BECSETCCAS BACCCCETCE BCCCCAABGA CCBCSABBCC CCCETCBBCC CBCCBBCSBC CACCCAGECE TCCTCCCCCT CCTAACCCSA SECERESCTA TATEGATCAC CCACCTCSEA TCTCCATCCA TAGETATCCC AAGETAATAS SACCTCCATA GGATTCBAGE 1810 1820 1830 1840 1850 1860 1870 1880 1990 1900 Teseccecte ccccectce seteseset seccccecte seccccecte cacterctees scaecceses ttccteses ttccteses 4310 4320 4330 4340 4350 4360 4370 4380 4390 4400 Secectatat Accasetses terretare tregetical accasetses aternatics accessed to tectore trectore trectore trectore accessed by another accased to the tractation of ACCOBECEAC BESECRASEC CCACCCCCCA CCEBEBECEAC CCETEBECEAC BESECRESE TECABBACCE CETEBECECC AABEACCECE ABECCCCCET 1910 1920 1930 1940 1950 1960 1970 1980 1990 2000 BCCBBECSEC CECCBETBEE TECECTBEEC CECTBECCCC CTCBBECAC CECTBECCCC CECCBETC CEECLABETC CTBEBECAC CEECCCSCC BCGECCACCC ABECBACCCS BCBACBBEEC BABECCCACC CECCACBBE BCBACCCSTS BCBACCEBE BCBACCCBE 4410 4420 4430 4440 4450 4460 4470 4480 4490 4500 CSTITIAATE GTAGAATAAC CIATAGGITA TIAACCIAET GSTBGAATAG SGTATTECAE CIGBGITATA ACCIATAGGIT ATATGAACC TAGAGBAAGG GCAAAATTAC CATCITAITE GATATCCAAT AATTGGATCA CCACCITAIC CCATAACGITC GACCCATATA TGGATATCCA TATATCTIGG ATCICCITEC 4510 4520 4530 4540 4550 4560 4570 4580 4590 4600 BAACECTATA BIBIATCCC TECEEECECT ACCECECCT CETTAC96T TBEETBAGEC CATECEECAC ECCAGECCE CB60818AC9 TB60ACECE CITG668ATAT CACATTA666 A686666686A T686686866A 666AAT6CCA AC66ACTC66 8TA6868618 6881C51866 8CECCACT6C ACC518668C 4610 4620 4630 4640 4630 4660 4670 4680 4690 4700 Cetecctac teacterica cotteeca titescare teacceater teaccate accession and the teacher CSCCSCCSCC AGSTCCT096 SCASCC0098 TTCCT09C5C TCC09609CA SCC98C58C CSCC001989 TCC8C1989C CSCT6CCCC CCC089798 SCS8C96C96 TCCA89ACCC CSTC8SCCCC AMBACC9C9 AG8CCCCCF CB8CC9CCC AG9CAACC9 SCAC0999C BASECCACC9 SCACSGAATE ACTGAACAGT GGAAACETET AAACCAGTCE ACTGECTACE ABCEGTSAAG SACCCAGTAC TEGACCBEAC ACGBAACAGE ECACCTETTA 2210 2220 2230 2240 2250 2260 2270 2280 2290 2300 Sebetagene cartaberae cartaberae cartaberae carbone carbone cartaberae acaberae acaberae acaberae acaberae cartaberae acaberae acaber 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800 STECCTCCAG CATGGTGGCT GCCTTTGGGA TGCATCACTT TGAGCCACTA AGCCCCCGTT GCTCGCCTTG CCTGCCTCAC CATGACACAC TAAGCCCCCTG CAGEGAGETC BCACCACCEA CEBAAACCCT ACETAETEAA ACTCEETEAT TCEEEECAA CEAECEEAAC BEACEEAATE BTACTETETE ATTCEEECAC 2310 2320 2330 2340 2350 2360 2370 2380 2390 2400 Teseccecte coccectes seteseses seccecete secceceec asterctese scaeceses tictesese teceseseca Accesesea sessesase coaccecce assecse costsecsa sceecesce tecaseace costsecce assecces 4810 4820 4830 4840 4850, 4860 4870 4880 4990 4900 CTAATCCATE ABCCCCCSCCT TTAGGAABCA CCACETCCCE SEGACBEAAB BBEACTTEGE ETEATTTTCT ATEYEBBEBT BBAAATATEA ECAMBAATAA BATTAGETAC TCSEEBCEGBA AATCCTTCET GETECAGEEC CCCTECCTTC CCCTGAACCC CACTAAAABA TACACCCCCA CCTTTATACT CBTTCTTATT 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 SCC888C58C CACCB8T86E TCCBCTBEEC CACTECCCE CTCBBETRE BEBETBECCC CACTEBERCAC CACTECICE CCCACCASET CTBBERCAC CSBCCCSCCE GCSBCCACCC ABBCSACCES GCSAC5888EC BASECCCACC CCCCAC586 BCSACCCSTE BCSAC565E BCSAC555E BCSAC5555 4910 4920 4930 4940 4950 4960 4970 4980 4990 5000 SGACSGCTCC TTATTAACCT GATCAGECCC GGAGTIGECT GTITEATCAC TAACCECGGG CCTBAAGAGG TIGACAAGAA BGGTCAAGGT TICGTCTGTG CCTGCCGAGG AATAATTGGA CTAGTEGBG CCTCAACGGA CAAAGTAGTG ATTSGGGCCC GGACTTCTCC AACTGTTCTT CCCAGTICCA AAGCAGACAC

A deletion in the EBV genome in Daudi cells

5010	5020	5030	5040	5050	5060	5070	5080	5090	5100	
TETTEAASEE	CASSSECTST	TEEETECATC	T66AAC86CT	TACCTC666T	AACTETTTEC	CATTAAAAAGG	TTOGGGATTA	SETTTASCCC	CTTTAGCTOC	
ACAACTTCCC	GTCCCCGACA	ACCCACGTAG	ACCTTECCEA	AT66A6CCCA	TTGACAAACG	STAATTTTCC	AACCCCTAAT	CCAAATCGGG	GAAATCGACG	
5110	5120	5130	5140	5150	5160	5170	5180	5190	5200	
CATTTCGAAC	CSSSSTETEC	AGATECAGET	CTCC666T66	SCAGGCAGTA	CSAGATETCA	CETTETETTE	TETTTECTEE	CACCCCTGTC	CTEECTETEE	
STAAASCTTS	SCCCCACACS	TCTACGTCCA	6A66CCCACC	CETCCETCAT	SCTCTACAST	6CAACACAAC	AGAAAGGAGG	GTSSSSACAS	GACCGACACC	
5210	5220	5230	5240	5250	5260	5270	5280	5290	5300	
CAAATGCGAC	CCTCATAGAG	TTETETTTCA	GETCTETETC	CTETTTTECE	STSSETTATT	TETTECETCA	STSTTTSCCA	SCTTATTTCC	CCASTTTTCA	
6TTTAC6CT6	GGAGTATCTC	AACACAAAAGT	CCAGACACAG	GACAAAACGC	CACCCAATAA	ASAASSSAGT	CACAAACEET	CGAATAAAGG	GETCAAAAST	
5310	5320	5330	5340	5350	5360	5370	5380	5390	5400	
CGTACTOGGG	CCT6T66ACA	CCTGAGGGAG	CEECCETTEE	TESETATETE	TTEGAATTEC	TCCCACCCTC	AATTTTCGCT	TECCTTCTTC	CCTTGTTAAC	
SCATGACCCC	GGACACCTGT	GGACTCCCTC	SCC66CAACC	ACCCATACAC	AACCTTAACG	ASS61666AS	TTAAAAGCGA	ACGGAAGAAG	SSAACAATTS	
5410	5420	5430	5440	5450	5460	5470	5480	5490	5500	
CTGATAGCAT	AGCCTCTAGG	TTTCCTTGTA	GETCTETTTE	GETTTETTEE	TTCACGTGGT	GCTAACTTGA	ATTTTTT66T	TTTCTAGTTC	CCTCTTAATT	
GACTATCETA	TC66A6ATCC	AAAGGAACAT	CCAGACAAAC	CCAAACAACC	AAGTGCACCA	CGATTGAACT	TAAAAAACCA	AAAGATCAAG	GGAGAATTAA	
5510	5520	5530	5540	5550	5560	5570	5580	5590	5600	
ACATTTETEC	CAGATCTTGT	ABAGCAAGAT	GECCTATICA	ACAAGEGAGA	TACTETTAEC	CCTETETATA	CSEGACAGTC	STGT6CAT66	AAATSGTACC	
TSTABACACS	STCTASSACA	TETESTICIA	CCREATAART	TRITCCCTCT	ATRACAATCE	SBACACATAT	SCCCTSTCAS	CACACGTACC	TTTACCATES	

570	5690	5680	5670	5660	5650	5640	5630	5620	5610
GAGGAGATA	TETETTECTT	TECETTATCA	ACTETAETTC	SCCAGAGGAC	TCCGCCTTTC	GAAACACCTC	ASCASCAAGA	TETTEGASCT	CTOCATCCTG
CTCCTCTATI	ACACAACGAA	ACCCAATAGT	TGACATCAAG	CEETCTCCTE	ASSCEGAAAG	CTTTGTGGAG	TCETCETTCT	ACAACCTCGA	SACGTAGGAC
580	5790	5780	5770	5760	5750	5740	5730	5720	5710
TAGAGATAT	TCAGTATTTT	GGATTTTAAC	ATGTOGATCT	CACACCEAAC	ATTTATAACA	CTTEGAACAG	TTTACAGAAA	TTCAGAGACA	TTGAACGAAA
ATCTCTATA	AGTCATAAAA	CCTAMAATTG	TACACCTAGA	eteteectte	TAAATATTET	GAACCTTETC	ANATETCTTT	AAGTCTCTGT	AACTTECTTT
5900	5890	5880	5870	5860	5850	5840	5830	5820	5810
TCCTTACTAT	ACCAGTCTAC	TTETETETA	CTECASSACA	SCATECATEC	AT66CCT66T	GTTGGCCTGG	TTEEECECEC	GACCCAAGCC	TCACCETEEA
ASSAATSAT	TESTCASATE	AACACAACAT	GACETCCTET	CSTACGTACG	TACCEGACCA	CAACCGGACC	AACCCGCGCG	CTEGETTCEE	AGTGGCACCT
6000	5990	5980	5970	5960	5950	5940	5930	5920	5910
GAAGACAAC	TACATTAATT	SCORCTOSTC	CATCAACAGG	TEETTEEATT	AAGGCCTSGA	GAAGCCAGCS	TOGGATETTA	TETCASTTCS	STTETEGACC
CTTCTSTTG	ATETAATTAA	CECCEACCAE	STASTISTCC	ACCAACCTAA	TTCC66ACCT	CTTCGGTCGC	ACCCTACAAT	ACAGTCAAGC	CAACACCTGG

Fig. 2. The 6005-bp complete DNA sequence of BamHI-H from B95-8 EBV DNA, shown in the 5' - 3' direction that corresponds to the conventional restriction enzyme map (Skare and Strominger, 1980). The BamHI site (repeated) is underlined. Bornkamm et al. (1980) report a second BamHI site in EBV DNA isolated from the M-ABA cell line in the region covered by the BamHI-H fragment; this results in a cleavage of 'normal' BamHI-H into two unequal-sized fragments. A change of the CG base pair to TA at position 1088 would account for the M-ABA data. The location of the small repetitive sequences (1) and the DS₁ sequence ($\mathbf{\nabla}$) are indicated. In our recombinant clone, 12 complete repeats were observed by partial digestion with NotI; the DNA sequence further identifies one incomplete (30%) repeat. This is similar to the observation of a non-unit repeat number for the large (BamHI-W) internal repeat (Hayward et al., 1982). A sequence comparison of the DS_L with the homologous DS_R from EcoRI-C (Bornkamm et al., 1980; B.G.Barrell, personal communication) reveals 99% homology between the two, with an overall difference of 1 bp (1044 relative to 1045 bp, respectively), and 11 mismatches. The site of the deletion in Daudi DNA (nucleotide 3978, see Figure 3) which removes a part of the DS₁ is indicated (box). The corresponding deletion in EBV DNA from P3HR-1 cells (Jeang and Hayward, 1983) occurs at nucleotide position 3599. There are a number of changes between the DNA sequence as determined here and the part of BamHI-H sequence recently reported (Jeang and Hayward, 1983). In seven cases, bases were not identified by Jeang and Hayward; their assignment is given here. In 10 cases, there are single base alterations between the two sequences. Since none of these changes results in the deletion or insertion of a termination codon they may be assumed to be of little significance. It should be noted, however, that there are some changes with respect to potential initiation codons, which could prove to be meaningful. Differences that produce frame-shifts are found, as noted: (i) position 4088-4089 bottom line, CA, is given as N by Jeang and Hayward (JH, position 10); (ii) position 3956-3958, CGA, is given as GT (JH, 140-141); (iii) position 2780, G, is AG (JH, 317-318); (iv) position 3637-3638, CC is C (JH, 460); (v) position 1550-1551, CC, is C (JH, 2420). One further and possibly significant difference is in the number of NotI repeats observed in BamHI-H. We find 12.3 as against 11.3 previously reported. This difference would be significant if the region is translated

6010 TTCCTBGATC

proline-rich translation product (data not shown). (ii) The second has potential coding sequences in all three reading frames, encompassing the NotI repeats (Figure 5A, nucleotides 1700 to \sim 4000). Of these, the sequence in frame 1 covers a region that could linearly specify a polypeptide of >560amino acids. However, transcriptional evidence suggests that the protein actually encoded by the NotI repeat region is derived from the other strand of the DNA (Jeang and Hayward, 1983). The deletion in the EBV genome in Daudi cells removes all of the open reading frames (Figure 5B) considered above. (iii) The third lies between 5523 and 6011 nucleotides (Figure 2). It has an ATG at position 5529 and allows the expression of a 161 amino acid polypeptide (see frame 3, Figure 5A). The ATG codon fulfils the criteria for an initiating codon (Kozak, 1981) since it is preceded by the sequence AAG and followed by G. This open reading frame continues into the adjacent BamHI-F fragment and can be seen to be conserved in the BamHI-WH fragment from the Daudi isolate. The weakness in suggesting this region as being actively transcribed is that the putative promoter sequences associated with it (TCAAT, 'CAT' box nucleotide position 5369 and GATA, 'TATA' box, position 5403) are not those commonly considered to be strong promoters. This might, on the other hand, explain the fact that no major mRNA has yet been mapped to this region. One interesting consequence of the deletion is that transcription of this open reading frame in Daudi cells could conceivably come under the control of the strong promoters in BamHI-W (GCAAT and TATA, see Jones and Griffin, 1983), probably normally used to transcribe the 3.0-3.1 kb mRNA (van Santen *et al.*, 1981). Thus, the deletion may potentially result in increased amounts of a protein expressed in part from the end of *Bam*HI-H, which might provide an explanation for some of the unusual properties associated with Daudi cells. This possibility is being explored.

From a similar analysis of the sequence on the other strand of the DNA (Figure 6), it can be seen that the only large open reading frames are the three in the region that encompasses the NotI repeat. The longest uninterrupted reading frame would accommodate a polypeptide nearly 700 amino acids long. The data of Jeang and Hayward (1983) map the 2.5-kb mRNA to this region of the genome. Our sequence data can accommodate this mRNA if the potential 'CAT' (either CCAAT at position 4032, or CCACT at 4018, Figures 2 and 6) and 'TATA' (GATA, position 3970) boxes are utilised as promoters to specify a message with a cap site (CC) at position 3941, an initiation codon (ATG) at position 3867 and a polyadenylation signal at 1470. Such a message would be transcribed initially from frame 3, but would require a splice to remove the termination codon (TGA) at position 3762. There are potential splice sites (Mount, 1982) within the sequence though none have yet been directly identified. Although the region around the NotI repeat is deleted in both P3HR-1 and Daudi DNAs, the former retains the putative weak promoters whereas the latter loses the GATA (TATA) box (see Figure 5). Based on these data, possible functional differences between P3HR-1 and Daudi EBV DNAs exist.

On first consideration, it would seem bizarre that a major

10	20	30	40	50	60	70	80	90	100	1810	1820	1830	1840	1850	1860	1870	1980	1890	1900
BGATCCCCCC	Accesccctt	CTCTCTGTCC	CCCTBCTCCT	CTCCAACCTT	Cectccaccc	Tagaccecag	CTTCT08CCT	CCCC99997CC	Accassccas	TCCGCCCCTA	Tatacca ss t	Gestegaect	Asstassatt	Cagetaggtt	CCTACT8886	Taccccccta	CCCTACCTTA	ABBTBCSCCA	CCCTTCCTCC
CCTABGBGGG	Tseccsseaa	Gasagacags	BEEACEAEEA	Baggttgbaa	Bcsagetgeg	Atct9998tc	GAAGACC8GA	66692229666	Testccsstc	AGGCGGGGAT	Atatsstcca	CCCACCTCGA	Tccatcctaa	Stegatecaa	GGATGACCCC	At soccc at	Begategaat	TCCACSCSGT	SBSAACGAGG
110	120	130	140	150	160	170	180	190	200	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
CCGGAGGGAC	CCCOOCAGCC	Ceeeceaetc	BCCTTCCCTC	TCCCCTOSCC	TCTCCTTCCC	BCCTCCCACC	CGA9CCCCCT	Cascitiscot	CCCCACCSSS	TTCCGTTTTA	Atestasaat	AACCTATAGG	TTATTAACCT	Agtggtggaa	Tabbetatte	Casctsssta	Tatacctata	Getatataga	Acctagagga
GGCCTCCCTG	OGOCCOTCOO	GCCCSCTCAE	CSSAAGGEAG	Assessacces	AGAOGAAGGG	CBGAGBETBE	9CTC96986A	Stobaaceba	BSSST89CCC	AAGGCAAAAT	Taccatctta	TTGGATATCC	AATAATTGGA	Tcaccacctt	Atcccataac	Stcsacccat	Atat oga tat	Ccatatatct	Tegatetet
210	220	230	240	250	260	270	280	290	300	2010	2020	2030	2040	2050	2060	2070	2080	2090	2100
TCCATCAGGC	C89CC89A98	BACCCC88C8	BCCCBBTBTC	Astecceet	SCAGCCGCCC	Astetees	TCCAGGCAAS	BBCBCCABCT	TTTCTCCCCC	A656AACCCT	Ataststaat	CCCTCCCCCC	CCTACCCCCC	CCTCCCTTAC	Bettecctba	BCCCATCCCC	Caccccaeca	CCCC8886518	Acetegcacc
AGGTAGTCCS	SCC69CCTCC	CT8866CC8C	CBBBCCACAG	Teassessa	CBTCBGCGGG	Teasasacss	AggtCCgttc	CCBCSBTCBA	AMAGAGGGGG	TCCCTT666A	Tatcacatta	S66A666666	66AT566666	SGAGGGAATS	Ccaacsbact	CSSSTABSSS	Gtebestcst	6666CCCCAC	Tecaccetee
310	320	330	340	350	360	370	380	390	400	2110	2120	2130	2140	2150	2160	2170	2180	2190	2200
CAGCCTGAGG	CCCAGTCTCC	Tetecactet	CTOTAAABTC	Caecticcea	CSCCCSTCCA	CB9CTCCC96	GCCCASCCTC	BTCCACCCCT	CCCCACBETS	CCBC6TGCCT	Tactbacttb	TCACCTITICC	ACATTIGETC	AGCTGACCGA	Tectceccac	TTCCT886TC	Atsacctosc	CT&T&CCTT&	TCCCGT86AC
STCSBACTCC	BEBTCABASE	Acaceteaca	BACATTTCAB	Stcsbasset	SCSSSCASST	9CC8A969CC	CSSGTCSGAG	CASETBOGGA	SEEE TECCAC	SSCSCAC96A	Atbactbaac	ASTOGAAACS	TETAAACCAS	TCGACTGGCT	Aceaeceete	AAGGACCCAG	Tactosaccs	BACAC&BAAC	A666CACCT6
410	420	430	440	450	460	470	480	490	500	2210	2220	2230	2240	2250	2260	2270	2280	2290	2300
Gacaggeceet	CTSTCCACCC	Beeccatccc	CSCCCCCTG	Tetccacccc	Agtcccatcc	Accesses	TTATOTGACC	CTTBGGCCTS	Beteceata	AATGTCCCTC	Cagcstssts	Sctsccttts	86AT8CATCA	CTTTEAGCCA	CTAAGCCCCC	Sttectcecc	TTOCCTOCCT	Caccateaca	Cactaageee
Ctotecggga	SACASST888	CCCBETAGEE	GCSGGGGAC	Acagetegee	Tcassacass	TCCCCCCTGA	AATACACTOG	GAACCCSGAC	Ceaegestat	TTACAGGGAG	Stcscaccac	CSACSGAAAC	CCTACSTAGT	GaaactCeet	Gattcgssss	Caacbaecee	AACEGACOGA	Stestactst	Gtgatteggg
510	520	530	540	550	560	570	580	590	600	2310	2320	2330	2340	2350	▼ 2360	2370	2380	2390	2400
Gactcccatg	TAAGCCTGCC	TCGAGTAGGT	GCCTCCAGAG	CCCCTTTTGC	CCCCCT68C6	GCCCASCCCS	ACCCCC888C	Occcccaaac	TTTGTCCAGA	Ctgctaatcc	Atgagecceg	CCTTTABBAA	Scaccacetc	CC6666AC96	AAGGEGEACTT	GGGGTGATTT	TCTATETEES	Betebaaata	TBABCAAGAA
Ctgagggtac	Attc88AC66	ABCTCATCCA	CBGAGGTCTC	BGGGAAAACG	S9566ACC8C	C996TC966C	T88899CCC6	Cooosottto	AAACAGETCT	Sacgattagg	Tactegggge	BBAAATCCTT	Ceteetscae	66CCCCT6CC	TTCCCCTGAA	CCCCACTAAA	Agatacaccc	Ccacctttat	Actebitett
610	620	630	640	650	660	670	680	690	700	2410	2420	2430	2440	2450	2460	2470	2490	2490	2500
Tetccasses	TCCCC SAGGS	TBABBCCCAS	CCCCCTCCC9	CCCCTGTCCA	Ctocccoot	CCCCCCABAA	BCCCCCAAAA	Stabagectc	ABBCCATECE	T aaggacssc	Tecttattaa	CCTGATCAGC	CCC66A6TT6	CCT&TTTCAT	Cactaacccc	506CCT6AA8	Absttbacaa	Gaabbetcaa	Betttcstct
Acagetcccc	AssesctCCC	ActccBBBTC	BBBBBBBBBBBBBBBBBBBBBBBBBBBB	SEGGACAGGT	Gacogogica	BBBBBBB	CSSSSSTTTT	Catctccbag	TCCSETACEC	Attcctsccs	Aggaataatt	GGACTAGTCG	666CCTCAAC	Seacaaasta	Stsattssss	CCC66ACTTC	Tccaactstt	Cttcccaett	Ccaaagcaga
710	720	730	740	750	760	770	780	790	800	2510	2520	2530	2540	2550	2560	2570	2580	2590	2600
Cocctotca	CCAGGCCTGC	Caaagagecca	Gatctaagge	CeeeAsAsec	ASCCCCAAAS	CBBBTBCABT	AACABGTAAT	CTCT86TA6T	Gattt9Bacc	Stststtsaa	SSSCASSSSC	Tettgeetec	Atctebaace	BCTTACCTCS	Setaactett	Teccattaaa	AGGTTGGGGA	TTAGGTTTAG	CCCCTTTAGC
Scossacagt	GBTCCSGACB	Stitctcsst	Ctagattccb	BCCCTCTCCB	TC9696TTTC	BCCCACBTCA	TTBTCCATTA	BASACCATCA	Ctanacct96	Cacacaactt	CCCSTCCCCS	Acaacccace	Tagaccttec	CGAATGGAGC	Ccattbacaa	Aceetaattt	TCCAACCCCT	AATCCAAATC	Begganatcs
810	820	830	840	850	860	870	980	890	900	2610	2620	2630	2640	2650	2660	2670	2680	2690	2700
CBAAATCTBA	Cactttagas	CTCT998896A	CTTTAAAACT	Ctaaaaatca	AAACTTTABA	Becsaateee	CBCCATTTTS	TCCCCAC9C8	CBCATAATBG	TBCCATTTC9	AACCBBBBBTB	Tecagateca	BetctccBBB	TBBBCABBCA	Stacbagats	TCACOTTOTO	TTOTCTTTCC	TCCCACCCCT	BTCCTBBCTB
BCTTTABACT	Stgaaatctc	BA6ACCTCCT	Gaaattttga	Gatttttagt	TTTBAAATCT	CCBCTTACCC	GCSSTAMAAC	A666678C9C	SCBTATTACC	ACBSTAAASC	TTBBCCCCAC	Acetetacet	Ccabageccc	Accesteest	Catectctac	Agtocaacac	AACAGAAAGG	A686T6666A	CASBACCBAC
910	920	930	940	950	960	970	990	990	1000	2710	2720	2730	2740	2750	2760	2770	2790	2790	2800
Cebacctase	CCTAAAACCC	CCASSAAGC8	Getctatget	Teectecect	Getgetatet	TTA GAGGGGA	AAAGASGAAT	AABCCCCCAB	Acagogaaat	TGGCAAATGC	Gacceteata	Sasttststt	TCAGGTCTGT	Stcctstttt	GCGGTGGGTT	Atticticcc	TCAGTGTTTS	CCASCITATI	TCCCCAGTTT
Scctsbatcc	SGATTTTSGG	Setccttcbc	Ccagatacca	Acceacecea	Cbacsataba	AATCTCCCCT	TTTCTCCTTA	TTCBGGGGTC	Tetcccctca	Accstttacs	Ctoccata	Ctcaacacaa	AgtCCAGACA	Cassacaaaa	CGCCACCCAA	Taaasaasse	AGTCACAAAC	Setcbaataa	Aggggtcaaa
1010	1020	1030	1040	1050	1060	· 1070	1080	1090	1100	2810	2820	2830	2840	2850	2860	2870	2880	2890	2900
BBBCTTBTTT	Steacttcac	CAAABETCAE	9900000099	Bettcscett	BCTAGGCCAC	CTTCTCABTC	CASCOCSTIT	Acetaagcca	BACABCANE	TCACBTACTG	9880CCT8T86	Acacctgagg	Baocesccet	Teotesetat	Stottobaat	Tectccacc	CTCAATTTTC	Octtoccttc	TTCCCTTBTT
CCCBAACAAA	Cacteaasts	STITCCAETC	00999110000	Ccanecscaa	CBATCCSGTS	SAMEMETCAS	STCOCSCAAA	Tgcattceet	CTETCETCEE	AGTGCATGAC	CCC66ACACC	Tgtggactcc	Ctcsccsoca	Accacccata	Cacaacctta	Acsassstes	Gagttaaaag	CGAACSBAAG	AASSGAACAA
1110	1120	1130	1 <u>140</u>	1150	1160	1170	1190	1190	1200	2910	2920	2930	2940	2950	2960	2970	2990	2990	3000
MATTETCAST	TCTABOBAGE	BEGACCACTS	CCCCTB <mark>ETAT</mark>	MANETEETCC	TBCAGCTATT	TCTBGTCBCA	TCAGAGCSCC	ABBABTCCAC	Acaaatstaa	AACCTGATAG	Catageetet	AGGTTTCCTT	Stagstetst	TTGGGTTTGT	Teettcacet	Getectaact	Tgaatttttt	Setttictae	TTCCCTCTTA
TTAACASTCA	AGATCCCTCC	CCCTBETGAC	BEBBACCATA	TITCACCABE	ACETCBATAA	AGACCABCST	AGTCTCSCSG	TCCTCABBTB	Tstttacatt	TTGGACTATC	Statebaga	TCCAAAGGAA	Catecagaca	AACCCAAACA	Accaaeteca	Ccacgattga	Acttaaaaaa	Ccaaaseatc	AAGGGAGAAT
1210	1220	1230	1240	1250	1260	1270	1290	1290	1300	3010	3020	3030	3040	3050	3060	3070	3080	3090	3100
Sagggggtct	TCTACCTCTC	CCTAGCCCTC	CBCCCCCTCC	AAGGACTCOG	Occcaettic	TAACTTTTCC	CCTTCCCTCC	CTCSTCTT9C	CCT8C8CCC8	Attacatte	Teccagatet	Tetabascaa	Satgecctat	TCAACAAGGE	Agatactett	ABCCCTBTGT	Atacessaca	BTCGTGTBCA	Teganateet
Ctcccccaga	Agatsbagag	BGATCBBGAG	BCBBBBBABB	TTCCTGAGCC	Cossicaaas	Att saaags	SGAASSEAGS	Bascasaacs	SBAC8C886C	Taatstaaac	Acegtetaga	Acateteett	Ctaccegata	AGTTGTTCCC	Tctatgacaa	TCGBGACACA	Tateccctet	CABCACACGT	Acctttacca
1310 9990CCACCTT CCC96T96AA	1320 Catcaccetc Stastsscas	1330 Octoactcco Coactoago	1340 CCATCCANSC SSTAGSTICS	1350 CTAGGGGAGA BATCCCCTCT	1360 CCSAAGTSAA 1360	1370 8800000088600000000000000000000000000	1380 CAACCC88CC 6TT888CC99	1390 C666CCCCCC 8CCC9666666	1400 Setatcseec Ccatascccb	3110 Accetecate Teseacetae	3120 Ctetetter Gacacaacct	3130 BCTABCABCA CBATCETCET	3140 Agagaaacac Tctctttgtg	3150 CTCTCC8CCT SAGA86C86A	3160 TTCGCCAGAG AASCGGTCTC	3170 Bacactetae Cteteacatc	3180 TTCTGCBTTA AAGACGCAAT	3190 TCATETETE ABTACACAAC	3200 Cttbabbaba Baactcctct
1410	1420	1430	1440	1450	1460	1470	1480	1490	1500	3210	3220	3230	3240	3250	3260	3270	3280	3290	3300
Casagetaas	Togactitaa	TTTTTCTBC	TAAGCCCAAC	Actecacae	ACCCAGGCAC	Acactacaca	CACCCACCC8	TBEBETCTCT	BTBTAATACT	TAATTBAACS	AAATTCABAB	Acatttacag	AAACTTGGAA	Cagatitata	Acacacace	AACATGTOGA	TCT66ATTTT	AACTCABTAT	TTTTABABAT
Stctccattc	Acctgaaatt	AAAAAAGACG	ATTCSGGTTG	Tgagetgetg	Tegetccete	Teteatetet	818881888C	ACCCCAGAGA	Cacattatga	Attaactt6C	TTTAAGTCTC	Tetanatetc	TTTGAACCTT	Stctaaatat	Teteteec	TTGTACACCT	Asacctaaaa	TTGAGTCATA	AAAATCTCTA
1510	1520	1530	1540	1550	1560	1570	1580	1590	1600	3310	3320	3330	3340	3350	3360	3370	3380	3390	3400
TTAAGGTTTG	CTCABBAGTS	Segecticit	Atteettaat	TCAGETETET	Cattitagec	Cettoeettt	Cattaasete	Tetcaccage	Teesteetac	Atticaccet	Seagacccaa	60011686606	Cecetteecc	Tegatebcct	Getecateca	Tecctecaee	Acattetett	Staaccastc	Tactcettac
AATTCCAAAC	BAGTCCTCAC	CCCCBAAGAA	Taaccaatta	AGTCCACACA	Stamatess	Scaacccaaa	Staattccac	Acaeteetcc	Acceaceate	Taaasteeca	CCTCT866TT	06644000060	BCBCAACCEE	Acctaccega	Ccacetacet	Aceeacetcc	Tetaacacaa	Cattestcas	At gaggaa tg
1610	1620	1630	1640	1650	1660	1670	1680	1690	1700	3410	3420	3430	3440	3450	3460	3470	3480	3490	3500
CTSGAGGTTA	TTCTATTOOG	Ataacbagab	BAGBAGGOGC	Tagagetccb	CBAGATTT99	Setaggessa	Occtcassas	Betcccctcc	Ataggettga	Tatsttstbs	Acctetcaet	TCGTGBBBATG	Tt abaascca	GCGAAGBCCT	98AT66TTB6	Attcatcaac	A9990290CTG	BTCTACATTA	Att baabaca
GACCTCCAAT	Angataaccc	Tattoctctc	CTCCTCCCCB	Atctccaggc	SCTCTAAACC	Ccatccscct	Cosastcctc	Ccaeegaage	Tatcccaact	Atacaacacc	Tegacaetca	AGCACCCTAC	Aatcttcggt	CGCTTCCGGA	CCTACCAACC	Taastastts	TCCCGCCGAC	CAGATGTAAT	Taacttctgt
1710 ACCASSAGGS TGGTCCTCCC	1720 Benebatter Cotcotnacc	1730 Octococco Ceaseceeee	1740 BATATACCTA CTATATOGAT	1750 BTODBTOBAS CACCCACCTC	1760 CCTABABETA SGATCTCCAT	1770 Betatccata Ccatagetat	1780 Sesticcati CCCAASSTAA	1790 Atcctobage Taggacctcc	1800 Tatcctaagc Ataggattcs	3510 Acattcctee Tetaabbacc	ATCC TABE								

Fig. 3. The 3508-bp DNA sequence from a novel *Bam*HI fragment found in the Daudi EBV strain (designated Daudi *Bam*HI-WH) that corresponds to a deletion of sequence, found at the junction of *Bam*HI fragments Y and H in B95-8 DNA, which results in a juxtaposition of DNA from the large internal repeat, *Bam*HI-W, to sequence in *Bam*HI H. The 'join' was identified by determining the primary structure of *PstI-PvuII* and *PstI-KpnI* restriction fragments, as illustrated (Figure 4). The total sequence was derived from data in Figure 2, as well as from the sequence of *Bam*HI W described previously (Jones and Griffin, 1983). The potential promoter signals (from *Bam*HI W) and the DS_L sequence are indicated (hatching), and the fusion point of *Bam*HI-W) with H noted (arrow). The deletion found in the Daudi EBV genome removes a number of open reading frames, all the *NotI* repeat sequences and part of the region (DS_L) that has homology with sequence (DS_R) in *Eco*RI-C. The resulting fusion of *Bam*HI W and H fragments does not create a new open reading frame. Further details are discussed in Figure 5.

viral protein should be encoded within a repetitive sequence, particularly one with such a high GC-content (nearly 85%) as the 125-bp *Not*I repeat (Jones and Griffin, 1983; Jeang and Hayward, 1983). Moreover, the number of nucleotides in the repeat also ensures that all triplets within the repeat are translated and re-iterated in the protein. However, it is interesting to note the recent report of a protein (*Plasmodium knowlesi* circumsporocoite) derived from DNA that has 12

copies of a 36-bp repeat unit which not only is translated but accounts for at least a third of the amino acid sequence of the total protein (Godson *et al.*, 1983).

This discussion has revolved around the potential coding sequences in an EBV transforming genome such as that found in the B95-8 strain of virus, and the consequences that may result from the deletion found in Daudi cells. Some of the properties of the latter might also, however, be a conse-



Fig. 4. (A) The EcoRI and BamHI restriction maps of linear virion B95-8 EBV DNA. Data for the EcoRI map are derived from Given and Kieff (1978) and those for BamHI from Skare and Strominger (1980). Daudi ccc DNA has an insertion of sequence in EcoRI-C similar to that observed with Raji EBV DNA (Raab-Traub et al., 1980; Rymo et al., 1981) and several viral strains derived from mononucleosis material (Fischer et al., 1981), as well as from recently established Burkitt's lymphoma lines (A Harris et al., in preparation). (B) An expansion of the BamHI W, Y, and H fragments that illustrates some of the restriction enzyme sites and some salient features in B95-8 DNA. Data from BamHI-W are from Jones and Griffin (1983), those from BamHI-Y are derived from our unpublished data and those of B.G.Barrell and colleagues (personal communication), and those from BamHI-H from Figure 2. The scale in kbp is indicated, as are the large internal repeats (LIR), the NotI repeats, and the leftward duplicated sequence (DS₁). The region shown as deleted in Daudi EBV DNA is taken from data in Figure 3 and that in the P3HR-1 strain from the published sequence of Jeang and Hayward (1983). Symbols used are: B = BamHI, P = PstI, PII = PvuII, H = HindIII, X = XhoI, K = KpnI. Potential transcriptional control sequences: \Box (promoter, left to right), and \dashv (right to left) and \downarrow AATAAA (left to right) and t (right to left). (C) A further expansion (note scale) of the portions of BamHI-W and -H extant in Daudi DNA, showing some of the features prominent in **B** that are lost in the deletion and others that are retained. Particularly notable is the relocation of the potential promoter (\rightarrow) from the large internal repeat (BamHI-W). The arrows (at bottom) indicate the area whose DNA sequence was determined in order to localise the precise sites of the deletion. Abbreviations are as in B.

quence of possible structural alterations. Jeang and Hayward (1983) have drawn attention to hairpin-like structures present in B95-8 DNA in the region under consideration (which would be absent in EBV DNA from Daudi cells) and Henry *et al.* (1983) have mapped a putative origin of replication (*ars* sequence) onto this general region. Further, an interesting structural feature may be created in Daudi EBV DNA by the juxtaposition of two otherwise non-contiguous regions of the genome. That is, a 46-bp long AC-rich site from one of the large internal repeats (*Bam*HI-W) (CCCAACACTCCACC-ACACCCAGGCACACACTACACACCCCG, nucleotides 1435 – 1480, Figure 3), with the potential for forming Z-DNA, becomes transposed from its normal pos-

itions in the genome to a site adjacent to the part of the EBV DS_L that remains in Daudi cells.

The discovery of a large (7.4-kb) deletion in the *Bam*HI-Y and -H fragments (see Figures 1 and 4B) of EBV episomal DNA in Daudi lymphoma cells clearly shows that the EVB deletion observed in P3HR-1 cells is not unique, as previously thought (see Introduction). Moreover, it exists in cells directly established from a tumour. The same deletion has been observed in the linear DNA of virions produced by Daudi cells (L.Karran and B.E.Griffin, unpublished results). It follows from studies on P3HR-1 that Daudi virions might be expected to be non-transforming. Preliminary experiments with EB virions from Daudi cells also indicate this to be the case (D.Crawford and B.E.Griffin, unpublished data), but these negative results require further confirmation.

Materials and methods

Cells

The EBV-transformed Daudi lymphoid cells, obtained from Professor G.Klein, Karolinska Institute, Stockholm, were the same as those studied by Rymo *et al.* (1979). They were propagated as suspension cultures in RPMI 1640 medium supplemented with 10% foetal calf serum, 100 U of penicillin and 100 μ g of streptomycin/ml. Cells were subdivided to 2 x 10⁵ cells/ml twice a week. B95-8 cells were the same as those used previously (Griffin *et al.*, 1981). Recombinant clones of EBV restriction enzyme fragments were taken from our EBV DNA library (Arrand *et al.*, 1981) and propagated as described.

Isolation and analysis of DNA

B95-8 virion DNA from EBV particles was isolated essentially as described by Adams (1975) from the spent medium of B95-8 cells that had been treated with 12-O-tetradecanoyl-phorbol-13-acetate (TPA), according to the procedure of zur Hausen *et al.* (1979). ccc DNA was prepared from Daudi cells as described by Griffin *et al.* (1981) except that for larger quantities of cells (~10¹⁰), the chromosomal DNA, suspended in 1 litre of buffer (pH 12.4), was fragmented by a brief (5–10 s) mix in a Waring blender. This considerably reduced the work-up time with no marked deleterious effect on the yield of ccc DNA. For further purification prior to cloning, Daudi ccc DNA was separated from mitochondrial DNA on a neutral sucrose gradient in the presence of ethidium bromide. Analysis of the fastest migrating species by gel electrophoresis subsequent to cleavage with *Bam*HI showed that it consisted mainly of EBV DNA.

For comparison of genomes, linear EBV DNA isolated from B95-8 virions and ccc DNA isolated from Daudi cells were cleaved with a 5-fold excess of the *Bam*HI restriction endonuclease (Boehringer Mannheim Ltd) and digested according to the manufacturer's directions. Fragments were separated by electrophoresis on 0.8% agarose gels in the presence of ethidium bromide, as previously described (Griffin *et al.*, 1981). For hybridisation studies, viral DNA in gels was denatured, transferred to nitrocellulose, and hybridised against suitable ³²P-labelled probes, according to the procedure of Southern (1975).

Molecular cloning and sequencing of DNA from Daudi cells

Episomal EBV DNA from Daudi cells was digested to completion with the restriction enzyme *Bam*HI, and the resulting fragments cloned into a *Bam*HI-cleaved, and phosphatase-treated, pAT153 vector, essentially as described by Arrand *et al.* (1981). The library of Daudi EBV clones obtained (~300) was probed with [³²P]nick-translated B95-8 *Bam*HI fragment H. Six positive clones were obtained. Four clones contained an insert of ~3500 bp, which was further shown to contain sequences from EBV *Bam*HI fragment W (the large internal repeat unit). One of these (designated 10G) was extensively mapped with a variety of restriction enzyme suing standard procedures. The remaining two clones were shown by restriction enzyme analysis to contain a larger insert that corresponded in size to a subfragment of *Eco*RI C; hybridisation against the analogous region of Raji DNA showed these to contain DS_R sequences.

The primary structure of the B95-8 *Bam*HI fragment H (clone number B7.12F) was determined by the random sonication, M13 cloning (Messing and Vieira, 1982) and dideoxy sequencing procedure (Sanger *et al.*, 1980), as pre-



Fig. 5. (A) Some of the potential coding areas in B95-8 BamHI fragment H, reading from left to right $(5' \rightarrow 3')$ on the conventional BamHI restriction enzyme map (Skare and Strominger, 1980), as derived from the DNA sequence (Figure 2). The corresponding sequence from Daudi BamHI-WH (from Figure 3) is shown for comparison in B. Some relevant transcriptional and structural features are noted. The number of base pairs (bp), taken from the sequences are indicated; reading frame 1 begins with the first nucleotide (G) of the BamHI recognition (GGATCC) signal, frame 2 with the second, etc. Vertical bars = termination codons; black horizontal lines = open reading frames; ● = 'CAT' or 'TATA' boxes; ■ = AATAAA; ▼ = limits of NotI Salient features found within BamHI-H are: (i) The long open reading frame present at the beginning of the sequence in frame 2 of B95-8 DNA could code for 372 amino acids (42 kd) and terminate at nucleotide 1118 (Figure 2) with TAA. This polypeptide has a high (~25%) proline content. The initiation codon for this potential reading frame must originate elsewhere ('to the left') on the genome. A polyadenylation site (AATAAA) lies at nucleotide 1156. (ii) The entire region covered by the Not 125-bp repeat contains no termination codons. The region adjacent to the repeats in all three reading frames also has coding capacity. (iii) The long open reading frame found at the end of the sequence in frame 3 (nucleotide 5529) has as its first codon an AUG initiation signal and within BarnHI-H could code for 161 amino acids. The putative promoter signals that precede this (TCAAT at nucleotide 5369 and GATA at nucleotide 5403) are not generally recognised as strong promoters. This potential coding sequence continues into BamHI-F. (B) The possible coding areas found in the Daudi BamHI-WH fusion fragment (in the left to right direction as above). The symbols used are the same as described in A. The sequence to the left of the vertical dashed line originates from BamHI-W. It is noteworthy that the only long open reading frame retained from BamHI-H is that described by (iii) above. The deletion removes ~ 190 bp from the DS₁.



Fig. 6. Comparison of the potential coding areas for Daudi BamHI-WH (A) and B95-8 BamHI fragment H (B), as derived from the DNA sequence (bottom to top, second strand, Figure 2), reading from right to left $(5' \rightarrow 3')$ on the conventional BamHI restriction enzyme map (Figure 4A). As in Figure 5, for the BamHI-H sequence frame 1 consists of amino acids that are initiated off the codon beginning with the first nucleotide (G) of the BamHI recognition sequence, etc. The symbols used are given in the legend to Figure 5. The only obvious, long open reading frames in B are those that surround the NotI repeats. These are removed by the deletion that creates the Daudi BamHI-WH fragment.

viously used for DNA sequence analysis of *Bam*HI-W (Jones and Griffin, 1983). Dideoxy sequencing was carried out using both $[\alpha^{-32}P]dATP$ and $[\alpha^{-33}S]$ thio-dATP labelling and 6% urea-polyacrylamide gels or 6% gradient polyacrylamide gels (Biggin *et al.*, 1983). To obtain the complete sequence, data were compiled from two libraries made from sonically fragmented DNA,

and an average of 250-300 bases were read from individual gels. Over 180 clones were analysed to give the sequence of 39 000 bases, from which the total *Bam*HI-H primary structures (Figure 2) was deduced. Sequence was derived from both strands for 96.6% of the fragment, and more than one clone was analysed for deducing 97.8% of the sequence.

Specific fragments from Daudi clone 10G were separated by agarose gel electrophoresis, isolated by electro-elution, and subcloned into suitable M13mp vectors. The *PstI-PvuII* and *PstI-KpnI* restriction fragments that spanned the *Bam*HI-W and -H junctions were sequenced, as above.

Acknowledgements

We thank Dr. B.G.Barrell for unpublished information on DNA sequences in Raji cells and Dr. T.Lindahl for much helpful discussion and critical reading of the manuscript. We are grateful to M.Hollis for help in preparing the manuscript.

References

- Adams, A. (1975) in Ablashi, D.V., Aaslestad, H.G. and de Thé, G. (eds.), Epstein-Barr Virus, Production, Concentration and Purification, International Agency for Research on Cancer, Lyon, pp. 129-146.
- Adams, A. (1979) in Epstein, M.A. and Achong, B.G. (eds.), *Epstein-Barr Virus*, Springer-Verlag, Berlin, pp. 155-183.
- Adams, A., Strander, H. and Cantell, K. (1975) J. Gen. Virol., 28, 207-217.
- Arrand, J. K., Rymo, L., Walsh, J.E., Björck, E., Lindahl, T. and Griffin, B.E. (1981) Nucleic Acids Res., 9, 2999-3014.
- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA, 80, 3963-3965.
- Bornkamm, G.W., Delius, H., Zimber, U., Hudewentz, J. and Epstein, M.A. (1980) J. Virol., 35, 603-618.
- Bornkamm, G.W., Hudewentz, J., Freese, U.K. and Zimber, U. (1982) J. Virol., 43, 952-968.
- Cheung, A. and Kieff, E. (1982) J. Virol., 44, 286-294.
- Erickson, J., Finan, J., Nowell, P.C. and Croce, C. (1982) Proc. Natl. Acad. Sci. USA, 79, 5611-5615.
- Fischer, D.K., Miller, G., Gradoville, L., Heston, L., Weststrate, W.M., Wright, J., Brandsma, J. and Summers, W.C. (1981) Cell, 24, 543-553.
- Fresen, K.-O., Merkt, B., Bornkamm, G.W. and zur Hausen, H. (1977) Int. J. Cancer, 19, 317-323.
- Gerber, P., Ablashi, D., Magrath, I., Armstrong, G., Anderson, P. and Trach, L., (1982) J. Natl. Cancer Inst., 69, 585-590.
- Gewert, D.R., Shah, S. and Clemens, M.J. (1981) Eur. J. Biochem., 116, 487-492.
- Given, D. and Kieff, E. (1978) J. Virol., 28, 524-542.
- Glaser, R., Boyd, A., Stoerker, J. and Holliday, J. (1983) Virology, 129, 188-198.
- Godson,G.N., Ellis,J., Svec,P., Schlesinger,D.H. and Nussenzweig,V. (1983) Nature, 305, 29-33.
- Griffin, B.E., Björck, E., Bjursell, G. and Lindahl, T. (1981) J. Virol., 40, 11-19.
- Hayward, S.D., Lazarowitz, S.G. and Hayward, G.S. (1982) J. Virol., 43, 201-212.
- Heller, M., Dambaugh, T. and Kieff, E. (1981) J. Virol., 38, 632-648.
- Henry, B.E., Raab-Traub, N.K. and Pagano, J.S. (1983) Proc. Natl. Acad. Sci. USA, 80, 1096-1100.
- Heston, L., Rabson, M., Brown, N. and Miller, G. (1982) Nature, 295, 160-163.
- Hinuma, Y., Konn, M., Yamaguchi, J., Wudarski, J., Bladeske, J.R. and Grace, J.T. (1967) J. Virol., 1, 1045-1051.
- Hudewentz, J., Delius, H., Freese, U.K., Zimber, U. and Bornkamm, E.W. (1982) *EMBO J.*, 1, 21-26.
- Hummel, M. and Kieff, E. (1982) Proc. Natl. Acad. Sci. USA, 79, 5698-5702.
- Jeang, K.-T. and Hayward, S.D. (1983) J. Virol., 48, 135-148.
- Jones, M.D. and Griffin, B.E. (1983) Nucleic Acids Res., 11, 3919-3937.
- King, W., Dambaugh, T., Heller, M., Dowling, J. and Kieff, E. (1982) J. Virol., 43, 979-986.
- Klein, E., Klein, G., Nadkarni, J.S., Nadkarni, J.J., Wigzell, H. and Clifford, P. (1968) Cancer Res., 28, 1300-1310.
- Kozak, M. (1981) Nucleic Acids Res., 9, 5233-5252.
- Lenoir, G.M., Preud'homme, J.L., Bernheim, A. and Berger, R. (1982) Nature, 298, 474-476.
- Luka, J., Kallin, B. and Klein, G. (1979) Virology, 94, 228-231.
- Manolov, G. and Manolova, Y. (1972) Nature, 237, 33-34.
- Messing, J. and Vieira, J. (1982) Gene, 19, 269-276.
- Miller, G., Robinson, J., Heston, L. and Lipman, M. (1974) Proc. Natl. Acad. Sci. USA, 71, 4006-4010.
- Mount,S.M. (1982) Nucleic Acids Res., 10, 459-472.
- Nilsson, K., Evrin, P.-E., Berggard, L. and Pontén, J. (1973) Nature New Biol., 244, 44-45.
- Pritchett, R., Pedersen, M. and Kieff, E. (1976) Virol., 74, 227-231.
- Raab-Traub, N., Dambaugh, T. and Kieff, E. (1980) Cell, 22, 257-267.
- Rabson, M., Gradoville, L., Heston, L. and Miller, G. (1982) J. Virol., 44, 834-844.

- Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol., 113, 237-251.
- Rymo, L. and Forsblom, S. (1978) Nucleic Acids. Res., 5, 1387-1402.
- Rymo, L., Lindahl, T. and Adams, A. (1979) Proc. Natl. Acad. Sci. USA, 76, 2794-2798.
- Rymo, L., Lindahl, T., Povey, S. and Klein, G. (1981) Virology, 115, 115-124.
 Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980)
 J. Mol. Biol., 143, 161-178.
- Silverman, R.H., Balkwill, F.R., Trowsdale, J. and Kerr, I. (1982) Eur. J. Biochem., 126, 333-341.
- Skare, J. and Strominger, J.L. (1980) Proc. Natl. Acad. Sci. USA, 77, 3860-3864.
- Southern, E.M. (1975) J. Mol. Biol., 98, 503-517.
- Stoerker, J., Holliday, J.E. and Glaser, R. (1983) Virology, 129, 199-206.
- Tovey, M.G., Dron, M. and Gresser, I. (1982) J. Gen. Virol., 60, 31-38.
- Twigg, A.J. and Sherratt, D. (1980) Nature, 283, 216-218.
- Van Santen, V., Cheung, A. and Kieff, E. (1981) Proc. Natl. Acad. Sci. USA, 78, 1930-1934.
- Weigel, R. and Miller, G. (1983) J. Virol., 125, 287-298.
- Zech, L., Haglund, U., Nilsson, K. and Klein, G. (1976) Int. J. Cancer, 17, 47-56.
- zur Hausen, H., Bornkamm, G.W., Schmidt, R. and Hecker, E. (1979) Proc. Natl. Acad. Sci. USA, 76, 782-785.

Received on 27 December 1983