DNA Sequence Relationship of the Baboon Endogenous Virus Genome to the Genomes of Other Type C and Type D Retroviruses

MAURICE COHEN,* NANCY RICE, ROBERT STEPHENS, AND CATHERINE O'CONNELL

Biological Carcinogenesis Program, Frederick Cancer Research Facility, Frederick, Maryland 21701

Received 17 August 1981/Accepted 9 November 1981

Baboon endogenous virus (BaEV) is a type C retrovirus present in multiple proviral copies in the DNA of baboons. Although interspecies antigenic determinants present on reverse transcriptase and *gag* proteins are shared among all mammalian type C viruses, no nucleic acid homology between BaEV and other type C viruses (except RD-114) has been found in conventional liquid hybridization experiments. In this study, we used restriction fragments of cloned BaEV DNA immobilized on nitrocellulose to test for relatedness with [³²P]cDNA's of various type C and type D viruses. We detected the following distant relationships previously found only through immunological and protein sequencing techniques: (i) eight type C viral cDNA's (the endogenous virus of rhesus monkeys, feline leukemia virus, simian sarcoma virus, gibbon ape leukemia virus, Rauscher murine leukemia virus, BALB-2, NZB, and RD-114) and two type D viral cDNA's (Mason-Pfizer monkey virus and squirrel monkey retrovirus) were able to hybridize with cloned BaEV DNA; (ii) the eight type C probes hybridized to restriction fragments spanning most of the BaEV genome, but only RD-114 hybridized to fragments within the 1.9 kilobases at the 3' end of the genome; (iii) the two type D probes hybridized primarily to fragments within the 1.9 kilobases at the 3' terminus and weakly or not at all elsewhere; and (iv) [³²P]cDNA's of several other oncornaviruses (mouse mammary tumor virus, equine infectious anemia virus, bovine leukemia virus, and reticuloendotheliosis virus) exhibited no homology with BaEV DNA. DNA sequence analysis has allowed us to orient the BaEV restriction map with the genetic map at both ends of the genome. Homologies between retroviral cDNA's and BaEV clone restriction fragments could thus be related to specific BaEV genes. Whereas type C cDNA's hybridized to fragments from *gag*, *pol*, and the *pol-env* junction, squirrel monkey retrovirus cDNA hybridized only to a fragment coding for the p15E portion of *env*. Mason-Pfizer monkey virus cDNA also hybridized within the p15E region, but exhibited homology to the 3' half of gp70 as well. These results are discussed relative to previously reported antigenic relatedness of retroviral proteins. The data suggest that BaEV represents an important link in oncornavirus evolution.

Baboon endogenous virus (BaEV) is a type C retrovirus present in multiple proviral copies in the DNA of baboons (10). With one exception (RD-114), no homology between BaEV and any of the other known mammalian type C or type D retroviruses has been found by liquid hybridization experiments under conditions of moderate stringency (12, 33, 37, 42, 43). In contrast, recent DNA sequence analyses of BaEV and AKR murine leukemia virus (AKR-MuLV) recombinant DNA clones revealed that their *gag* p30 genes contain a sequence of at least 35 identical nucleotides (14a; R. Stephens, R. Mural, and M. Cohen, unpublished data). Also, immunological and amino acid sequence analyses of retrovirus proteins have shown that: (i) all

mammalian type C retroviruses share interspecies antigenic determinants for reverse transcriptase (25); (ii) BaEV shares interspecies *gag* protein determinants with all other mammalian type C retroviruses (4, 5, 20); (iii) the amino acid sequences of BaEV and Rauscher MuLV (R-MuLV) p30 proteins are significantly related (16, 29; I. C. Henderson et al., unpublished data); and (iv) BaEV shares envelope gp70 antigenic determinants with the gp70's of two type D retroviruses, Mason-Pfizer monkey virus (MPMV) and squirrel monkey retrovirus (SMRV) (19, 35).

We initially characterized the BaEV genome by restriction analysis of BaEV unintegrated linear DNA (14) and the related proviruses of

baboon cells and tissues (13). The characterization was extended to other BaEV isolates by analysis of unintegrated BaEV DNAs (6) and through cloning of these DNAs into λ vectors (27). We recently reported the isolation of infectious and noninfectious BaEV proviral recombinant DNA clones derived from a baboon producer cell strain (14a). In this report we have reexamined the question of nucleic acid homology between BaEV and other mammalian type C and type D retroviruses. Through the use of BaEV proviral clone DNA in Southern blotting-hybridization experiments, we have developed an approach which has enabled us to detect distant relationships previously detected only by immunological and protein sequencing techniques.

MATERIALS AND METHODS

Viral probes. The following retroviruses were provided by the Viral Resources Laboratory, Frederick Cancer Research Center: BaEV (M7) from A204 cells; simian sarcoma virus/simian sarcoma-associated virus [SSV-1 (SSAV)] from NC-37 cells; gibbon ape leukemia virus (GaLV-H) from 6g-1 cells; SMRV from D-17 cells; MPMV from CMMT cells; RD-114 from RD cells; BALB-2 from Cf2Th cells; NZB from Cf2Th cells; endogenous virus of rhesus monkeys (MMC-1) from Cf2Th cells; R-MuLV from JLS-V9 cells; and feline leukemia virus (FeLV) (A, B, C) from RD cells (7). The procedure for isolation of 70S viral RNA was described previously (14). Randomly primed [32 P]cDNA probes were synthesized from the 70S viral RNAs as reported earlier (14, 38). Under these conditions, cDNA's prepared with BaEV or reticuloendotheliosis virus (REV-A) 70S RNA were judged to be representative probes since they hybridized to DNA fragments throughout their respective genomes. Similarly, RD-114 cDNA detected the same BaEV genomic fragments that were detected with a BaEV probe. All other retrovirus cDNA's were prepared identically and, therefore, were judged also to be representative.

Genomic BaEV clones. Genomic BaEV recombinant DNA clones were isolated as described elsewhere (14a). Two full-length clones (14A2 and 7B2) were used in these experiments. Although DNA of clone 14A2 but not of clone 7B2 is infectious in transfection assays (14a), the two proviruses are colinear and contain no gross DNA heterogeneities with respect to each other.

Restriction enzyme digestion and blot hybridization. BaEV clone DNAs were cleaved by standard single and double restriction enzyme digestions, electrophoresed, and blotted as previously described (14, 34). From the known BaEV restriction map (13, 14), the completeness of digestion was monitored simply by the appearance of the expected restriction fragments and the absence of incompletely digested ones. Unless otherwise stated, 0.1 μ g of BaEV clone 7B2 DNA was loaded in each lane of the agarose gel. Blotted filters were wetted in $3\times$ SSC (SSC: 0.15 M NaCl plus 0.015 M sodium citrate) and annealed overnight at 67°C in a solution containing approximately 10^7 cpm of the respective viral cDNA per ml and $5\times$ SSC, $10\times$ Denhardt's solution ($1\times$ Denhardt's: 0.02% polyvinyl-

pyrrolidone, 0.02% Ficoll, 0.02% Pentax bovine serum albumin), 0.1% sodium dodecyl sulfate, 20 μ g of polyriboadenylic acid per ml, and 20 μ g of denatured coal fish DNA per ml. After hybridization, filters were individually washed at 67°C in $3\times$ SSC containing $10\times$ Denhardt's solution, 0.1% sodium dodecyl sulfate, and 0.1% NaPPi. The final wash was at 67°C in $1\times$ SSC containing 0.1% NaPPi. Hybridization of the various cDNA probes was accompanied by a control hybridization with BaEV cDNA probe to confirm digestion patterns. Autoradiography of washed blots was done at -70°C with or without intensifying screens (Du Pont Cronex Lightning Plus) by use of Kodak XR film.

Stringency and quantitation. Since washing filters after hybridization in buffer of progressively lower ionic strength might result in denaturation of very unstable hybrids or loss of immobilized DNA from the filter, or both, we tested the effect of ionic strength in the posthybridization treatment. It was found that filters bearing BaEV [32 P]cDNA hybridized to BaEV DNA gave about the same band intensities when washed at 67°C in $4\times$, $2\times$, or $1\times$ SSC, that the intensities were slightly lower after washing in $0.4\times$ SSC, and that they were significantly lower after washing in $0.1\times$ SSC. Repeating the experiment with a heterologous [32 P]cDNA, that of SSV, gave essentially the same result. The important observation was that all bands detectable after washing in $4\times$ SSC were also apparent, albeit fainter, after washing in $0.1\times$ SSC. Thus, our standard washing condition of $1\times$ SSC at 67°C did not result in loss of otherwise detectable hybrids.

Band intensity after hybridization of BaEV DNA with any of the heterologous cDNA's was very weak compared with the intensity generated by hybridization with BaEV cDNA. In an attempt to quantitate this difference, we progressively reduced the amount of BaEV DNA per gel lane, looking for the amount which, after hybridization with BaEV cDNA, would give bands of similar intensity to those generated by 0.1 μ g of BaEV DNA hybridized with SSV cDNA. That amount turned out to be 0.001 μ g of BaEV DNA. Thus, by this rather imprecise measure there appears to be a difference factor of approximately 100 in the extents of hybridization in the heterologous and homologous cases; failure to detect relatedness of these viruses to BaEV DNA by conventional liquid hybridization is therefore understandable.

RESULTS

In the experiments described below, DNA from a BaEV recombinant clone that was isolated from a library of baboon cell DNA fragments (14a) was utilized to test for homologies with other type C and type D retrovirus [32 P]cDNA's. BaEV clone DNA was cleaved by one or two restriction endonucleases, electrophoresed in agarose gels, and transferred to nitrocellulose filters where the fragments were assayed for homology with the various labeled viral cDNA's by hybridization.

Our results, summarized in Fig. 1, indicate that: (i) eight type C and two type D viral cDNA's hybridized to a BaEV clone DNA even

FIG. 1. Summary of results. *Top*: Genetic map of BaEV genomic clone DNAs. Map coordinates are given in parentheses. Alignment of the genetic and restriction maps was obtained by DNA sequence analysis of a *gag* gene fragment (14a) and of a p15E-containing M13 subclone (Stephens et al., in preparation). *Middle*: Pertinent restriction enzyme cleavage sites in clones 14A2 and 7B2; the *Bam*HI site at map position 1.4 is present in clone 14A2 but not in 7B2. Each clone contains an approximately 0.55-kb long terminal repeat at each end (cross-hatching), and each is attached to baboon DNA flanking sequences (see Fig. 2). Right end of the proviral genome is at map position 8.8. Italicized letters A through H represent regions of the genome between restriction cleavage sites. *Bottom*: Summary of restriction fragments detected by retroviral $[^{32}\text{P}]\text{cDNA}$'s. Symbols: -----, weak hybridization; ———, intermediate hybridization; , very strong hybridization.

though (with the exception of RD-114) relatedness to BaEV was not detected by conventional liquid hybridization; (ii) the eight type C probes hybridized primarily to restriction fragments spanning all except the 1.9 kilobases (kb) at the 3' end of the genome but in differing patterns and intensities of annealing; (iii) the two type D probes hybridized primarily to restriction fragments spanning the 1.9 kb at the 3' end of the genome with little or no hybridization elsewhere in the genome; (iv) although readily detectable, hybridization of the heterologous cDNA's was very weak when compared with hybridization with a BaEV cDNA probe; and (v) $[^{32}\text{P}]\text{cDNA}$'s of several other oncornaviruses exhibited no homology with BaEV DNA. Each of these points will be considered in some detail below. To facilitate interpretation of the data, we present in Fig. 2 a map of the BaEV DNA that shows relevant restriction sites and lengths of fragments generated in single and double enzyme digests. The genome has been divided into regions A through H (Fig. 1 and 2).

Type C viruses. As shown in Fig. 3, MMC-1

and FeLV cDNA probes gave very similar patterns of hybridization to BaEV DNA. The two internal *Xho*I fragments and the two internal *Bam*HI fragments were detected (see Fig. 2), and in each case stronger hybridization occurred to the fragment derived from the 3' half of the genome. After digestion with *Bam*HI plus *Xho*I, these cDNA's hybridized to the 3.75-kb (genome regions B plus C) and 1.4-kb (E) fragments with stronger hybridization to the latter, but no hybridization was detected to the 0.9-kb (A) fragment located at the 5' end of the genome or to the 1.8-kb (H) fragment located at the 3' end of the genome. Digestion with *Bam*HI plus *Eco*RI confirmed and extended these results, showing that hybridization was detectable only to the 2.65-kb (B) and 1.85-kb (D plus E) fragments and not to the internal 1.1-kb (C) fragment or to the 3.5-kb (F through H) fragment which includes the 3' end of the genome. In several repetitions of this experiment, hybridization was always stronger to the 1.85-kb (D plus E) fragment than to the 2.65-kb (B) fragment. Neither increasing the exposure time nor using a 10-fold

FIG. 2. BaEV clone restriction cleavage sites. Restriction enzyme cleavage of BaEV recombinant clone DNAs in single and double enzyme digestions released genomic fragments of the given lengths. Map shows flanking *EcoRI* sites for clone 7B2: approximately 13 kb at the 5' end; 1.6 kb at the 3' end. Wavy lines represent baboon DNA flanking sequences. Other information same as in legend to Fig. 1.

higher mass of DNA per lane resulted in the appearance of any additional hybridizing bands. Thus, homology was detected over regions *B*, *D*, and *E*, with strongest hybridization to *E*.

A broader spectrum of homology was apparent when BaEV DNA was annealed with [³²P]cDNA's of SSV, GaLV, and R-MuLV (Fig. 3), as well as with cDNA of the xenotropic virus of New Zealand black mice (data not shown). In addition to detecting both internal *Bam*HI fragments and both internal *Xho*I fragments, these probes also detected the 1.1-kb *Bam*HI-*Eco*RI (*C*) fragment. The 3.5-kb *Bam*HI-*Eco*RI (*F* through *H*) fragment which spans the 3' end of the genome was detected with all three probes but with a relatively weak signal. cDNA of BALB-2, a murine xenotropic virus (data not shown), hybridized to the same restriction fragments within regions *B* through *E* but, like MMC-1 and FeLV, revealed strongest homology with the 1.4-kb *Bam*HI-*Xho*I fragment (region *E*).

Many studies have revealed that the endogenous cat virus, RD-114, is the type C retrovirus that is most closely related to BaEV. The relatedness of their genomes is sufficiently high to be detectable by conventional liquid hybridization,

where about 20% homology has been observed (9, 28). As shown in Fig. 4, this homology is distributed over the entire BaEV genome since all restriction fragments detectable with BaEV [³²P]cDNA were also observed after hybridization with an RD-114 probe. The intensity of annealed bands varied in proportion to the fragment size except that the 1.9-kb fragment at the 3' end of the genome hybridized weakly (see the *Bam*HI plus *Eco*RI 3.5-kb band, regions *F* through *H*; the *Bam*HI plus *Xho*I 1.7-kb band, *F* plus *G*; and the 1.8-kb *Xho*I plus *Eco*RI band, *H*). This region of the genome, which contains the coding sequence for a portion of the *env* gene (see Fig. 1), appears quite specific to BaEV by comparison with the other type C viruses. In contrast, the central parts of the genome represented by the 2.65-kb *Bam*HI-*Eco*RI fragment (*B*) and the 1.85-kb *Bam*HI fragment (*D* plus *E*) are the regions of the BaEV genome that are apparently related to all of the type C viruses shown in Fig. 1.

Several retrovirus cDNA's, including those of mouse mammary tumor virus (courtesy of L. Arthur, Frederick Cancer Research Center), bovine leukemia virus, and equine infectious anemia virus, did not hybridize to BaEV DNA. In

FIG. 3. Relatedness of mammalian type C probes to BaEV clone DNA. BaEV recombinant clone 7B2 was cleaved with various restriction enzymes in single and double enzyme digests and electrophoresed in 1.0% agarose gel as described in Materials and Methods. The DNA was denatured in situ, transferred to nitrocellulose sheets, and hybridized with [³²P]cDNA of the indicated retrovirus RNA. Additional large restriction fragments (as in the *Xho*I digest) are the result of partial enzymatic digestion.

FIG. 4. Homology between RD-114 and BaEV. Details of digestion, electrophoresis, and blotting same as in legend to Fig. 3. Hybridization was with BaEV or RD-114 probe. Although not observable in the figure, hybridization of RD-114 cDNA to the 0.45-kb *Bam*HI plus *Xho*I fragment could be seen in the original film.

addition, cDNA of reticuloendotheliosis virus, REV-A (the replication competent helper component of the transforming virus, REV-T), failed to hybridize to BaEV DNA even when annealed in 1.5 M Na⁺ (7.5× SSC) at 60°C for 18 h. This result is noteworthy because REV-A cDNA is able to hybridize with cloned DNA of the endogenous colobus type C virus, CPC-1 (29a). Since the macaque type C virus (MMC-1) and CPC-1 are quite closely related (33) and since MMC-1 cDNA hybridizes to BaEV (Fig. 3), hybridization of REV-A cDNA to BaEV might have been predicted. No homology was detected, however, in several repetitions of the experiment.

Type D viruses. The cDNAs of MPMV, a horizontally transmitted type D retrovirus of the rhesus monkey, and of SMRV, an endogenous type D retrovirus of the New World squirrel monkey, were also able to hybridize to BaEV

DNA, but they did so in a pattern unlike that observed for any of the type C viral cDNA's. These probes detected homologies with BaEV DNA mainly near the extreme 3' end of the genome, the region which hybridized poorly with all of the type C viral cDNA's. As shown in Fig. 5A, there was detectable but comparatively weak hybridization of MPMV [³²P]cDNA to the internal *Bam*HI fragments (B through E) and to the 5.1-kb *Xho*I fragment (A through D), but quite strong hybridization with the 3.1-kb *Xho*I fragment (E through G). After digestion of BaEV DNA with *Bam*HI plus *Eco*RI, the MPMV probe weakly detected the 2.65-kb (B), 1.1-kb (C), and 1.85-kb (D plus E) fragments, but strongly detected the 3.5-kb fragment which spans the 3' end of the genome (F through H). The SMRV probe also detected the 3.1-kb (E through G) and 3.5-kb (F through H) 3' end fragments in *Xho*I digests and *Bam*HI-*Eco*RI digests, respectively, but did not detect other genomic fragments. These results were supported and extended by the experiment shown in Fig. 5B. Both MPMV and SMRV cDNA probes revealed strong homology with the *Xho*I-*Eco*RI 3.1-kb (E through G) and the *Bam*HI-*Xho*I 1.7-kb (F plus G) fragments. As before, other genomic fragments hybridized weakly with MPMV cDNA and not at all with SMRV cDNA. Since our stock of MPMV was not cloned, we cannot rule out the possibility that it was contaminated with minor amounts of a type C virus of rhesus monkey, the original source of MPMV.

The region of strongest homology between BaEV DNA and the two type D retroviruses was further localized by hybridization after cleavage with *Xho*I plus *Bcl*I. As shown in Fig. 6, *Bcl*I cleaved the 3.1-kb *Xho*I fragment into 2.2-kb and 0.9-kb pieces, the latter encompassing all but the last 0.2 kb at the 3' end of the genome. (Digestion was incomplete, presumably because *Bcl*I is sensitive to methylation, and this site appears to be partially methylated during growth in *Escherichia coli* K-12.)

SMRV cDNA hybridized to the 0.9-kb fragment (G) but not to the 2.2-kb fragment (E plus F) in *Xho*I-*Bcl*I double digests (Fig. 6). Thus, the intensely hybridizing bands seen with SMRV [³²P]cDNA in Fig. 5 are derived entirely from this 0.9-kb region (G) which immediately precedes the BaEV 3' terminus. cDNA of MPMV, the other type D virus tested, gave a broader pattern of hybridization, showing homology to both the 2.2-kb (E plus F) and the 0.9-kb (G) *Xho*I-*Bcl*I fragments (Fig. 6). As was shown in the *Bam*HI-*Xho*I digest in Fig. 5, MPMV hybridized strongly only to the 1.7-kb fragment (F plus G). To map the homology of MPMV to BaEV in this region in more detail, we isolated the 3.1-kb *Xho*I fragment (E through G) from an agarose gel

FIG. 5. Relatedness of type D retrovirus probes to BaEV clone DNA. Details of digestion, electrophoresis, and blotting same as in legend to Fig. 3. Hybridization was with MPMV or SMRV probe. As in Fig. 3, extra bands, especially in *XhoI* digests are the result of partial enzymatic cleavages. In the *XhoI* plus *EcoRI* digest probed with MPMV, the three bands just above the 3.1-kb fragment are of lengths 3.6 kb (fragment A plus B), 4.7 kb (fragments C through G, the result of incomplete *XhoI* digestion), and 5.1 kb (fragments A through D, the result of incomplete *EcoRI* digestion). The weakly hybridizing band in the *XhoI* plus *EcoRI* digest probed with SMRV is the 4.7-kb partial *XhoI* product.

after electrophoresis. When the fragment was cut with *Bam*HI plus *Bcl*II, MPMV [32 P]cDNA was found to hybridize to both the 0.8-kb *Bam*HI-*Bcl*II fragment (F) and the 0.9-kb *Bcl*II-*Xho*I fragment (G) (data not shown).

The *Bcl*II site at map position 7.7 defines the approximate 3' limit of type C probe hybridization. The 2.2-kb (E plus F) and 0.9-kb (G) *Bcl*II-*Xho*I fragments were detected after hybridization with either BaEV cDNA or RD-114 cDNA (data not shown). In striking contrast, the other type C retrovirus probes hybridized only to the 2.2-kb (E plus F) fragment and not at all to the 0.9-kb (G) fragment. This is shown for FeLV, R-MuLV, and GaLV in Fig. 6; an identical pattern of hybridization was observed with SSV, MMC-1, and NZB cDNA probes as well (data not shown). These data confirm that the primary nucleic acid homology between BaEV and the various type C viruses extends to the right of the *Xho*I site at map position 5.5, but that it does not extend beyond the *Bcl*II site at position 7.7. The weak hybridization in Fig. 3 of R-MuLV, GaLV, and SSV cDNA's to the *Bam*HI-*Eco*RI 3.5-kb (F through H) fragment indicates the low, although

detectable, degree of relatedness of these genomes to BaEV in region F. Thus, the primary type C retrovirus homologies to BaEV are to the left of map position 6.9.

DISCUSSION

In this report we describe the detection of nucleic acid relatedness between the genomes of BaEV and those of other mammalian retroviruses and map the regions of homology in the BaEV genome. Our success in detecting such homologies was partially due to utilization of cloned BaEV DNA fragments (14a) for hybridization with 32 P-labeled viral cDNA probes in Southern blotting experiments. Although hybridization to BaEV restriction fragments with heterologous cDNA's was weak in comparison with that observed with BaEV probes, the results were quite reproducible. cDNA probes of a particular retrovirus made at different times and annealed in different experiments repeatedly detected some, but not other, BaEV restriction fragments, usually with the same relative intensities. These observations were not due to anomalies in the blotting of DNA, since parallel

FIG. 6. Homology of type C and type D retrovirus probes to the *env* region of a BaEV clone DNA.

digests annealed with BaEV probes revealed fragments hybridizing in proportion to their lengths.

The qualitative data are summarized in Fig. 1. All eight type C viruses contain nucleic acid homologies with BaEV. RD-114 is related to BaEV throughout its genome, but is less related near the 3' terminus. R-MuLV, GaLV, NZB, and SSV revealed homologies in restriction fragments located throughout the central parts of the genome but not at either terminus. BALB-2, FeLV, and MMC-1, as with the Rauscher-gibbon ape group, hybridized only in the central portion of the genome, but unlike that group revealed their strongest hybridization to the *Bam*HI-*Xho*I fragment *E*. Further, MMC-1 and FeLV genomes showed no relatedness to BaEV in the *Eco*RI-*Bam*HI fragment *C*. Except for RD-114, the type C retroviruses revealed little (R-MuLV, NZB, GaLV, SSV) or no (FeLV, MMC-1, BALB-2) detectable homology with BaEV to the right of region *E* (Fig. 1). In striking contrast to the type C viruses, the two type D retroviruses showed a converse pattern of hybridization: with MPMV the homology was primarily to the right of region *E*, and with SMRV the homology was to the right of region *F*. Finally, several retroviruses were found to be unrelated to BaEV. These are reticuloendotheliosis virus, a type C virus of fowl, mouse mammary tumor virus, a type B virus, and the unclassified bovine leukemia virus and equine infectious anemia virus.

We have oriented the BaEV restriction map with the genetic map at both ends of the genome through DNA sequence analysis (14a; Stephens et al., in preparation). The DNA sequence of an M13 subclone of a BaEV clone DNA fragment revealed the subclone to contain a 228-base pair insert beginning with the *Bcl*II site at map position 7.7. This fragment was found to contain the coding sequence for the amino terminus of the BaEV p15E envelope protein and to begin 10 nucleotides upstream in the linker peptide region between the gp70 and p15E (Fig. 1; T. D. Copeland et al., in preparation). Thus, the *Bcl*II site between region *F* and *G* can be located at the gp70/p15E border. We have previously shown that the BaEV *Bam*HI site at map position 1.4 is the probable coding sequence for *gag* p30 amino acids 32 to 34, which places the 5' end of the *gag* gene at approximately 0.63 in the physical map (14a).

Thus, homologies between the type C and type D retrovirus genomes and that of BaEV

Details of digestion, electrophoresis, and blotting same as in legend to Fig. 3. Hybridization was with indicated probe.

could be assigned to homologies within specific BaEV genes (Fig. 1). RD-114 revealed relatedness with BaEV in restriction fragments within the *gag*, *pol*, and *gp70* portion of the *env* gene, but less (although detectable) relatedness in fragments spanning the 3' end of the *gp70* and *p15E* envelope genes. Although the RD-114 and BaEV genomes were reported to be only approximately 20% homologous by nucleic acid hybridization (9, 28), they are known to share antigenic determinants for *gag* proteins (32) and reverse transcriptase (25); further, the BaEV reverse transcriptase was partially inhibited by antisera prepared against RD-114 polymerase (31). Their envelope proteins are also closely related since the viruses show reciprocal interference and some cross neutralization in tissue culture (21, 22). Electron microscope heteroduplex analysis between BaEV RNA and short RD-114 cDNA transcripts revealed detectable homologies in the *gag* and *pol* gene regions (24). Further, the strong stop cDNA sequences of the two viruses were found to be related, having 72% of their nucleotides corresponding (26).

We have also detected nucleic acid homologies between BaEV and seven other type C retroviruses in both the *gag* and *pol* genes and in fragment *E* (see Fig. 1), which spans the *pol* gene-*env* gene junction. Three retroviruses, MMC-1, FeLV, and BALB-2, revealed their strongest homology to this BaEV fragment. It is noteworthy that the two xenotropic mouse viruses, BALB-2 and NZB, differ in the strength of their hybridization to fragment *E*, a difference which may correlate with the extensive restriction site differences noted between their genomes (S. K. Chattopadhyay et al., *Virology*, in press). Previous immunological studies showed that the BaEV structural proteins (products of the *p12*, *p10*, and *p30 gag* genes) share interspecies antigenic determinants with the analogous proteins from all other mammalian type C viruses tested (4, 5, 20). But whereas *gp70* antigenic determinants are shared among SSV, FeLV, GaLV, and murine type C viruses, SSV shares no detectable *gp70* determinants with BaEV or the closely related RD-114 (23). Thus, the above-noted homologies of the several type C viral probes in region *E* are due either to close homology within the 3' part of the BaEV *pol* gene or to previously undetected relatedness with the BaEV *env* gene. The very weak homologies of R-MuLV, NZB, GaLV, and SSV detected near the 3' end of the BaEV *gp70 env* gene (region *F*) have no correlates in immunological studies.

Our results show that sequence relatedness between BaEV and the two type D retroviruses, MPMV and SMRV, is present primarily in the *env* gene. Again, the results are both supported

by and in contrast to those of previous immunological studies. Whereas we detected homology between BaEV and MPMV primarily in the envelope region (in the *p15E* and in the 3' 40% of the *gp70*), weak but reproducible hybridization was observed in restriction fragments throughout the *gag* and *pol* genes as well (Fig. 5 and 6). However, neither the *gag* proteins (1, 15, 35, 44) nor the reverse transcriptase (25, 30) of MPMV share antigenic determinants with the respective type C viral proteins, including those of BaEV.

With SMRV, on the other hand, we detected homology with BaEV only in the *p15E* portion of the *env* gene. No hybridization was observed in the *gp70* coding region even though the *gp70* of SMRV is antigenically related to that of BaEV (19). Thus, an even closer immunological cross-reactivity should exist between the BaEV and SMRV *p15E* proteins than between their *gp70*'s. Consistent with the above finding, the major structural proteins of SMRV were reported to be unrelated to those of mammalian type C viruses (1, 15, 17); however, a recent report showed some cross-reactivity between the major structural protein of SMRV and the murine type C virus, R-MuLV (3).

It is clear from the above results that the use of recombinant clone DNA for detecting cross-species relatedness in Southern experiments is an approach that is both highly sensitive and revealing. The use of a clone DNA was of paramount importance since it enabled us to immobilize purified and highly concentrated BaEV DNA fragments on the filters where they were available for hybridization. The detection of homologies in such experiments, but not in conventional liquid hybridization studies, is probably based on two major factors. First, the background to signal ratio on filters after hybridization is typically very low. In contrast, the background in liquid hybridization experiments is usually several percent of the total radioactivity, regardless of whether duplex formation is assayed by S1 nuclease resistance or binding to hydroxylapatite. Therefore, hybrid formation by a few percent of the probe, such as we estimate is occurring here (see Materials and Methods), is difficult to detect reliably in such experiments. Second, for a hybrid region to be detected in liquid hybridization studies, it must be long enough and sufficiently well matched to escape S1 digestion or to bind to hydroxylapatite. Southern blotting experiments impose no such constraints. Indeed, the difficulty in thermally denaturing filter-bound hybrids suggests that the nitrocellulose may actually somewhat stabilize a hybridized molecule. As in hydroxylapatite experiments, a long unhybridized tail would serve to amplify the apparent extent of hybridization.

We know neither the length nor the degree of

pairing in the heterologous duplexes we have detected. Although some could be both long (>100 base pairs) and well matched, it is unlikely that they all are, for only about four such regions would be sufficient to render homologies detectable by liquid hybridization. We do know of one example of a short region of very close homology. We recently reported (14a) that the BaEV and an AKR-MuLV recombinant DNA clone contain a DNA sequence of at least 35 identical nucleotides within their p30 regions. Finally, we may also be detecting hybrids with significant degrees of mismatching, and these could, in principle, be quite long. In fact, very low-stringency hydroxylapatite experiments revealed limited annealing of BaEV probe to macaque and murine type C and langur type D viral RNAs (11).

Could the homologies that we observed be the result of nonspecific binding of probe to BaEV clone restriction fragments? Several lines argue against this possibility: (i) cDNA probes from four other retroviruses (mouse mammary tumor virus, REV-A, equine infectious anemia virus, and bovine leukemia virus) did not anneal to BaEV genome restriction fragments although two of the probes (mouse mammary tumor virus and REV-A) were positive in hybridization to their respective DNAs; (ii) each of the eight type C and two type D probes reproducibly hybridized to some but not to other BaEV fragments; and (iii) we know of one case where weak nucleic acid homology was not detected in our hybridization experiments. We sequenced a portion of the BaEV p15E envelope gene (Stephens, O'Connell, and Cohen, unpublished data; see Fig. 2) and compared it to the nucleotide sequence of the Moloney MuLV p15E gene (36) using the computer program ALIGN (18). In the 180 nucleotide long portion of the Moloney MuLV gene corresponding to the p15E amino acids 63 through 123, the program revealed an alignment score of 9.06 (the probability of this being a random score would be 4.5×10^{-20}). The percentage of nucleotide correspondences was 66%. Thus, a Moloney MuLV cDNA molecule spanning just this portion of the genome would contain a 34% mismatch relative to the BaEV DNA sequence. Indeed, as summarized in Fig. 1, cDNA's of three other murine type C retroviruses, R-MuLV, NZB, and BALB-2, showed no detectable homology to the BaEV p15E coding region. Since a probe having a 34% mismatch would not be expected to form a stable hybrid under our hybridization conditions, these results further rule out the possibility that very weak homologies or nonspecific interactions account for the hybridization data.

In general, our results support and extend the relationships observed between BaEV and other

mammalian type C and type D retroviruses from immunological studies. In addition, we have detected previously unsuspected relatedness between BaEV and several type C retroviruses within the gp70 portion of the *env* gene and between BaEV and MPMV within portions of the *gag* and *pol* genes. Another region of particularly strong sequence relatedness was detected between BaEV and the genomes of FeLV, MMC-1, and BALB-2 in a fragment (*E*; see Fig. 1) which spans the 3' end of the *pol* gene and the 5' half of the gp70. On the other hand, nucleic acid hybridization did not detect the reported antigenic relatedness between BaEV and SMRV in the gp70 portion of the *env* gene (19).

How might nucleic acid hybridization detect viral interrelatedness that was not observed by immunological techniques? First, analogous proteins of two different species may contain considerable amino acid sequence homology (homology which may be reflected in the DNAs as conserved nucleotide sequences), but if the conserved homologies are not antigenically accessible, the native proteins would not cross-react immunologically. Second, since antigenic reactive sites are only six to seven amino acid residues in length (2), other important and possibly conserved features may be present on the surface of a protein that do not involve antigenic determinants.

On the other hand, conservation of antigenic specificities between retroviruses need not correlate with conservation of nucleotide sequences encoding those amino acids. The genetic code is sufficiently versatile to allow extensive DNA divergence without altering the amino acid sequence. Also, as shown by Atassi and co-workers for myoglobin, two or more regions of a protein that are separated in linear sequence may nevertheless interact in three-dimensional structure to form an antigenic reactive site (2). This possibility further increases the likelihood that some antigenic relatedness is not reflected in the nucleic acid sequence.

It now appears that oncornaviruses of different species are evolutionarily related. They all contain common functional and genetic elements, including long terminal repeats and genes coding for *gag*, *pol*, and *env* proteins. Furthermore, retroviral genes are endogenous in the genomes of a diverse group of vertebrates, e.g., birds, mammals (including many primates), and reptiles (1; P. Andersen, personal communication). The possibility exists, therefore, that the original "protovirus" (39, 40) appeared early in the evolution of vertebrates. In addition, retroviruses have apparently also undergone horizontal transspecies spread several times (possibly many times) in evolution. For example, the SSV/GaLV retrovirus group is thought to have

originated from endogenous viruses of mice (8), and RD-114, an endogenous cat virus, appears to have originated from a progenitor of today's BaEV (9). Thus, any attempted clarification of evolutionary relationships of retroviruses cannot ignore the possibility of transspecies virus spread.

Although these relationships are complex, one aspect is apparent: BaEV (and RD-114) shares *gag* and *pol* gene homologies with the type C retroviruses and *env* gene homologies with the type D retroviruses tested. With few exceptions (3, 41), the type C and type D viruses are otherwise quite unrelated. Therefore, as previously discussed (19), BaEV may represent an important link in oncornavirus evolution. Precise nucleotide sequence comparisons between primate type C and type D envelope genes should reveal whether recombination has occurred in their ancestry.

ACKNOWLEDGMENTS

We are grateful to R. V. Gilden, A. Rein, and T. I. Bonner for helpful discussions and to C. V. Benton for providing virus. We thank JoAnne Beheler and Stephanie Simek for expert technical assistance, and Jeannie Clarke, Linda Fawley, and Margaret Fanning for help in manuscript preparation.

This research was sponsored by the National Cancer Institute under contract NO1-CO-75380 with Litton Bionetics.

LITERATURE CITED

- Andersen, P. R., M. Barbacid, S. R. Tronick, H. F. Clark, and S. A. Aaronson. 1979. Evolutionary relatedness of viper and primate endogenous retroviruses. *Science* 204:318-321.
- Atassi, M. Z. 1975. Antigenic structure of myoglobin: the complete immunochemical anatomy of a protein and conclusions relating to antigenic structures of proteins. *Immunochemistry* 12:423-438.
- Barbacid, M., L. K. Long, and S. A. Aaronson. 1980. Major structural proteins of type B, type C, and type D oncoviruses share interspecies antigenic determinants. *Proc. Natl. Acad. Sci. U.S.A.* 77:72-76.
- Barbacid, M., J. R. Stephenson, and S. A. Aaronson. 1976. *gag* gene of mammalian type C viruses. *Nature (London)* 262:554-559.
- Barbacid, M., J. R. Stephenson, and S. A. Aaronson. 1977. Evolutionary relationships between *gag* gene-coded proteins of murine and primate endogenous type C RNA viruses. *Cell* 10:641-648.
- Battula, N., and G. J. Todaro. 1980. Physical map of infectious baboon type C viral DNA and sites of integration in infected cells. *J. Virol.* 36:709-718.
- Benton, C. V., H. M. Hodge, and D. L. Fine. 1978. Comparative large-scale propagation of retroviruses from Old World (Mason-Pfizer monkey virus) and New World (squirrel monkey virus) primates. *In Vitro* 14:192-199.
- Benveniste, R. E., R. Callahan, C. J. Sherr, V. Chapman, and G. J. Todaro. 1977. Two distinct endogenous type C viruses isolated from the asian rodent *Mus cervicolor*: conservation of viro gene sequences in related rodent species. *J. Virol.* 21:849-862.
- Benveniste, R. E., and G. J. Todaro. 1974. Evolution of C-type viral genes: inheritance of exogenously acquired viral genes. *Nature (London)* 252:456-459.
- Benveniste, R. E., and G. J. Todaro. 1976. Evolution of type C viral genes: evidence for an asian origin of man. *Nature (London)* 261:101-108.
- Bonner, T. I., and G. J. Todaro. (1980). The evolution of baboon endogenous type C virus: related sequences in the DNA of distant species. *Virology* 103:217-227.
- Bryant, M. L., C. J. Sherr, A. Sen, and G. J. Todaro. 1978. Molecular diversity among five different endogenous primate retroviruses. *J. Virol.* 28:300-313.
- Cohen, M., N. Davidson, R. V. Gilden, R. M. McAllister, M. O. Nicolson, and R. M. Stephens. 1980. The baboon endogenous virus genome. II. Provirus sequence variations in baboon cell DNA. *Nucleic Acids Res.* 8:4423-4440.
- Cohen, M., M. O. Nicolson, R. M. McAllister, M. Shure, N. Davidson, N. Rice, and R. V. Gilden. 1980. Baboon endogenous virus genome. I. Restriction enzyme map of the unintegrated DNA genome of a primate retrovirus. *J. Virol.* 34:28-39.
- Cohen, M., A. Rein, R. M. Stephens, C. O'Connell, R. V. Gilden, M. Shure, M. O. Nicolson, R. M. McAllister, and N. Davidson. 1981. Baboon endogenous virus genome: molecular cloning and structural characterization of non-defective viral genomes from DNA of a baboon cell strain. *Proc. Natl. Acad. Sci. U.S.A.* 78:5207-5211.
- Colcher, D., Y. A. Teramoto, and J. Schlom. 1977. Interspecies radioimmunoassay for the major structural proteins of primate type-D retroviruses. *Proc. Natl. Acad. Sci. U.S.A.* 74:5739-5743.
- Copeland, T. D., L. E. Henderson, E. S. Vanlaningham-Miller, J. R. Stephenson, J. R., G. W. Smythers, and S. Oroszian. 1981. Amino- and carboxyl-terminal sequences of proteins coded by *gag* gene of endogenous baboon and cat type C viruses. *Virology* 109:13-24.
- Dahlberg, J. E., S. R. Tronick, and S. A. Aaronson. 1980. Immunological relationships of an endogenous guinea pig retrovirus with prototype mammalian type B and type D retroviruses. *J. Virol.* 33:522-530.
- Dayhoff, M. O. 1976. Atlas of protein sequence and structure, vol. 5, suppl. 2, p. 1. National Biomedical Research Foundation, Washington, D.C.
- Devare, S. G., R. E. Hanson, and J. R. Stephenson. 1978. Primate retroviruses: envelope glycoproteins of endogenous type C and type D viruses possess common interspecies antigenic determinants. *J. Virol.* 26:316-324.
- Gilden, R. V. 1977. Biology of RNA tumor viruses, p. 435-542. In D. Nayak (ed.), *Molecular biology of animal viruses*, vol. 1. Dekker, New York.
- Hellman, A., P. T. Peebles, J. E. Strickland, A. K. Fowler, S. S. Kalter, S. Oroszian, and R. V. Gilden. 1974. Baboon virus isolate M7 with properties similar to feline virus RD-114. *J. Virol.* 14:133-138.
- Henderson, I. C., M. M. Lieber, and C. J. Todaro. 1974. Mink cell line Mv1Lu (CCL64). *Virology* 60:282-287.
- Hino, S., J. R. Stephenson, and S. A. Aaronson. 1975. Antigenic determinants of the 70,000 molecular weight glycoprotein of Woolly monkey type C RNA virus. *J. Immunol.* 115:922-927.
- Hu, S., N. Davidson, M. O. Nicolson, and R. M. McAllister. 1977. Heteroduplex study of the sequence relations between RD-114 and baboon viral RNAs. *J. Virol.* 23:345-352.
- Krakower, J. M., and S. A. Aaronson. 1978. Radioimmunologic characterization of RD-114 reverse transcriptase: evolutionary relatedness of mammalian type C viral *pol* gene products. *Virology* 86:127-137.
- Lovinger, G. G., and G. Schochetman. 1980. 5' terminal nucleotide sequences of type C retroviruses: features common to terminal noncoding sequences of known eucaryotic messenger RNAs. *Cell* 20:441-449.
- Noda, M., M. Wagatsuma, T. Tamura, T. Takano, and K. Matsubara. 1981. Structure of the baboon endogenous virus genome: cloning of circular virus DNA in bacteriophage λ . *Nucleic Acids Res.* 9:2173-2186.
- Okabe, H., R. V. Gilden, M. Hatanaka, J. R. Stephenson, R. E. Gallagher, R. C. Gallo, S. R. Tronick, and S. A.

- Aaronson. 1976. Immunological and biochemical characterization of type C viruses isolated from cultured human AML cells. *Nature (London)* 260:264-266.
29. Oroszlan, S., L. E. Henderson, T. D. Copeland, A. M. Schultz, and E. M. Rabin. 1980. Processing and structure of murine leukemia virus *gag* and *env* gene encoded polyproteins, p. 219-232. In G. Koch and D. Richter (ed.), *Biosynthesis, modification, and processing of cellular and viral polyproteins*. Academic Press, Inc., New York.
 - 29a. Rice, N. R., T. I. Bonner, and R. V. Gilden. 1981. Nucleic acid homology between avian and mammalian type C viruses: relatedness of reticuloendotheliosis virus cDNA to cloned proviral DNA of the endogenous colobus virus CPC-1. *Virology* 114:286-290.
 30. Scolnick, E., W. Parks, and G. J. Todaro. 1972. Reverse transcriptases of primate viruses as immunological markers. *Science* 177:1119-1121.
 31. Sherr, C. J., M. M. Lieber, R. E. Benveniste, and G. J. Todaro. 1974. Endogenous baboon type C (M7); biochemical and immunologic characterization. *Virology* 58:492-503.
 32. Sherr, C. J., and G. J. Todaro. 1974. Radioimmunoassay of the major group specific protein of endogenous baboon type C viruses: relation to the RD-114/CCC group and detection of antigen in normal baboon tissues. *Virology* 61:168-181.
 33. Sherwin, S. A., and G. J. Todaro. 1979. A new endogenous primate type C virus isolated from the Old World monkey *Colobus polykomos*. *Proc. Natl. Acad. Sci. U.S.A.* 76:5041-5045.
 34. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.
 35. Stephenson, J. R., S. Hino, E. W. Garrett, and S. A. Aaronson. 1976. Immunological cross-reactivity of Mason-Pfizer monkey virus with type C RNA viruses endogenous to primates. *Nature (London)* 261:609-611.
 36. Sutcliffe, J. G., T. M. Shinnick, I. M., Verma, and R. A. Lerner. 1980. Nucleotide sequence of Moloney leukemia virus: 3' end reveals details of replication, analogy to bacterial transposons, and an unexpected gene. *Proc. Natl. Acad. Sci. U.S.A.* 77:3302-3306.
 37. Tainsky, M. A. 1981. Analysis of the virogenes related to the rhesus monkey endogenous type C retrovirus in monkeys and apes. *J. Virol.* 37:922-930.
 38. Taylor, J. M., R. Illmensee, and J. Summers. 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. *Biochim. Biophys. Acta* 442:324-330.
 39. Temin, H. M. 1970. Malignant transformation of cells by viruses. *Perspect. Biol. Med.* 14:11-26.
 40. Temin, H. M. 1980. Origin of retroviruses from cellular moveable genetic elements. *Cell* 21:599-600.
 41. Thiel, H. J., E. M. Broughton, T. J. Matthews, W. Schäfer, and D. P. Bolognesi. 1981. Interspecies reactivity of type C and D retrovirus p15E and p15C proteins. *Virology* 111:270-274.
 42. Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr. 1978. MAC-1, a new genetically transmitted type C virus of primates: "low frequency" activation from stump-tail monkey cell cultures. *Cell* 13:775-782.
 43. Todaro, G. J., R. Callahan, U. R. Rapp, and J. E. DeLarco. 1980. Genetic transmission of retroviral genes and cellular oncogenes. *Proc. R. Soc. London Ser. B* 210:367-385.
 44. Tronick, S. R., J. R. Stephenson, and S. A. Aaronson. 1974. Immunological properties of two polypeptides of Mason-Pfizer monkey virus. *J. Virol.* 14:125-132.