JOURNAL OF VIROLOGY, Jan. 1985, p. 32–39
0022-538X/85/010032-08$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 53, No. 1

# Molecular Basis of Host Range Variation in Avian Retroviruses

ANDREW J. DORNER, JONATHAN P. STOYE, AND JOHN M. COFFIN*

*Department of Molecular Biology and Microbiology and Cancer Research Center, Tufts University School of Medicine, Boston, Massachusetts 02111*

Received 30 July 1984/Accepted 10 September 1984

**Previous genetic analysis has localized the region of the Rous sarcoma virus (RSV) *env* gene responsible for host range specificity to that encoding the middle one-third of gp85. To better understand the host range determinants, the relevant regions of the genomes of infectious molecular clones of the transformation-defective Prague strain of RSV, subgroup B (Pr-RSV-B) and Rous-associated virus 0 (RAV-0) (subgroup E) were sequenced and compared with the sequence of Pr-RSV-C. This comparative analysis identified two variable regions of low amino acid sequence homology flanked by highly conserved amino acid sequences. The first variable region (hr1) begins at base 5654 in the Pr-RSV-C sequence and encodes 32 amino acids. The second variable region (hr2) begins at base 5846 and encodes 27 amino acids. To test the role of the variable regions in host range specificity, we determined the sequence of this region of the *env* gene of NTRE-4, a recombinant virus between Pr-RSV-B and RAV-0 which exhibits an extended host range. This analysis revealed that the recombinant subgroup-encoding region of NTRE-4 is composed of 200 bases of RAV-0 sequence, including hr2, flanked by sequences which are otherwise of Pr-RSV-B origin. This study indicates that hr1 and hr2 are the domains of gp85 responsible for host range determination in avian retroviruses.**

The retroviruses of chickens related to Rous sarcoma virus (RSV) can be divided into five host range subgroups, A through E, defined by the ability to infect genetically defined chicken cells, interference patterns, and neutralizing antibody cross-reaction (8, 11, 25, 28). Several more subgroups have been identified in viruses isolated from other species of galliform birds (27). The ability of a virus to infect a cell is dependent upon the presence on the cell surface of a specific receptor site for that viral subgroup. The virion envelope glycoproteins gp85 and gp37, the products of the *env* gene, are responsible for the subgroup-specific interaction with the host cell receptor (24). Resistance of cells to superinfection by viruses of the same subgroup specificity as defined by interference patterns is probably the result of blockage of the receptor by newly synthesized viral glycoprotein expressed at the cell surface (21).

In the chicken, three autosomal loci have been identified which control cell susceptibility to infection by viruses of subgroups A, B, and C. These loci are designated *tv-a*, *tv-b*, and *tv-c*, respectively (27). Susceptibility is always dominant over resistance, implying that they directly encode the receptors. The *tv-a* and *tv-c* loci are genetically linked, although they code for independent receptors (19). Whereas some genetic studies have indicated that the *tv-b* locus consists of multiple alleles encoding the receptors for both subgroup B and subgroup E viruses (5), other genetic evidence suggests that an independent locus, *tv-e*, encodes the subgroup E receptor (18). Resistance to subgroup B viruses is associated with resistance to subgroup E viruses, and preinfection of chicken cells with subgroup B or D virus prevents superinfection by subgroup E virus. However, this interference is nonreciprocal (27). Chicken cells resistant to subgroup B viruses also exclude subgroup D viruses almost completely and the two subgroups exhibit reciprocal interference, indicating that they probably recognize the same receptor encoded by the *tv-b* locus (8). Thus, subgroup B, D, and E viruses recognize closely related receptors. Analysis of the susceptibility of other avian species to avian leukosis virus has shown that other species do not exhibit patterns of resistance identical to those described for chicken cells. For example, turkey or quail embryo fibroblasts which are susceptible to infection by subgroup E virus are resistant to infection by subgroup B and D viruses (24). This is in contrast to the situation described for chicken cells in which susceptibility to subgroup E virus is always associated with susceptibility to subgroup B and D viruses. Thus, these receptors are not identical among galliform species.

T1 oligonucleotide fingerprint analysis has identified a region within the *env* gene which segregates strictly with the subgroup phenotype in crosses between subgroups (3, 12). By identifying the T1 oligonucleotides associated with this region of the *env* gene within the nucleotide sequence of the Prague strain of RSV, subgroup C (Pr-RSV-C), it was possible to position the sequences coding for subgroup specificity roughly between bases 5490 and 5950 (10). This places the host range determinants within the middle one-third of gp85 and points to the domain of gp85 which specifically interacts with the cell surface receptors.

Although the host range coding region usually segregates as an inviolate unit, some recombinant viruses have been isolated which exhibit a host range extended beyond that of the parental viruses (23). One such virus, NTRE-4, is a recombinant between transformation-defective (*td*) Pr-RSV-B and the endogenous Rous-associated virus 0 (RAV-0). Analysis of this virus indicated that it recognizes both the subgroup B receptor on chicken cells and the subgroup E receptor on turkey cells. This composite host range is associated with a unique recombination event within the subgroup-encoding region of the *env* gene. A single subgroup E-specific oligonucleotide is present within an *env* gene which is otherwise composed entirely of subgroup B-specific oligonucleotides. This recombination event is an exception to previous observations of the strong linkage of oligonucleotides within the subgroup-encoding region (22) and further associates these sequences with host range determination.

To identify the host range determinants on gp85, we have obtained molecular clones of NTRE-4 and *td*Pr-RSV-B DNA

* Corresponding author.

and conducted a comparative nucleotide sequence analysis of the region of the *env* gene between bases 5266 and 6060 of these viruses and of RAV-0. This region encodes amino acids 8 through 280 in the deduced amino acid sequence of the mature gp85. The gp85 of Pr-RSV-C is 341 amino acids in length (10). This sequence analysis encompassed all of the $T_1$ oligonucleotides which segregate with subgroup specificity.

Comparison of the deduced amino acid sequences of this region of gp85 revealed two domains of low amino acid homology between virus subgroups within a highly conserved amino acid sequence. We have termed these variable regions hr1 and hr2. Analysis of the sequence of NTRE-4 showed that the crossovers within the *env* gene generated a recombinant gp85 which carried hr1 of subgroup B and hr2 of subgroup E. Thus, our comparative sequence analysis has identified two small variable domains on gp85, both of which are involved in host range determination.

## MATERIALS AND METHODS

**Cells and viruses.** Chicken embryo fibroblasts (C/E; chf$^-$, gs$^-$, V$^-$) were prepared from fertilized eggs (SPAFAS, Norwich, Conn.). Turkey embryo fibroblasts (T/BD) were prepared from fertilized eggs (Orlopp Enterprises, Arosi, Calif.). QT-6 cells are a continuous line derived from a methylcholanthrene-induced fibrosarcoma of Japanese quail (17). Pr-RSV-B, its transformation-defective deletion mutant (*td*Pr-RSV-B), the endogenous RAV-0, and the recombinant virus NTRE-4 have been extensively characterized in this laboratory (3, 23). The molecular clone of RAV-0 (pRAV-0) was a gift of P. N. Tsichlis. The molecular clone of Pr-RSV-C (pATV-8) was a gift of R. Guntaka (13). Both of these clones yielded infectious virus indistinguishable from the parental type after introduction into appropriate avian host cells (P. Norton and L. Lieberman, personal communication).

**Molecular cloning of *td*Pr-RSV-B and NTRE-4.** Subconfluent monolayers of C/E cells were infected with Pr-RSV-B and the infection was monitored by focus formation. Subconfluent monolayers of QT-6 cells were infected with NTRE-4 and the infection was monitored by the presence of sedimentable reverse transcriptase activity. For large-scale purification of viral DNA, uninfected and virus-infected cells were mixed in a ratio of 10:1 and grown for 20 h in roller bottles. The unintegrated viral DNA from $2 \times 10^8$ to $3 \times 10^8$ cells was isolated by the method of Hirt (9), and the supernatant was treated with pronase and pancreatic RNase essentially as described previously (7). The purified supernatant was enriched for closed circular DNA by a single extraction with acidic phenol (pH 4.2) (30).

Because preliminary analysis of proviral DNA had indicated the presence of a unique *Sal*I recognition site within the *env* gene sequences encoding gp85, this site was utilized in the cloning procedure. A total of 10% of the purified Hirt supernatant DNA was digested with *Sal*I and ligated with 1 μg of *Sal*I-digested L47.1 λ DNA (14) overnight at 13°C under standard conditions (15). Ligated DNA was packaged in vitro into phage particles from a sonicated extract of BHB2690 (prehead donor) and a freeze-thaw extract of BHB2688 (packaging protein donor) (15). Phage was plated onto *Escherichia coli* 1160, a *recA*$^+$ P2 lysogen, at approximately $1.5 \times 10^4$ PFU per 10-cm plate. Recombinant phage were identified by plaque hybridization by the method of Benton and Davis (1), using DNA from plasmid pATV-8 labeled by nick translation as a probe. Nick translation was performed under conditions recommended by the radioisotope supplier (New England Nuclear). After several rounds of plaque purification on *E. coli* 1160, recombinant phage DNA was prepared from plate lysates (15) of infected *E. coli* LE392, analyzed by restriction enzyme digestion, and tested for infectivity as described below. Viral DNA inserts were subcloned into the *Sal*I site of pBR322 and tested for infectivity as described below. Infectious plasmid clones of *td*Pr-RSV-B (pPrB-1) and NTRE-4 (pNT-4) were used for subsequent DNA sequence analysis.

**Transfection of cloned viral DNA and $T_1$ oligonucleotide fingerprinting of rescued virus.** Viral inserts were excised from vector DNA by *Sal*I digestion and ligated with T4 DNA ligase to produce nonpermuted circular copies of the viral DNA. C/E or T/BD cells were transfected by either the calcium phosphate method (7) modified such that the final concentration of $Na_2HPO_4$ in the HEPES (*N*-2-hydroxyethylpiperazine-*N*'-2-ethanesulfonic acid)-buffered saline was 1 mM instead of 7 mM or by the DEAE-dextran procedure (6). Production of infectious virus was assayed by testing for sedimentable reverse transcriptase activity. $T_1$ oligonucleotide fingerprints of $^{32}P$-labeled RNA of virus rescued by transfection were prepared essentially as described before (2).

**Nucleotide sequence analysis.** DNA fragments were labeled at the 5' termini with [$\gamma$-$^{32}P$]ATP and T4 polynucleotide kinase after removal of 5'-phosphate groups with calf intestinal alkaline phosphatase according to standard procedures (15). DNA fragments were labeled at the 3' termini, using the Klenow fragment of DNA polymerase I (15). Radiolabeled DNA was usually subjected to secondary restriction enzyme digestion to produce fragments with a single radiolabeled terminus and were purified by polyacrylamide gel electrophoresis (16). Alternatively, the complementary strands were denatured and the strands were separated by polyacrylamide gel electrophoresis (15). DNA sequence analysis was performed by the chemical cleavage method of Maxam and Gilbert (16) with the following alterations: the second precipitation after the base modification reactions was performed by dissolving the pelleted DNA in 150 μl of 10 mM Tris-hydrochloride (pH 7.5) and adding 4 volumes of 95% ethanol containing 60 mM sodium acetate; in place of the series of lyophilizations after the piperidine cleavage reactions, the reactions were precipitated by the addition of 4 volumes of 95% ethanol containing 60 mM sodium acetate and the pelleted DNA was washed with 70% ethanol and lyophilized. Analysis of sequence data was performed by using a dot matrix program written by J.M.C. (unpublished data).

**Enzymes and radioisotopes.** Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories and used under conditions recommended by the supplier. T4 DNA ligase and T4 polynucleotide kinase were from New England Biolabs. Calf intestinal alkaline phosphatase was purchased from Boehringer Mannheim Biochemicals. The Klenow fragment of *E. coli* DNA polymerase I, DNA sequencing reagents, and all radioisotopes were from New England Nuclear Corp.

## RESULTS AND DISCUSSION

**Characterization of molecular clones of *td*Pr-RSV-B and NTRE-4.** Initial restriction mapping of recombinant phage revealed a range of sizes of viral inserts resulting from the cloning of Pr-RSV-B and NTRE-4 DNA. Although the original virus stock contained transformation-competent Pr-RSV-B, none of the cloned viral DNA analyzed presented arestriction pattern consistent with a full-length Pr-RSV-B genome. DNA from several of the largest phage clones of Pr-RSV-B was tested for infectivity by transfection of C/E

cells. Within 7 days after transfection, the production of infectious virus was detected by testing for sedimentable reverse transcriptase activity (data not shown). Morphological transformation of transfected cells was not observed, indicating that the molecular clones of Pr-RSV-B we obtained represented transformation-defective deletion mutants.

Genomic RNA of virus rescued by transfection was subjected to $T_1$ oligonucleotide fingerprint analysis. This analysis confirmed the identity of the virus and revealed the absence of *src*-specific oligonucleotides 6A and 5, which map approximately in the middle of the *src* gene, and the retention of *src*-specific oligonucleotide 6B (data not shown). This composition was consistent with the observation that virus stock obtained after a few passages contains a large proportion of virus which lacks the portion of the *src* gene encoded by oligonucleotides 6A and 5 (4). Restriction mapping of the infectious plasmid clone pPrB-1 DNA indicated that approximately 1 kilobase pair of 3′ *src* sequences had been deleted in this molecular clone of *td*Pr-RSV-B (Fig. 1A).

DNA from phage clones of NTRE-4 was introduced into T/BD cells by transfection, and rescued virus was examined by fingerprint analysis. The oligonucleotide fingerprint of the rescued virus showed the unique characteristic of NTRE-4: the presence of a single subgroup E-specific oligonucleotide (oligonucleotide 06) within the *env* gene otherwise containing only subgroup B-specific oligonucleotides (data not shown). The infectious plasmid clone pNT-4 exhibited a restriction map similar to but distinguishable from that of pPrB-1 (Fig. 1A). pNT-4 DNA lacked an *Xho*I site located 3′ of the *env* gene in pPrB-1 DNA presumably because these sequences are lost with the complete *src* gene deletion of NTRE-4. The position of an *Sst*I site in the *gag* gene of pNT-4 DNA which is not present in pPrB-1 corresponds to a site in pRAV-0 DNA. Its presence in pNT-4 is in agreement with the $T_1$ oligonucleotide fingerprint which showed that NTRE-4 was generated by several crossovers at the 5′ end of the genome within the *gag* and *pol* genes (23).

As a final check on the identity of the molecular clones, the host range of the rescued viruses was investigated. NTRE-4 exhibited an extended host range with the ability to infect both C/E cells and T/BD cells, whereas *td*Pr-RSV-B infected C/E cells and was unable to infect T/BD cells (data not shown). Thus, the molecular clones of NTRE-4 and *td*Pr-RSV-B had the same biological activity as the original virus stocks.

**Comparison of the deduced amino acid sequences of gp85.** Nucleotide sequence analysis was performed on the region of the *env* gene between an *Xho*I site at base 5258 and a unique *Sal*I site at base 6059 (nucleotide assignments according to the sequence of Pr-RSV-C [20]) by the strategy shown in Fig. 1B for *td*Pr-RSV-B, NTRE-4, and RAV-0. The complete sequences determined by this analysis are presented in Fig. 2 and 3. This region of the *env* gene contains the subgroup B-specific oligonucleotides 9, 2, 33, 19, 28, and 35 in Pr-RSV-B and the subgroup E-specific oligonucleotides 012, 07, 02, 04, 013, and 06 in RAV-0. Oligonucleotide 308 in Pr-RSV-C and RAV-0 is allelic to oligonucleotide 9 in Pr-RSV-B. Located near the 5′ end of these sequences is oligonucleotide 4 which is found in all strains of avian leukosis virus. Preliminary sequence analysis has located oligonucleotide 402 in RAV-0 approximately 20 bases after the *Sal*I site, which marks the 3′ terminus of the sequence analysis (data not shown). Oligonucleotide 402 is not subgroup specific. Thus, the nucleotide sequence presented here encompasses all of the oligonucleotides reported to segregate with the subgroup phenotype in crosses between subgroups (3, 12) and so must include all of the information relevant to host range determination. Initial studies involving the construction of recombinant virus by the exchange of restriction fragments have also shown that the host range determinants are located in sequences 5′ of the *Sal*I site (Coffin, unpublished data).

Presented in Fig. 4 is the complete amino acid sequence of the gp85 of Pr-RSV-C, beginning with the aspartic acid residue located at the amino terminus of the mature glycoprotein (10), compared with the deduced amino acid sequences of the portion of gp85 of Pr-RSV-B and of RAV-0 deter-

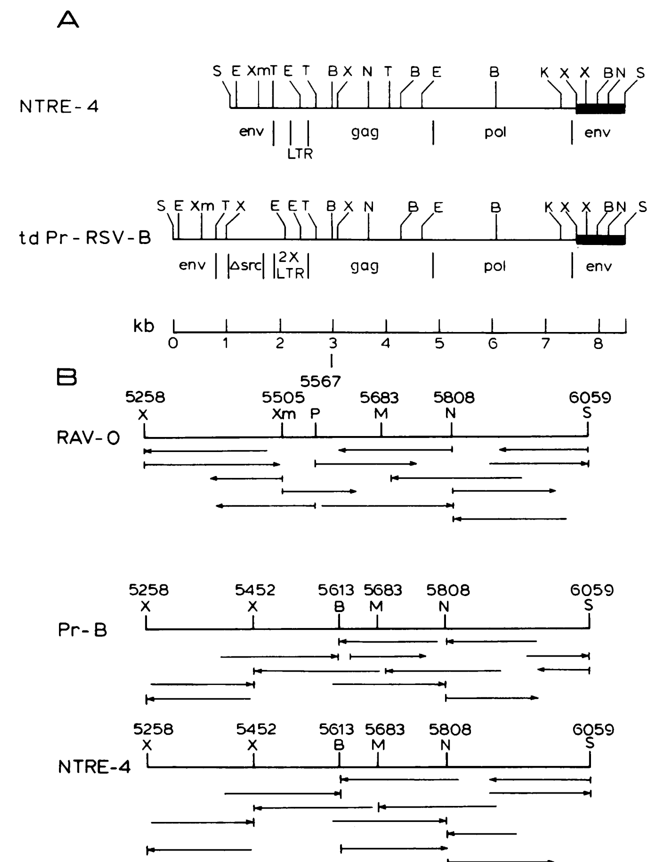


FIG. 1. Structure of pPrB-1 and pNT-4 DNAs. (A) Infectious clones of *td*Pr-RSV-B (pPrB-1) and NTRE-4 (pNT-4) DNA permuted at the unique *Sal*I site in pBR322 were analyzed by cleavage with the restriction enzymes shown. The genetic structure of the virus genome is indicated below the restriction map. Clone pNT-4 DNA contains a single long terminal repeat (LTR), whereas clone pPrB-1 DNA has two long terminal repeats in tandem as determined by the *Eco*RI digestion pattern. Note the deletion of approximately 1 kilobase from the 3′ end of the *src* gene in pPrB-1. The region of the genome subjected to sequence analysis is indicated by the thick line. (B) Sequencing strategy of the host range encoding region of the *env* gene. Indicated are the restriction sites utilized in the sequence analysis of the regions of RAV-0, *td*Pr-B (*td*Pr-RSV-B), and NTRE-4. Nucleotide assignments of restriction sites are numbered from the first base of the recognition sequence according to the sequence of Pr-RSV-C (20). Vertical lines indicate the labeled end of a fragment, and horizontal arrows indicate the extent and orientation of the sequencing run. Restriction enzyme designations: X, *Xho*I; Xm, *Xmn*I; T, *Sst*I; P, *Pvu*II, S, *Sal*I; N, *Nde*I; M, *Msp*I; K, *Kpn*I; E, *Eco*RI; B, *Bam*HI.

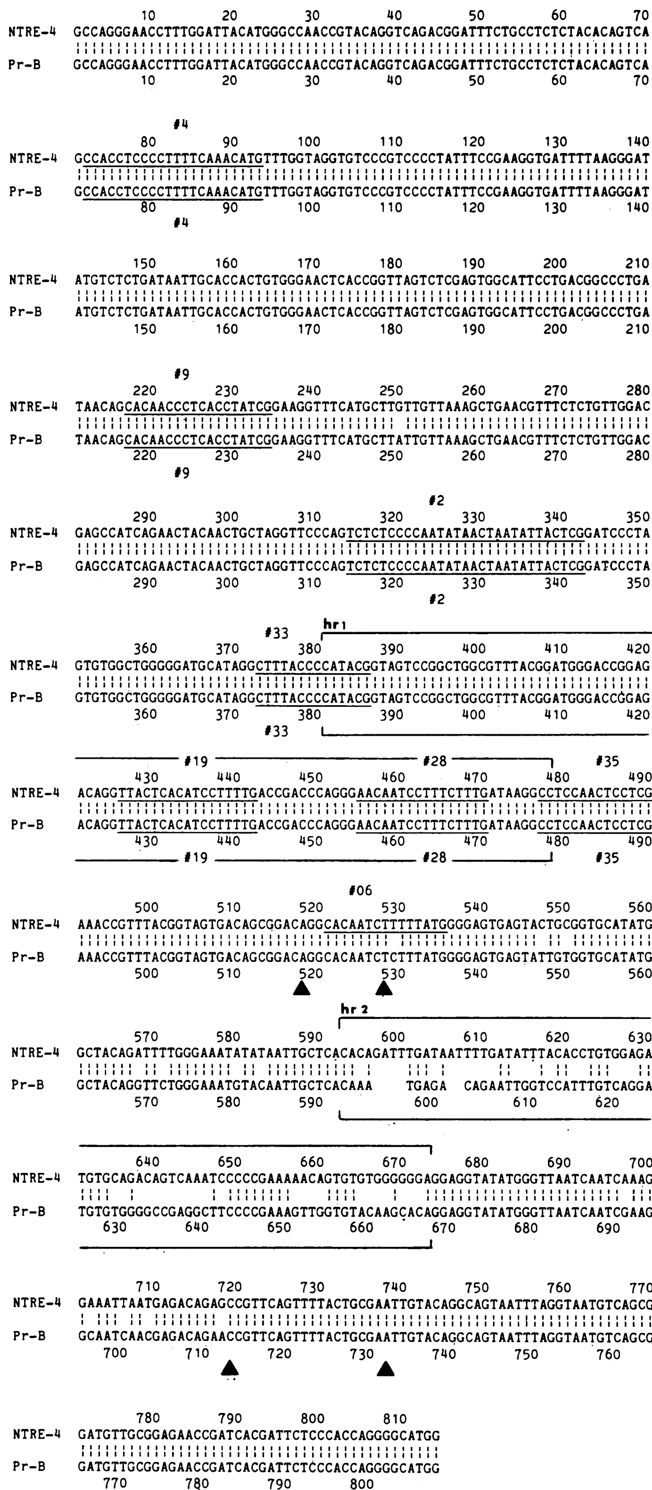

FIG. 2. Comparison of nucleotide sequence of NTRE-4 with that of Pr-RSV-B. The first base of these sequences corresponds to base 5266 in the Pr-RSV-C sequence and the sequence terminates at base 6061. This portion of the *env* gene encodes the 273-amino acid region of gp85 shown in Fig. 4. RNase $T_1$ oligonucleotides present within this part of the genome are underlined and numbered according to Coffin et al. (3). The sequences corresponding to the host range-specific regions hr1 and hr2 are boxed. Triangles delineate the regions within which the recombination events which generated NTRE-4 must have occurred.

mined by our sequence analysis. This comparative analysis encompassed amino acid positions 8 through 280 of the gp85 of the three different host range subgroups and included the middle one-third of the protein implicated in host range determination. A consensus amino acid sequence of gp85 can be constructed of amino acid residues shared among the different subgroup sequences. Such a sequence highlights regions of gp85 which exhibit subgroup-specific variation by the appearance of lacunae in the consensus sequence at positions where there is no common residue between at least two subgroups.

Examination of these sequences revealed large blocks of complete amino acid homology among all three subgroups. The amino-terminal region of gp85 displays a zone of highly conserved amino acid sequence extending from the proline residue at position 8 to the aspartic acid residue at position 57, in which the only two differences are the same conservative changes from isoleucine residues in RAV-0 and Pr-RSV-C to valine residues in Pr-RSV-B. Beginning at position 58 and continuing to position 76 is a region in which the amino acid sequence is no longer completely homologous among all three subgroups. Of 19 positions within this region, 10 exhibit an amino acid residue shared among all three subgroups. Six positions in the consensus sequence indicate homology only between subgroups C and E, including two positions marked by deletions in the subgroup B sequence. Beginning with the glycine residue at position 77 is a 60-amino acid stretch of 80% amino acid homology among all three subgroups in which the consensus sequence still remained intact.

The only amino acid sequence differences which showed subgroup-specific variation were found to form two small regions within the middle one-third of gp85 between amino acids 136 and 234. Following a conserved proline residue at position 136 and corresponding to bases 5654 to 5731 in the Pr-RSV-C sequence, the consensus sequence breaks down. At 12 of 32 positions there is no common residue between at least two subgroups. This region was termed hr1 (for host range region 1). Immediately preceding hr1 there lies a short region of amino acid differences in which there is agreement only between two of three subgroups. More than half of these differences are conservative amino acid changes, and because they did not display variation according to subgroup, this region was not included within hr1. After hr1 is a region of almost complete amino acid homology encompassing 38 residues. Beginning at position 207 and corresponding to bases 5846 to 5917, a second region of variable amino acid sequence exists, termed hr2. Within hr2 11 of 27 amino acid positions fail to exhibit homology among subgroups. hr2 ends at amino acid position 233 and is followed by a conserved amino acid sequence up to the end of our sequence analysis at amino acid position 280. The variability of hr1 and hr2 within the highly conserved sequence of gp85 implicates these two regions in the determination of host range. An ancillary role of other amino acid differences such as those following hr2 between residues 258 and 277 is also possible.

Within hr1 the amino acid sequences of the gp85 of subgroups B and E display more homology with each other than with that of subgroup C. Fourteen of 16 amino acid residues shared between any two subgroups are common to subgroups B and E. Five amino acids are homologous among all three subgroups, whereas 11 positions exhibit subgroup-specific variation. This suggests that the host range determinants of subgroup B and E viruses are more closely related to each other than to those of subgroup C

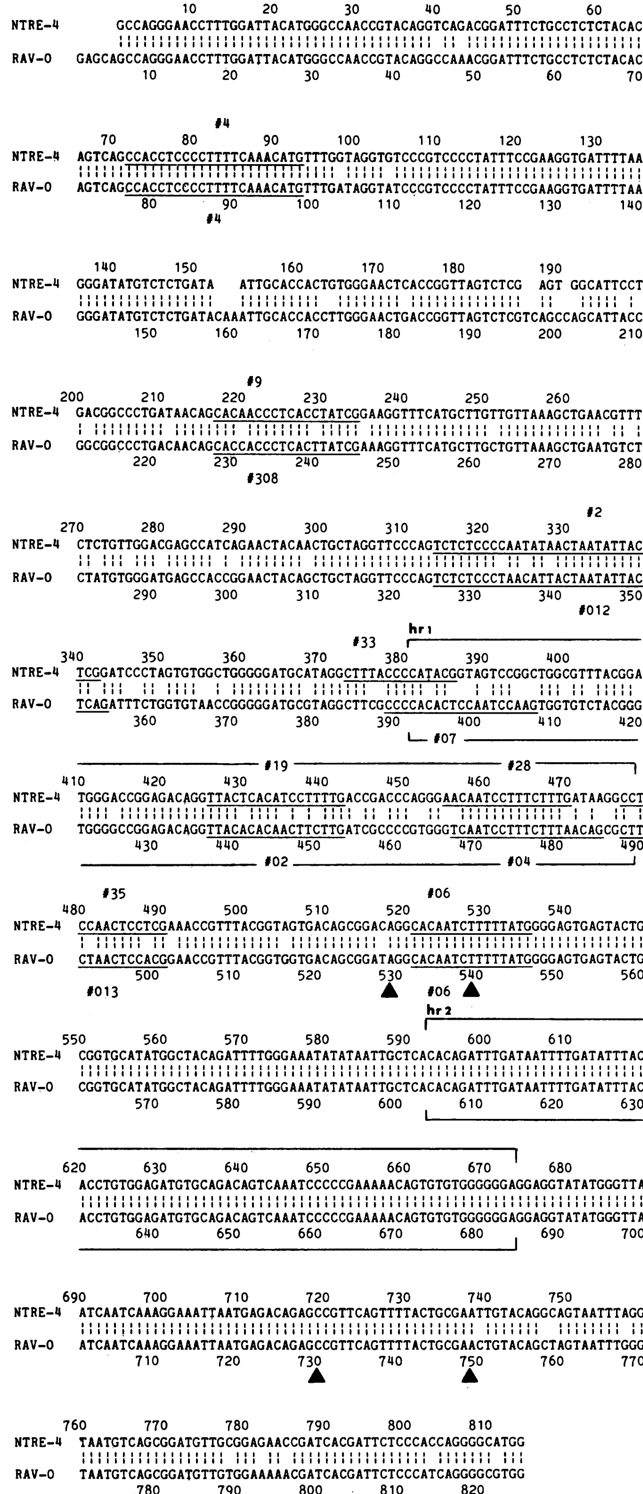

FIG. 3. Comparison of nucleotide sequence of NTRE-4 with that of RAV-0. The first base of the RAV-0 sequence corresponds to base 5261 of the Pr-RSV-C sequence and the sequence ends at base 6061. All conventions are as described in the legend to Fig. 2.

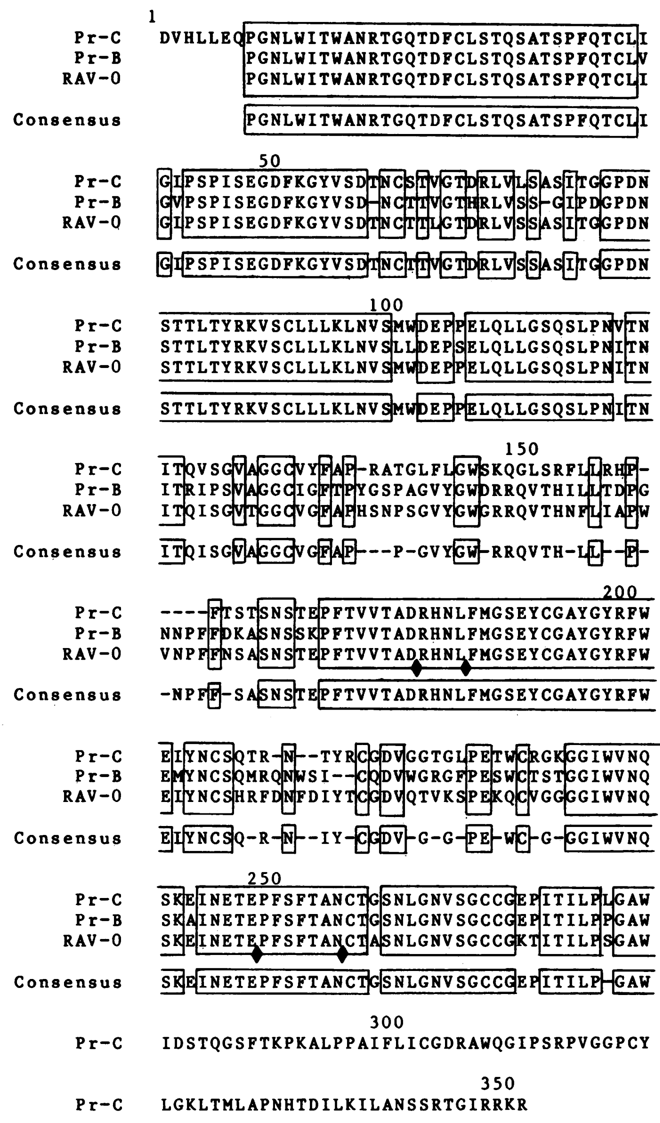

FIG. 4. Comparison of amino acid sequences of gp85 of three different subgroups. The amino acid sequence of gp85 of Pr-RSV-C is shown in its entirety (10). The deduced amino acid sequences of the portion of gp85 determined by this sequence analysis of Pr-RSV-B and RAV-0 are aligned within the Pr-RSV-C sequence. Amino acids are numbered from the amino-terminal aspartic acid of the mature gp85 (10). Boxed areas indicate positions in the gp85 sequence at which all three subgroups have a common amino acid. Hyphens indicate deletions in the aligned sequence. The consensus sequence was derived as follows: boxed regions indicate positions of amino acid homology among all three subgroups; unboxed amino acid residues are common between any two subgroups; hyphens mark positions in the consensus sequence where there is no amino acid homology. Hyphens in the consensus sequence thus identify regions of subgroup-specific divergent amino acid sequence among the virus subgroups. Diamonds indicate the regions of the crossovers which generated the recombinant *env* gene of NTRE-4.

viruses. The homology between hr1 of subgroup B and hr1 of subgroup E correlates with the results of genetic analysis which indicates that subgroup B and E viruses recognize closely related receptors (5) and supports the presumptive role of these variable regions in host range determination by interaction with the cellular receptors. hr2 displays more variation in amino acid sequence between subgroups than hr1, and an obvious relationship between hr2 of different subgroups cannot be discerned.

It is interesting that outside of the two variable regions the subgroup E amino acid sequence is more homologous with

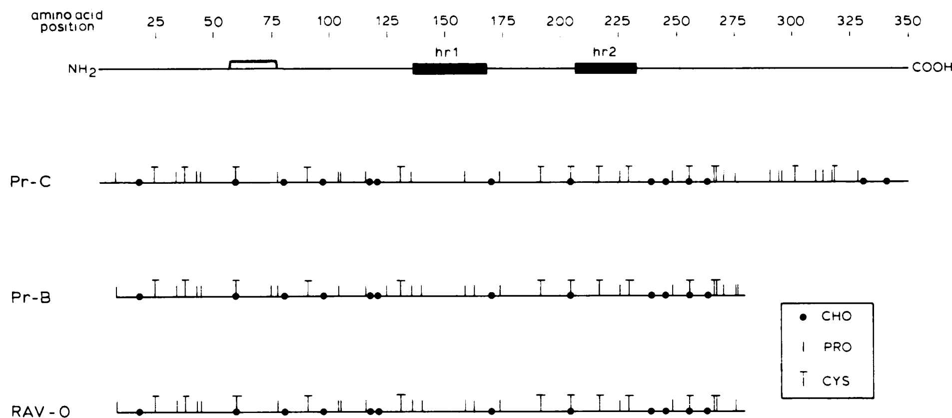


FIG. 5. Topography of the envelope glycoprotein gp85. The upper schematic representation of gp85 shows the positions of hrl and hr2 on the molecule. The open bar indicates the location of the small region of Pr-RSV-B divergent from that of RAV-0 and Pr-RSV-C. The amino acid positions are numbered as in Fig. 4. The representations of the gp85 of Pr-RSV-C, Pr-RSV-B, and RAV-0 below indicate the positions of cysteine and proline residues and of potential glycosylation sites on the protein. The subgroup C-specific gp85 is repesented in its entirety, whereas the gp85 molecules of subgroup B and E specificity are represented within the region determined by this sequence analysis. Symbols: ı, proline; T, cysteine; ●, glycosylation site of the type Asn-X-Thr/Ser.

subgroup C than with subgroup B. This is particularly striking in the region from amino acids 58 through 76. Outside of hrl and hr2 20 of 30 amino acids shared between any two subgroups represent homology between subgroups E and C. This may be the result of homogenization by frequent recombination of sequences not required for host range specificity. It cannot be ruled out, however, that these sequences are involved in host range determination, but this would imply that the interaction of subgroup B viruses with their cellular receptor is different in some way from that of subgroup E and C viruses.

Alternatively, the divergence of the subgroup B amino acid sequence may be related to the reported cytopathogenicity of avian leukosis viruses of subgroups B and D. A study of naturally cytopathic retroviruses of subgroup B indicated that the absence of superinfection resistance resulted in massive reinfection and cell death, whereas viruses of subgroups A, C, and E are not cytopathic (29). It has also been observed that repeated passage of Pr-RSV-B in cultures of chicken cells selects for transformation-defective variants which are highly cytopathic in culture (4). The variant viruses have sustained point mutations in oligonucleotides 4 and 9 which are located in the 5′ region of the *env* gene (see Fig. 2). These oligonucleotides encode portions of the highly conserved amino-terminal portion of gp85. It is possible that this region of gp85 determines the flexibility of the protein structure which is necessary for cell membrane-bound gp85 to interact with adjacent cell surface receptors and establish superinfection resistance. Conceivably, mutations within the amino-terminal region result in a protein conformation incompetent to interact in this manner. The variation of the subgroup B sequence from that of the other subgroups from amino acids 58 to 76 within this region suggests a role for these differences in the cytopathic effect of subgroup B viruses.

**Identification of recombinant *env* sequences in NTRE-4.** Totest the role of hrl and hr2 in host range variation, the organization of the recombinant *env* sequence of NTRE-4 was determined. The gp85-encoding region of NTRE-4 was generated by two crossovers in the vicinity of bases 5770 and 5970 within regions of high nucleotide sequence homology (Fig. 2 and 3). Except for a single transition which does not produce an amino acid difference, the PR-RSV-B and NTRE-4 nucleotide sequences are completely homologous through the hrl-encoding region and then diverge for approximately 200 bases before returning to complete homology (Fig. 2). The 5′ end of this divergent region is marked by the presence of the subgroup E-specific oligonucleotide 06 in the NTRE-4 sequence. When the nucleotide sequence of RAV-0 was compared with that of NTRE-4, a region of complete homology including the hr2-encoding region was found to correspond exactly to the region of divergence from the Pr-RSV-B sequence (Fig. 3).

Thus, the comparative sequence analysis identified the recombination events within the *env* gene which generated the extended host range virus, NTRE-4, as two crossovers within regions of high sequence homology. These led to the insertion of approximately 200 bases of RAV-0-derived sequence, including the hr2-encoding region, into sequences completely homologous to that of Pr-RSV-B. RAV-0-derived sequence codes for the peptide region beginning around the arginine residue at position 182 through hr2 to the vicinity of the asparagine residue at position 255 (see Fig. 4). Since the host range of NTRE-4 encompasses that of both parental viruses, the subgroup B- and E-specific regions of its gp85 must both contribute directly to host range specificity. The subgroup E-specific portion of gp85 consists only of hr2, demonstrating a role for this region. The subgroup B-specific sequences include not only hrl but also the region from amino acids 58 through 76. For reasons discussed above, it is more likely that hrl is the region relevant to host range determination, but confirmation of this point requires the analysis of in vitro recombinant viruses. In any case, combination of variable regions is clearly responsible for the ability of the gp85 of NTRE-4 to recognize both the subgroup B receptor on chicken cells and the subgroup E receptor on turkey cells. This recombinant *env* gene organization underscores the role of more than one region in determining host range.

**Topography and function of gp85.** Figure 5 presents a schematic diagram of the primary structure of gp85 marking the positions of cysteine and proline residues and of potential glycosylation sites as deduced from the nucleotide sequences of subgroup C, B, and E viruses. hrl (32 amino acids) and hr2 (27 amino acids) are positioned approximately in the middle of the protein and are separated by a highly

conserved region of 38 amino acids. Within the sequence of gp85 covered by our comparative analysis, there are 12 glycosylation sites of the type Asn-X-Thr/Ser which are conserved among all three subgroups. In the sequences of all three subgroups the variable regions are bounded by, but do not contain, potential glycosylation sites.

The positions of cysteine and proline residues are also highly conserved throughout gp85. Particularly striking is the conservation of cysteine and proline residues within the variable regions. Within hr2 there are two cysteine residues and one proline residue which are conserved among all three subgroups. Within hr1, the subgroup B and E sequences display three common proline residues, whereas a fourth common proline residue is located immediately adjacent to the left-hand terminus of hr1. The subgroup C sequence, which contains less homology with the sequences of the other two subgroups within hr1, shares a single proline residue within the variable region and also has the proline residue located at the left-hand end of hr1. The high degree of conservation of these topographic markers on gp85 implies their importance in determining the conformation which gp85 adopts on the virion surface and indicates that the variable regions of all three subgroups assume similar orientations on the mature glycoprotein.

The conformation of gp85 may expose hr1 and hr2 on the surface of the molecule in a structure similar to that proposed for the hemagglutinin glycoprotein monomer of influenza virus (26). In such a structure, the variable determinants are located side by side at the tops of stem structures, forming a single binding site which is conformationally free to interact with receptor molecules. Although the mechanism by which the variable regions on gp85 recognize the appropriate receptors is unknown, several possibilities can be envisaged. In a manner similar to that of the hemagglutinin protein, hr1 and hr2 may form a single binding site. Alternatively, a two-step mechanism may be involved in which one variable region initiates contact with a cell surface receptor, bringing the second variable region into the proper orientation to complete the recognition process. The recombinant structure of the gp85 of NTRE-4 ($hr1_B$-$hr2_E$) shows that two variable regions of different subgroups can interact to recognize both types of receptor. Such an event may be possible in this instance due to the similarity of the subgroup E and B receptors and the homology of $hr1_B$ and $hr1_E$. In this regard, it is of interest to note that the variant *env* sequence found in a Pr-RSV stock as described by Hunter et al. (10) contains approximately 180 bases of complete homology with RAV-0, including the $hr2_E$-encoding region corresponding to bases 5810 through 5975. However, the rest of the sequence is divergent from that of RAV-0 and Pr-RSV-C and it is possible that this variant represents a similar type of recombinant between another avian leukemia virus strain and endogenous viral sequences. Further insight into the molecular basis of host range variation will result from the in vitro construction and analysis of recombinant viruses expressing novel combinations of hr1 and hr2.

## ACKNOWLEDGMENTS

We thank C. van Beveren for the modified sequencing protocol, B. Mermer and P. Norton for helpful discussions, P. N. Tsichlis and R. Guntaka for gifts of cloned viral DNA, and M. Bostic-Fitzgerald and S. Morrison for technical assistance.

This work was supported by Public Health Service grant R01-CA-17659 from the National Cancer Institute (to J.M.C.) and by a Damon Runyon-Walter Winchell Cancer Fund Fellowship, DRG-677 (to A.J.D.). J.P.S. was supported by a long-term fellowship from the European Molecular Biology Organization.

## LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
2. **Billeter, M. A., J. T. Parsons, and J. M. Coffin.** 1974. The nucleotide sequence complexity of avian tumor virus RNA. Proc. Natl. Acad. Sci. U.S.A. **71**:3560–3564.
3. **Coffin, J. M., M. Champion, and F. Chabot.** 1978. Nucleotide sequence relationships between the genomes of an endogenous and an exogenous avian tumor virus. J. Virol. **28**:972–991.
4. **Coffin, J. M., P. N. Tsichlis, C. S. Barker, S. Voynow, and H. L. Robinson.** 1980. Variation in avian retrovirus genomes. Ann. N.Y. Acad. Sci. **354**:410–425.
5. **Crittenden, L. B., and J. V. Motta.** 1975. The role of the *tvb* locus in genetic resistance to RSV (RAV-0). Virology **67**:327–334.
6. **Cullen, B. R., A. M. Skalka, and G. Ju.** 1983. Endogenous avian retroviruses contain deficient promoter and leader sequences. Proc. Natl. Acad. Sci. U.S.A. **80**:2946–2950.
7. **DeLorbe, W. J., P. A. Luciw, H. M. Goodman, H. E. Varmus, and J. M. Bishop.** 1980. Molecular cloning and characterization of avian sarcoma virus circular DNA molecules. J. Virol. **36**:50–61.
8. **Duff, R. G., and P. K. Vogt.** 1969. Characteristics of two new avian tumor virus subgroups. Virology **39**:18–30.
9. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
10. **Hunter, E., E. Hill, M. Hardwick, A. Bhown, D. E. Schwartz, and R. Tizard.** 1983. Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its products. J. Virol. **46**:920–936.
11. **Ishizaki, R., and P. K. Vogt.** 1966. Immunological relationship among envelope antigens of avian tumor viruses. Virology **30**:375–387.
12. **Joho, R. H., M. A. Billeter, and C. Weissman.** 1975. Mapping of biological functions on RNA of avian tumor viruses: location of regions required for transformation and determination of host range. Proc. Natl. Acad. Sci. U.S.A. **72**:4772–4776.
13. **Katz, R. A., C. A. Omer, J. H. Weiss, S. A. Mitsialis, A. J. Faras, and R. V. Guntaka.** 1982. Restriction endonuclease and nucleotide sequence analysis of molecularly cloned unintegrated avian tumor virus DNA: structure of large terminal repeats in circle junctions. J. Virol. **42**:346–351.
14. **Loenen, W. A. M., and W. J. Brammer.** 1980. A bacteriophage lambda vector for cloning large DNA fragments made with several restriction enzymes. Gene **20**:249–259.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
17. **Moscovici, C., M. G. Moscovici, H. Jiminez, M. M. C. Lai, M. J. Hayman, and P. K. Vogt.** 1977. Continuous tissue culture cell lines derived from chemically induced tumors of Japanese quail. Cell **11**:95–103.
18. **Pani, P. K.** 1976. Further studies in genetic resistance of fowl toRSV (RAV-0): evidence for interaction between independently segregating tumor virus b and tumor virus e genes. J. Gen. Virol. **32**:441–453.
19. **Payne, L. N., and P. K. Pani.** 1971. Evidence for linkage between genetic loci controlling response of fowl to subgroup A and subgroup C sarcoma viruses. J. Gen. Virol. **13**:253–259.
20. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
21. **Tozowa, H., H. Bauer, T. Graf, and H. Gelderblom.** 1970. Strain-specific antigen of the avian leukosis sarcoma virus group 1. Isolation and immunological characterization. Virology **40**:530–539.

22. **Tsichlis, P. N., and J. M. Coffin.** 1980. Recombinants between endogenous and exogenous avian tumor viruses: role of the C region and other portions of the genome in the control of replication and transformation. J. Virol. **33:**238–249.
23. **Tsichlis, P. N., K. F. Conklin, and J. M. Coffin.** 1980. Mutant and recombinant avian retroviruses with extended host range. Proc. Natl. Acad. Sci. U.S.A. **77:**536–540.
24. **Vogt, P. K.** 1977. Genetics of RNA tumor viruses, p. 341–455. *In* H. Fraenkel-Conrat and R. R. Wagner (ed.), Comprehensive virology, vol. 9. Plenum Press, New York.
25. **Vogt, P. K., and R. Ishizaki.** 1965. Reciprocal patterns of genetic resistance to avian tumor viruses in two lines of chickens. Virology **26:**664–672.
26. **Webster, R. G., W. G. Laver, G. M. Air, and G. C. Schild.** 1982. Molecular mechanisms of variation in influenza virus. Nature (London) **296:**115–121.
27. **Weiss, R.** 1982. Experimental biology and assay, p. 210–260. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Weiss, R. A.** 1969. The host range of Bryan strain Rous sarcoma virus synthesized in the absence of helper virus. J. Gen. Virol. **5:**511–528.
29. **Weller, S. K., A. E. Joy, and H. M. Temin.** 1980. Correlation between cell killing and massive second round superinfection by members of some subgroups of avian leukosis virus. J. Virol. **33:**494–506.
30. **Zasloff, M., G. D. Ginder, and G. Felsenfeld.** 1978. A new method for the purification and identification of covalently closed circular DNA molecules. Nucleic Acids Res. **5:**1139–1152.