DNA Sequence of the Bryan High-Titer Strain of Rous Sarcoma Virus: Extent of env Deletion and Possible Genealogical Relationship with Other Viral Strains

TEENA L. LERNER AND HIDESABURO HANAFUSA*

The Rockefeller University, New York, New York 10021

Received 11 August 1983/Accepted 18 October 1983

The genetic structure of the Bryan high-titer strain of Rous sarcoma virus (BH-RSV) was analyzed by using a molecular clone obtained from proviral DNA. DNA sequencing of the pol-src junction of BH-RSV revealed that the env sequence was almost entirely absent; only six base pairs following the pol termination codon remained. Beginning at nucleotide 7 (relative to the end of pol), a 91-base pair sequence identical to the 91 base pairs immediately upstream from src in other strains of RSV was found. The helper virus-related sequence of about 100 base pairs, which is present as a direct repeat in the 5′ and 3′ regions flanking src in other RSVs, was present only on the 3′ side of src in BH-RSV. The 3′ end of BH-RSV, from the last 16 base pairs of src through the U1 region, was virtually identical to a region downstream of env through U3 in the nontransforming helper virus Rous-associated virus-2, suggesting that BH-RSV may have been derived by recombination between Rous-associated virus-2 and cellular src DNA. The possibility that the original RSV may have been a defective transforming virus and a precursor of the nondefective RSV strains is discussed.

MATERIALS AND METHODS

Molecular cloning of BH-RSV proviral DNA fragments. A clone of 3Y-1 rat cells infected with BH-RSV was obtained from S. Kawai and grown in minimal essential medium as described previously (19). Total cellular DNA was extracted as described previously (26). DNA (350 μg) was completely digested with EcoRI and then fractionated on a sucrose gradient, and fractions containing the 5-kilobase (kb) viral pol-src fragment were identified. Insert DNA (0.075 μg) was ligated to 1 μg of EcoRI arms of the phage vector, lambda gt WES (2:1 molar ratio of arms to insert), and packaged in an in vitro lambda packaging system. Details of procedures for in vitro packaging, screening, phage purification, amplification, and preparation of phage DNA have been previously described (1, 16, 27). DNA fragments were purified by electrophoresis in agarose or polyacrylamide gels and electropheluted in dialysis tubing containing electrophoresis buffer.

The 5-kb EcoRI fragment from BH-RSV was purified from a lambda phage clone and subcloned into pBR322 according to standard procedures (27). For initial screening of clones, plasmid DNA of antibiotic-resistant colonies was isolated by the rapid alkaline lysis technique (2).

Enzymes and enzyme reactions. Restriction endonucleases were purchased from New England Biolabs or Bethesda Research Laboratories and were used as recommended by the manufacturer or as described previously (27). Usually, at least a twofold excess over the recommended amounts of enzyme was used for digestion of chromosomal or closed circular plasmid DNA. Nick translation, ligation with T4 DNA ligase (New England Nuclear Corp.), removal of 5′ phosphates with bacterial alkaline phosphatase (Bethesda Research Laboratories), and 5′ end-labeling of DNA fragments with T4 polynucleotide kinase were performed according to standard procedures (27).

Gel electrophoresis of DNA and filter hybridization. For agarose gel electrophoresis, DNA samples were run on
submerged horizontal gels in electrophoresis buffer (0.04 M Tris-acetate [pH 7.8], 0.005 M sodium-acetate, 0.001 M EDTA). For sizing restriction fragments smaller than 1 kb or as a preparative step in obtaining a labeled DNA fragment for sequencing, electrophoresis on polyacrylamide slab gels was performed. The concentration of polyacrylamide was 5 or 8% in TBE buffer (50 mM Tris [pH 8.3], 50 mM boric acid, 1 mM Na₂EDTA) as described by Maxam and Gilbert (28). Southern transfer (41) of DNA from agarose gels to nitrocellulose filters (Schleicher & Schuell, Inc.), hybridization to radioactive probes, and washing of filters were carried out as previously described (26).

**DNA sequencing and analysis of sequence data.** The chemical cleavage method of Maxam and Gilbert (28) with the hydroxylamine modification of Rubin and Schmid (33) for cleavage at cytosines was used throughout. All regions were sequenced at least twice (on both DNA strands, in most regions); the sequences of restriction enzyme cleavage sites were confirmed by DNA sequencing.

Analysis of DNA sequences for the presence of restriction enzyme cleavage sites or for identification of regions of extensive homology was facilitated by use of the computer programs SEARCH and SEQFIT, respectively (42).

**RESULTS**

**Molecular cloning of a 5-kb pol-src fragment of BH-RSV.** The EcoRI sites in the U₁ region and in gag are conserved in BH-RSV, whereas the EcoRI site in env is absent (26). Consequently, EcoRI digestion of BH-RSV proviral DNA yields two fragments, 2.5 and 5.0 kb (Fig. 1). By using total cellular DNA from 3Y-1 rat cells transformed by BH-RSV (19), the 5-kb EcoRI fragment which contains the region from pol to U₁ was molecularly cloned into a lambda phage vector and then subcloned into pBR322. The plasmid clone will be referred to as pH-beta.

**Restriction enzyme cleavage analysis of BH-beta.** To determine the extent of missing env sequences in BH-RSV, we digested pH-beta DNA with restriction enzymes which have cleavage sites in env (37, 44, 45). The BH-RSV cleavage sites were compared with published sequences (37, 44, 45) and used to construct a deletion map. The results of the restriction enzyme cleavage analysis are shown in Fig. 1.

BH-RSV fragments which were contained wholly in pol or src were found to be the same size as the corresponding fragments of SR-RSV or PR-RSV (data not shown), with the following exceptions. (i) An XhoI site, not found in PR-RSV or SR-RSV, was present in the pol gene of BH-RSV, ca. 800 base pairs (bp) from the 5' end of the 5-kb insert. (ii) A PvuII site in pol, present in PR-RSV, was not found in BH-RSV. (iii) A second BamHI site, mapping near the 3' end of the PR-RSV pol gene, was not present in BH-RSV. This site is also not present in SR-RSV (7). However, an additional BamHI site mapping near the 5' end of the pol gene was found in pH-beta. This site was not present in SR-RSV or PR-RSV; it was also not present in BH-RSV alpha (data not shown), a variant of BH-RSV which is defective in pol (12).

Restriction enzyme sites which map in either env or the intercistronic region between env and src were absent (data not shown). The NcoI site, shown in Fig. 1 to map at the border between the intercistronic region and src, was present in BH-RSV. This cleavage site includes the ATG initiation codon of the PR- and SR-RSV src genes (37, 45); thus, the presence of the NcoI site in BH-RSV suggested that the srccoding region was intact in BH-RSV.

The size of the pol-src junction fragment, bounded by the KpnI site in pol and the NcoI site at the 5' end of src, was determined to be ca. 300 bp (data not shown). The homologous fragment of PR-RSV or SR-RSV would be ca. 2.2 kb. If the 3' end of the BH-RSV pol gene is similar to that of PR-RSV, we expect that 200 bp of pol gene sequences would follow the KpnI site, allowing for no more than 100 bp of env or intercistronic sequences in BH-RSV.

**Sequence of the pol-src junction in BH-RSV.** The 300-bp KpnI-NcoI fragment was partially sequenced by the method of Maxam and Gilbert (28). The NcoI 5' end was labeled with 32P in a kinase reaction, and sequencing in the reverse direction (noncoding strand) from the initiation codon of src revealed that there were only 97 bp between the end of pol and the beginning of src in BH-RSV (Fig. 2B and 2C). The nucleotides preceding src in BH-RSV were found to be essentially identical to the 91 nucleotides preceding src in PR-RSV and SR-RSV (37, 45), in recombinant rASV1441 (44), and in c-src (46).

Immediately preceding this 91-bp region in BH-RSV was the sequence identified as the end of the pol gene, including the stop codon TAA and six nucleotides which follow pol in PR-RSV. The pol-src junction sequence was confirmed by sequencing with a DdeI site as the labeled end; partial

![Fig. 1. Structure of BH-RSV proviral DNA compared with nondefective RSV DNA. The restriction map of nondefective RSV DNA was compiled from previous reports (7, 15). Restriction enzyme cleavage sites on the nondefective RSV genome (37, 45), between the EcoRI site in gag and the EcoRI site in U₁, were examined for their presence in BH-RSV. The env gene was drawn to a larger scale. A (+) or (−) below a position indicates whether the site was present or absent in BH-RSV. The dotted line marking an enzyme site indicates a site which was present in BH-RSV but not present in PR-RSV. An extra BamHI site which was present in BH-RSV beta only (see text) is not shown; it mapped between X and Xb in the 5' portion of pol. Enzymes: E, EcoRI; Hp, HpaI; H, HindIII; X, XhoI; Xb, Xbal; B, BamHI; Bε, B*; N, NcoI; P, PvuII; A, AccI; Bg, BglII; K, KpnI; B, BglII.](http://jvi.asm.org/Downloaded from http://jvi.asm.org)
sequencing of the 3' end of the BH-RSV pol gene (about 150 bp) confirmed that the sequence was essentially identical to that of PR-RSV (data not shown).

3' end of BH-RSV. Preliminary restriction enzyme mapping analysis of the BH-RSV 5-kb EcoRI fragment indicated that the last 500 bp of this fragment were similar to the corresponding region of SR-RSV, except for a short region immediately after the src gene (Fig. 3). Hybridization experiments (data not shown) indicated that BH-RSV did not contain the sequence termed "F3" (the region immediately following src in SR-RSV) (46) or the sequence termed "F2" or "E" (the sequence found on the 5' side of src in SR-RSV and on the 3' side of src in PR-RSV) (37, 45).

The 3' end of the BH-RSV, from the last 200 bp of src to the EcoRI site in the U3 region, was sequenced according to the strategy depicted in Fig. 3; the results are shown in Fig. 4. The BH-RSV sequence was basically similar to the corresponding SR-RSV sequence, except in the region 118 nucleotides immediately after the stop codon of src. The sequence immediately after src in BH-RSV was unrelated to other known retroviral sequences. However, the Rous-associated virus-2 (RAV-2) viral genome has recently been sequenced by Bizet et al. (2a), and there is extensive sequence similarity in this region between BH-RSV and RAV-2. The RAV-2 DNA sequence, which was kindly provided to us, is included in Fig. 4. The homologous portion in RAV-2 is the region which begins ca. 70 bp downstream of the env gene. In BH-RSV, the homologous portion begins with the last 16 bp of the src gene (13 bp of coding sequence plus the termination codon). Included in these 16 bp are the last 8 bp of v-src (CTGAAATAG), which are not found in the region of c-src DNA which contains the 3' end of v-src (46) (Fig. 4).

Conservation of a helper virus-related sequence in avian retroviruses. One copy of "F1," the sequence present as a direct repeat on either side of src in SR-RSV and PR-RSV (6, 37, 45, 46), was found on the 3' side of src in BH-RSV (Fig. 4). (We have divided the F1 sequence into regions F1-A through F1-D as described below.) Since the F1 sequence is also present between env and U3 in Rous-associated virus-0 (RAV-0) (17), F1 is believed to be of avian leukosis virus (ALV) origin. The F1 sequence is present immediately upstream from U3 in every avian retrovirus for which the sequence in this region is known (22, 24, 34, 37, 39, 45). Figure 5 shows the F1 sequence present in
RAV-0, PR-RSV, SR-RSV, avian myeloblastosis virus, Fujinami sarcoma virus, Y73, and BH-RSV. In addition, F1 is also present in endogenous viral sequences ev1 and ev2 and in RAV-2 (A. M. Skalka et al., personal communication). Comparison of the F1 sequences in these viruses indicated that the sequence seemed to be composed of discrete blocks of conserved sequences which we termed F1-A through F1-D.

Figure 6 illustrates the relationship between the flanking regions of BH-RSV and those of two other RSVs and a helper virus. The first copy of the SR-RSV and PR-RSV F1 sequence (on the 5' side of src) seems to be complete, however, the second copy of F1 (on the 3' side of src) in these sarcoma viruses lacks the F1-A region. In PR-RSV, SR-RSV, and BH-RSV, the ca. 120 nucleotides immediately after src are unrelated to each other, and in BH-RSV this sequence is homologous to a sequence in RAV-2 as described above. The origin of the sequence immediately downstream from src in the other two sarcoma viruses is, at present, unknown.

**DISCUSSION**

**Structure of BH-RSV.** We estimate the size of the env deletion in BH-RSV, as compared with nondefective RSV, to be ca. 1.9 kb. The sequence data are in good agreement with previous size estimates based on restriction enzyme mapping analysis (2 kb) (38) but are slightly more extensive than the estimate based on oligonucleotide fingerprinting analysis (1.5 kb) (8).

Sequencing revealed the gp85 and gp37 coding sequences to be completely absent from BH-RSV. Only the portion of the env signal sequence which overlaps with the 3' end of pol (in another reading frame) (37), as well as 6 bp which follow the pol terminator, was present in BH-RSV. The fact that BH-RSV viral particles are produced (11, 31) even though the viral genome cannot direct the synthesis of any env-related gene products conclusively demonstrates that the env product is not required for viral assembly. For enveloped viruses, env proteins may play a role in assembly, as was proposed (3). However, alternative mechanisms of assembly which do not require the presence of viral glycoprotein must also exist.

The putative splice acceptor site for env mRNA maps within the 3' end of pol (37). BH-RSV may, therefore, synthesize two species of 21S src-containing mRNA which differ by about 135 nucleotides. The slight difference in size of these RNAs may have precluded their detection in previous experiments (14). The larger mRNA species would not be expected to direct the synthesis of a protein in the env reading frame since this frame is interrupted by several termination codons in the pol-src junction.

In addition to lacking env, BH-RSV does not contain the F1 sequence upstream from src. The region between pol and src in BH-RSV, therefore, does not share any homology with ALV, and this may explain the inability of BH-RSV to readily recombine with ALV to generate a nondefective RSV (20, 47).

Takeya and Hanafusa reported that the sequence of the c-src exons is highly conserved in the v-src genes of SR-RSV and rASV1441, and that the sequence conservation includes 91 bp of noncoding information immediately preceding c-src (45, 46). It was therefore proposed that position −91 relative to the start of c-src represents the site at which recombination occurred between a helper virus and the c-src gene to generate RSV (45). It is noteworthy that the pol-src junction in BH-RSV occurs at the very same position, −91 relative to src (Fig. 2A), and essentially the same 91-bp nucleotide sequence is conserved in BH-RSV. This may imply a requirement for this entire region since it contains a splice acceptor "consensus sequence" (25) which has been proposed to be the acceptor site for src mRNA (37, 44). Alternatively, structural features may be inherent in the DNA sequence in the vicinity of −91 which facilitate recombination at that site. As noted previously (45), the presence of four consecutive TCTGs in this vicinity of c-src DNA is striking.

To determine whether there is an inherent feature in the DNA sequence which renders the env gene susceptible to complete deletion, we began to examine the pol-src junction of another env-defective RSV, SR-NY8, an env deletion mutant derived from SR-RSV (21). Our preliminary studies suggest that the size of a pol-src junction fragment of SR-NY8 is about 100 bp longer than the corresponding fragment of BH-RSV (data not shown). We have not yet determined whether the extra sequence information present in SR-NY8 maps in pol, env, or 5' of src.

In our designation of the F1 region and its subdivision into F1-A through F1-D, we considered the RAV-0 DNA sequence (17) as our reference. F1 was found to be highly conserved among many retroviruses (Fig. 5), suggesting a functional role for this sequence. On the other hand, the sequence called F2 or E has been seen in only two nondefective RSVs. It is no longer clear whether the F2 sequence in retroviruses has any biological function since this sequence is absent from BH-RSV. Similarly, since BH-RSV contains only one copy of F1, a second copy found in SR-RSV and PR-RSV may not be necessary. Hughes and Kosik (personal communication) observed that in vitro mutagenesis of SR-RSV was performed such that the F2 sequence and one copy of F1 (upstream of src) were deleted, transforming activity and replication of the virus were not affected.

**Origin of BH-RSV.** We sought to determine which of the following models best accounts for the generation of BH-RSV: (i) BH-RSV was derived from nondefective RSV by
FIG. 4. DNA sequence of the 3’ end of BH-RSV. The sequence of the region of BH-RSV depicted in Fig. 3 was compared with other sequences: c-src, the 3’ end of the c-src gene (46); c-src’, a sequence 900 bp downstream of the 3’ end of the c-src (46); RAV-2, the sequence of the noncoding region between env and U₃ in RAV-2 (2a); SR-RSV, the sequence of the homologous region of SR-RSV (45) as depicted in Fig. 3. Amino acids shown represent carboxyl terminal residues of c-src’s. *, termination codon of src. The boxes indicate proposed recombination sites. F1-A through F1-D, see text for explanation. PPT, polypurine tract. *, same nucleotide as BH-RSV. Blank space indicates nucleotide not present; inserted nucleotides are indicated.

deletion, (ii) BH-RSV and nondefective RSV were each generated by independent recombination events between c-src and a helper virus, or (iii) BH-RSV (or a similar defective virus) was the original RSV generated by a recombination event between c-src and a helper virus, and nondefective RSVs were derived from the defective virus by acquisition of env. Analysis of the sequence data could not rigorously exclude any of the above possibilities; however, the striking sequence relationship between c-src, RAV-2, and BH-RSV favors the recombination models (ii) and (iii).

No extensive sequence homologies were observed between the pol-src junction which would clearly support a model in which an env deletion occurred by homologous recombination. However, some short homologies were observed. The sequence immediately after pol in PR-RSV (TTATATTCTATTAT) shares 10 out of 15 bp with a sequence upstream from src (TAATATTCTATTAT) in RAV-2 (Fig. 2D). A homologous recombination occurring within this region would delete env and generate the sequence of the BH-RSV pol-src junction. Similarly, the sequence immediately after pol in PR-RSV (TCTATATTCTATTAT) shares 12 out of 19 bp with the sequence upstream from src in c-src DNA (CTCTATATTCTATTATGCTAT). Recombination between c-src and the sequence after pol in ALV (and a 6-bp deletion) would generate the BH-RSV pol-src junction. Presently, PR-RSV is the only avian retrovirus for which the pol-env sequence is known; as other sequence data become available, more obvious sequence homologies may become evident. On the other hand, illegitimate recombination may have been involved in the generation of BH-RSV (9).

The sequence of the 3’ end of BH-RSV failed to demonstrate a relationship between BH-RSV and either of the two nondefective viruses, PR-RSV or SR-RSV (Fig. 4 and 6). The observed homologies between RAV-2, c-src, and BH-RSV in this region support the hypothesis that BH-RSV arose via direct recombination between c-src and RAV-2.

Takeya and Hanafusa proposed that a recombination event between the last exon of c-src (c-src in Fig. 4) and a sequence 900 bp downstream of c-src (c-src in Fig. 4) may have occurred to generate the 3’ end of v-src (46). However, the resulting recombined sequence still lacks the last 8 bp of the v-src coding sequence. An additional recombination...
FIG. 5. Comparison of the Fl sequence of RAV-0 with that of several avian transforming retroviruses. RAV0, RAV-0 sequence shown is the 3' end of env through U3 (17). Z, sequence between the termination codon of env and start of Fl in RAV-0. Other regions are the same as described in legends to Fig. 2 and 4. 5'SR and 5'PR, 5' side of src in SR-RSV and PR-RSV, respectively. Sequence shown is the 3' end of env through the env-src intercistronic region (37, 45). 3'BH, 3'SR, and 3'PR: 3' side of src in BH-RSV, SR-RSV, and PR-RSV, respectively. The sequence shown is downstream of src through U1 (Fig. 1 and 6) (37, 45). In avian myeloblastosis virus (AMV), Y73, and Fujinami sarcoma virus (FSV), the sequence shown is downstream of the onc sequence through U1 (22, 24, 34, 39). *, same nucleotide as RAV-0. Blank spaces indicate nucleotides which were not present. Inserted nucleotides are indicated.
event utilizing the 9- out of 11-bp homology between c-src’ and RAV-2 would specifically generate the 3’ end of the BH-RSV genome, i.e., the last 8 bp of src as well as the next ~250 bp (or more) after src.

Since RAV-2 was isolated from a BH-RSV stock (10), one might argue that it was RAV-2 which obtained the 3’ end sequence from BH-RSV. However, Bizub et al. (2a) found extensive homology in this region between RAV-2 and Y73 virus, a virus isolated in a different time and place (18). This strongly suggests that the sequence found downstream of env in RAV-2 was present in natural helper viruses rather than derived from RSV. Since the sequence comparisons in Fig. 4 suggest a homologous recombination between the 3’ end of c-src and RAV-2, it would also be of interest to examine the pol-env region of RAV-2 for evidence of recombination events which may have occurred between the 5’ end of c-src and RAV-2.

Although all strains of RSV are considered to be derived from Rous “chicken tumor #1” (32), an accurate pedigree for most strains is only partially documented (29, 40). Furthermore, it was reported that there were no longer any remaining samples of chicken tumor #1 (40). The oldest available sample is RSV-29, which had been frozen in 1929 and recovered in 1963 (40). RSV-29, the virus which is thus probably the closest to chicken tumor #1, appears to be defective in replication (P. J. Simons, personal communication; H. Hanafusa, unpublished results). The Harris strain which is derived from the same line as RSV-29 is also replication defective (30). Together with the defectiveness of BH-RSV, these findings support the hypothesis that RSV was originally defective, like many other transforming retroviruses. The nondefective SR-RSV and PR-RSV strains may represent another independent isolate or may have been acquired in the laboratory as a result of in vivo passage of the virus through heterologous hosts or interaction with novel helper viruses, or both.

The differences in sequence organization surrounding the src genes of the three transforming viruses shown in Fig. 6 may be due to their different in vivo and in vitro passage histories. Schwartz et al. propose a structure of the ancestral RSV which contains two copies of all the sequence elements; they suggest that each strain arose by deletion of certain elements (37). We propose another possibility: the ancestral RSV was similar in structure to BH-RSV and contained only few of these sequence elements. During passage of the virus, new flanking sequences were inserted. For instance, a rare recombination event between the –91 SA region of BH-RSV and the 5’ end of the ALV U3 region would generate the structure found in the 5’ side of src in PR-RSV (Fig. 6). The generation of a nondefective RSV genome presumably involves two independent recombinations between retroviruses and the c-src (43, 46). Thus, the idea that BH-RSV is a precursor of nondefective RSV is consistent with this scheme.

ACKNOWLEDGMENTS

We thank S. Kawai for the gift of BH-RSV-transformed 3Y1 rat cells. T. Takeya for advice in sequencing, and A. M. Skalka and S. H. Hughes for providing unpublished data on RAV-2 sequences and on in vitro-constructed RSV mutants, respectively.

This work was supported by Public Health Service grants CA14935 and CA18213 from the National Cancer Institute. T. L. L. was a recipient of Public Health Service training grant T32AI07233 from the National Institutes of Health.

LITERATURE CITED


