The nucleotide sequence of an untranslated but conserved domain at the 3' end of the avian sarcoma virus genome

A.P.Czernilofsky, W.DeLorbe, R.Swanstrom, H.E.Varmus, J.M.Bishop, E.Tischer* and H.M.Goodman*

Department of Microbiology and Immunology, and *Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Received 10 April 1980

ABSTRACT

The genomes of numerous avian retroviruses contain at their 3' termini a conserved domain denoted "c". The precise boundaries and function of "c" have been enigmas. In an effort to resolve these issues, we determined the sequence of over 900 nucleotides at the 3' end of the genome of the Schmidt-Ruppin subgroup A strain of avian sarcoma virus (ASV). We obtained the sequence from a suitable fragment of ASV DNA that had been cloned into the single-stranded DNA phage M13mp2. Computer-assisted analysis of the sequence revealed the following structural features: i) the length of "c" - 473 nucleotides; ii) the 3' terminal domain of src, ending in an amber codon at the 5' boundary of "c"; iii) terminator codons that preclude continuous translation from "c"; iv) suitably located sequences that may serve as signals for the initiation of viral RNA synthesis and for the processing and/or polyadenylation of viral mRNA; v) a repeated sequence that flanks src and that could facilitate deletion of this gene; vi) repeated sequences within "c"; and vii) unexplained homologies between sequences in "c" and sequences in several other nucleic acids, including the 5' terminal domain of the ASV genome, $tRNA^{Trp}$ and its inversion, the complement of $tRNA^{Trp}$ and its inversion, and the 18S RNA of eukaryotic ribosomes. We conclude that "c" probably does not encode a protein, but its sequence may nevertheless serve several essential functions in viral replication.

INTRODUCTION

The haploid genome of avian sarcoma virus (ASV) is a single-stranded RNA composed of ca. 9500 nucleotides (1). This RNA fills two roles during the viral life cycle. First, it serves as mRNA for the synthesis of several viral polypeptides (1); accordingly, the RNA is capped (2) and polyadenylated (3) at the 5' and 3' termini, respectively. Second, the viral genome is the template for synthesis of viral DNA by reverse transcriptase (1,4); the product of this synthesis is integrated into the genome of the host cell and transcribed into progeny viral RNA (1,5). Two structural features of the viral genome have been implicated in the synthesis of viral DNA: a molecule of $tRNA^{Trp}$ located near the 5' end of the viral RNA serves as primer

for reverse transcriptase (6); and a sequence of 16-21 nucleotides, repeated directly at the two ends of the viral genome (7,8,28), mediates early events in propagation of the nascent DNA chain (4,7-9,28; R.S., manuscript in preparation). The integrated DNA (or "provirus") of ASV is coextensive with viral RNA but, in addition, contains a 330bp terminal redundancy composed of nucleotide sequences derived from both the 3' and 5' ends of the viral genome (10,11). The redundant sequences may serve either or both of two purposes: they may mediate integration; and they may harbor the promoter for viral RNA synthesis.

The genome of ASV contains four identified genes, arranged as follows on the viral RNA: 5'- gag-pol-env-src -3' (1). The genes gag, pol and env are required for viral replication, whereas src is devoted exclusively to neoplastic transformation of the host cell (12). In addition, structural studies have identified a conserved domain located to the right of src, composed of ca. 500 nucleotides, and denoted "c" because of its presence in numerous strains of avian retroviruses (13,14).

Two distinctive functions have been attributed to "c". On the one hand, recombination analysis indicated that the sequence may be important for viral replication (15,16). On the other hand, some investigators have sought to implicate "c" in leukemogenesis by avian leukosis viruses, which otherwise have no apparent genetic locus devoted to tumorigenesis (1,17,18). In order to explore the structural boundaries and possible functions of "c", we have determined the sequence of over 900 nucleotides at the 3' end of the genome of the Schmidt-Ruppin subgroup A strain of ASV (SR-A ASV).

Our experimental strategy exploited the fact that we had previously cloned the entire genome of SR-A ASV into both phage and plasmid vectors (19). From these clones, we isolated a 3.2kb Eco RI fragment of DNA that contained all but 65 nucleotides of "c", the entirety of src, and ca. 1.0kb to the left of src. We then cloned this DNA into the single-stranded DNA phage M13mp2 (20,21) in order to facilitate sequence analysis by the "chain terminator technique" developed by Sanger and colleagues (22). Supplementary data were also obtained by the sequencing procedure of Maxam and Gilbert (23), and with DNA isolated from plasmid sub-clones of SR-A ASV DNA. Our results indicate that "c" is probably not expressed by translation into a protein. In

addition, we have found suitably located nucleotide sequences that may serve as signals for the initiation of viral RNA synthesis and the polyadenylation of viral RNA. We encountered unexpected and non-random homologies between sequences in "c" and sequences in $tRNA^{Trp}$, its polynucleotide complement, and the 5' end of the ASV genome. Several of our conclusions conform to a recent report that contained the sequence of 217 nucleotides in the "c" domain of the Schmidt-Ruppin subgroup D strain of ASV (24).

MATERIALS AND METHODS

Materials

Most of the materials have been described previously (19-23,25). All of the restriction enzymes were from New England Biolabs Inc. and were used as recommended. The synthetic dodecanucleotide d(TCACGCAGTTGT) was obtained from Collaborative Research. This oligonucleotide is complementary to a sequence within the lac region of M13mp2 and can be used as a universal primer for the sequencing of DNA joined to the lac region at the Eco RI cloning site (20,21). The Klenow fragment of polymerase I and polynucleotide kinase were from Boehringer Mannheim, and the α^{32}P-dATP and γ^{32}P-ATP from Amersham (specific activity ca. 400 mCi/mmol and 3000 Ci/mmol respectively).

Cloning procedure: The viral DNA of SR-A was cloned previously into the vector λgtWES·λB (19). Treatment of the cloned DNA with Eco RI generates the following fragments: 2.0kb, 3.9kb, 3.2kb, 0.33kb, and 0.26kb. We used for our cloning purposes the fragment of 3.2kb. DNA for cloning was usually obtained by Eco RI treatment of the recombinant phage DNA and subsequent separation of the fragments on a 1% agarose gel. The DNA was localized by staining with ethidium bromide and by reference to appropriate markers and was then recovered by electroelution (25). DNA was cloned into M13mp2 and recovered as recombinant phage genome or replicative form as described previously (20,21,25). The phage vector and its host, E. coli strain JM101, were provided by J. Messing. All work with recombinant DNA was performed according to NIH guidelines in certified P2 and P3 facilities.

Hybridization procedures: For hybridization with DNA transferred to nitrocellulose filters we used two different probes: i) $cDNA_{rep}$, which was prepared as described previously (26,27), using the genome of Prague-C ASV as template. This cDNA contains mainly single-stranded

DNA of negative polarity (ie., complementary to the ASV genome) and was used for screening recombinant plaques and for identification of the polarity of the single-stranded viral DNA cloned into M13mp2. ii) A probe for nucleotide sequences represented in the 3.2kb Eco RI fragment of ASV DNA was prepared by transcribing the denatured fragment with reverse transcriptase in the presence of random oligodeoxynucleotide primers, as described previously (19,27,28). DNAs were transferred from agarose gels to nitrocellulose filters and hybridized with radioactive cDNAs (29).

Isolation of primers for sequencing reactions: The 3.2kb fragment of ASV DNA was isolated from the lambda clone described above, from a sub-clone in pBR322, or from the replicative form of a sub-clone in M13mp2. The purified fragment was cleaved with suitable restriction endonucleases; the products of digestion were fractionated by electrophoresis in gels of polyacrylamide or agarose, located by staining with ethidium bromide and by reference to standard markers, and recovered from slices of the gel by electroelution (25).

Sequencing techniques: We generally used chain termination with dideoxynucleotides exactly as described by Sanger et al (22), except that we used α^{32}P-dATP as the radioactive precursor in all reactions. At the conclusion of the polymerase reactions, primers were usually removed from the newly synthesized DNA by cleavage with the appropriate restriction endonuclease (one unit of enzyme per 20 microliters of reaction mixture, five minutes at 37^{o}). Three fragments were also sequenced by the Maxam-Gilbert technique (23), which was used without any modification. The DNA fragments were labeled on the 5'-end with γ^{32}P-ATP and polynucleotide kinase under standard conditions (23), or they were labeled on their 3'-ends by copying the 5' "overhangs" of restriction sites with polymerse I (Klenow fragment) under the same conditions as we used for the sequencing reaction described above.

RESULTS

Isolation of ASV DNA for sequencing: molecular cloning in the single-stranded DNA phage M13mp2.

We have previously cloned DNA representing the genome of SR-A ASV in the vector gtWES B (19). Hydrolysis of the chimeric DNA with Eco RI cleaves the ASV component four times. Among the products of

cleavage is a 3.2 kb DNA fragment that extends from the 65th residue of "c" (numbering from the 3' terminus of the viral RNA genome) to ca. 1.0kb to the 5' side of src (19). In work to be reported elsewhere, we used transfection of cultured fibroblasts to demonstrate that this DNA fragment is capable of directing the synthesis of the protein encoded in src, and thus, of transforming recipient cells to a neoplastic phenotype (P. Luciw et al., manuscript in preparation); hence, the structure and function of the src gene (and by inference, the remainder of the DNA fragment) survived cloning and propagation in prokaryotic vectors.

We isolated the 3.2kb DNA fragment as described under Materials and Methods; analysis by electrophoresis in an agarose gel demonstrated that the isolated fragment was homogeneous and pure (Figure 1a, lane 1). The fragment was then cloned into the Eco RI site of M13mp2. In order to document the composition of positive clones, we isolated the replicative form DNA of recombinant phage, cleaved the DNA with Eco RI, and demonstrated that the products of cleavage included a 3.2kb fragment that hybridized with ASV $cDNA_{rep}$ (Figure 1a, lane 2). We concluded that the viral DNA had survived the sub-cloning without sustaining an appreciable deletion.

In preliminary studies of a set of sub-clones, we found that the single-stranded genome of the chimeric phage invariably contained only the negative strand of ASV DNA; the predominance of one polarity in recombinant clones of M13 has been a recurrent problem in the use of this vector (unpublished observations of the authors). In an effort to obtain both polarities of ASV DNA in the phage genome, we isolated the replicative form of one of the recombinant clones, cleaved the DNA with Eco RI, religated the DNA and used it to transform the bacterial host JM 101. Clones containing viral DNA were again isolated, and some were found to contain a 3.2kb Eco RI fragment that hybridized with $cDNA_{rep}$ (data not shown).

The presence of inserted DNA in the genomes of these new phage clones was demonstrated by electrophoresis of the single-stranded DNA in an agarose gel; the genomes of phage derived from apparently recombinant clones migrated appreciably slower (Figure 1b, lanes 2 and 3) than the genome of the phage itself (Figure 1b, lane 1). The electrophoretic mobility of the chimeric DNAs conformed to the anticipated size of 10.4kb.

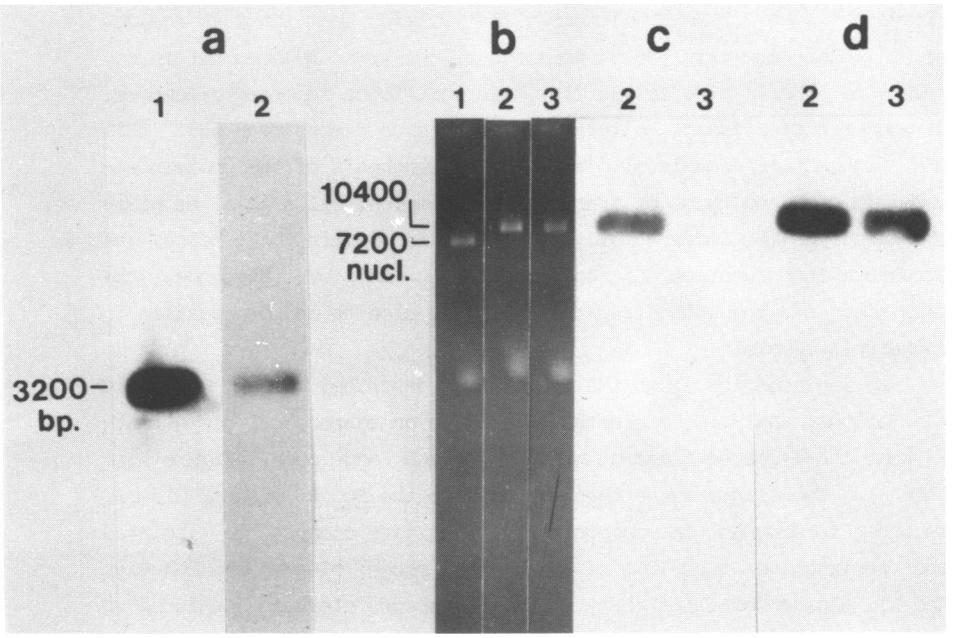

Figure 1. Preparation of cloned DNA for sequence analysis.

DNAs were isolated, cleaved with restriction endonucleases, fractionated by electrophoresis in agarose gels, transferred to nitrocellulose filters and analysed by molecular hybridization as outlined under Materials and Methods.

Panel a. Sub-cloning the Eco RI 3.2kb fragment of ASV DNA. Lane 1, the 3.2kb fragment prior to sub-cloning into M13mp2. The 3.2kb fragment of ASV DNA was purified following cleavage of recombinant DNA containing the entire ASV genome with Eco RI. A portion of the isolated fragment was analysed by electrophoresis in a gel of 1% agarose. The figure illustrates an autoradiogram obtained following transfer of the DNA to a nitrocellulose filter and hybridization with $cDNA_{rep}$. Lane 2, the 3.2kb fragment following sub-cloning into M13mp2. The DNA fragment analysed in lane 1 was sub-cloned into M13mp2 at the Eco RI site. Replicative form of the sub-clone was isolated, cleaved with Eco RI and analysed as in lane 1. The figure illustrates an autoradiogram obtained following hybridization of the immobilized DNA with $cDNA_{rep}$.

Panel b. Detection of inserts in the genome of M13mp2. Single-stranded DNA was extracted from phage, fractionated by electrophoresis in a gel of 1% agarose and detected by staining with ethidium bromide. Lane 1, the genome of M13mp2. Lanes 2 and 3, the genomes of phage containing inserts of ASV DNA.

Panel c. Detection of recombinant clones of M13mp2 containing positive strands of ASV DNA. The DNAs of lanes 2 and 3 in panel b were transferred to nitrocellulose filters and hybridized with $cDNA_{rep}$; the figure illustrates autoradiograms obtained subsequent to the hybridization.

Panel d. Detection of sub-clones of M13mp2 containing negative

strands of ASV DNA. The filters used in panel c were subsequently hybridized with radioactive DNA representing both strands of the 3.2kb Eco RI fragment of ASV DNA; the figure illustrates an autoradiogram obtained after the hybridization.

We determined the polarity of the ASV inserts in the genomes of chimeric phage by hybridizing the fractionated DNA with either $cDNA_{rep}$ (which reacts principally with ASV nucleic acids of positive polarity) or a radioactive and denatured preparation of the original 3.2kb ASV DNA fragment (which will react with ASV nucleic acids of both polarities). Representative results are illustrated in Figure 1c and d: the genome of one recombinant phage hybridized with $cDNA_{rep}$ and thus contained ASV DNA of positive polarity (Figure 1c); another recombinant genome hybridized only with the cDNA made with the 3.2kb fragment and thus contained ASV DNA of negative polarity (Figure 1d, right lane). Subsequent analyses of nucleotide sequence confirmed these assignments. Assured that we now possessed recombinant clones that could be used for sequencing both strands of the ASV DNA fragment under study, we proceeded with the sequence analysis.

Nucleotide sequence of the "c" domain of the ASV genome.

Figure 2 illustrates our experimental strategies for determining the nucleotide sequence reported below. We relied principally upon chain termination with dideoxynucleotides, as described by Sanger and colleagues (22). We attempted to improve the accuracy of our results by using at least two independent means to reproduce all of the sequence: usually, we sequenced overlapping regions by using two different primers in the terminator technique; in all regions, the sequence was determined in both polarities; and in a few instances, we obtained supplementary data with the procedure of Maxam and Gilbert (23). The 3.2kb fragment of DNA used in the present analyses does not include the last 65 nucleotides at the 3' terminus of the ASV genome. The sequence of these residues has been determined in a separate study which will be reported elsewhere (R.S., manuscript in preparation), and similar results have been obtained previously for the Prague subgroup C strain of ASV (unpublished results of A.P.C. and J. Shine, and personal communication from D.Schwartz and W. Gilbert); for clarity, these data will be included in the discussion that follows.

The data obtained as outlined above provided us with the sequence of 936 nucleotides at the 3' end of the SR-A ASV genome. Figure 3 il-

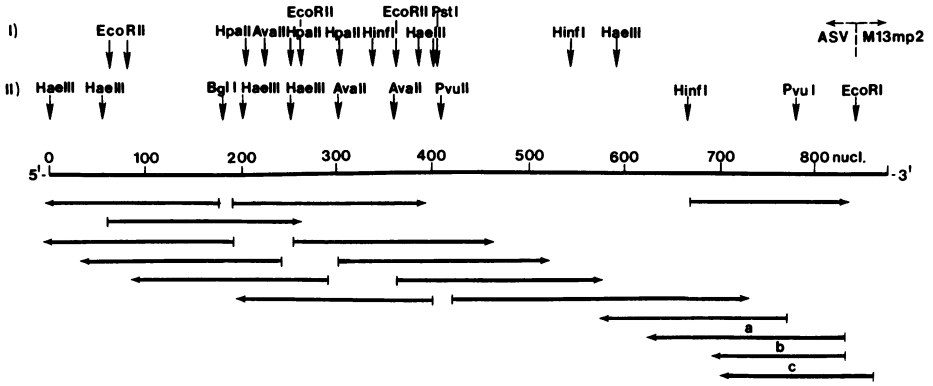

Figure 2. Strategy for determining the nucleotide sequence of "c".

The map coordinates begin arbitrarily with the first nucleotide at the 5' boundary of the sequence determined in the present study and extend beyond the junction between the cloned ASV DNA and the Eco RI site of the M13 vector DNA. The map is aligned according to the polarity of the ASV genome (plus strand). Restriction sites are located by arrows above the map: row II gives the positions of sites used in sequencing the DNA, row I gives additional sites deduced from the final sequence. The arrows below the map define the regions sequenced with individual primers and fragments. The direction of the arrows designates the polarity of DNA used as template in the terminator procedure: to the left, minus strand; to the right, plus strand. Additional designations: a, a sequence determined with the Maxam-Gilbert technique; the fragment was labelled at the 3' terminus; b, a sequence determined with the Maxam-Gilbert technique; the fragment was labelled at the 5' terminus; c, a sequence obtained by using a synthetic primer that initiates DNA synthesis 20 nucleotides to the right of the junction between ASV DNA and M13 DNA.

lustrates this sequence and joins it to the sequence found at the 5' end of ASV RNA (8,30); we chose this alignment for illustration in order to facilitate discussion of important structural and functional aspects of the nucleotide sequence. The genome of ASV has a direct terminal redundancy (7,8); this sequence appears once in Figure 3 (positions 917-936), linking the 3' and 5' domains of the viral genome, in accord with the structure of DNA transcribed from viral RNA (R.S., manuscript in preparation).

Structural features of the sequence.

i.) The topography of the sequence was not disturbed by molecular cloning.

The SR-A ASV DNA from which we prepared our sub-clones has been

mapped extensively by site-specific cleavage with restriction endonucleases (19). Our sequence contains all of the sites located previously by this means. In the process of preparing suitable primers for sequencing with the terminator technique, we identified a number of additional restriction sites; these also appear at appropriate positions in the the final sequence. Since we prepared most of the primers from the original clone of DNA in lambda phage rather than from our sub-clones in M13, we are further assured that the sub-cloning did not cause major deletions or rearrangements in the DNA. Figure 2 summarizes the locations of cleavage sites for most of the available restriction endonucleases; a complete catalogue of restriction sites within the sequence is available from the authors on request.

ii.) Correlations with previous data.

Previous reports have described oligonucleotides obtained from the "c" region by hydrolysis of ASV RNA with T1 RNase (13-15). The most characteristic of these is an oligonucleotide whose composition has proven to be remarkably constant among numerous strains of avian retroviruses (13-15,31,32). The sequence illustrated in Figure 3 contains this oligonucleotide at positions 904 to 915, a location that is in accord with previous, less precise efforts to map the oligonucleotide on the viral genome (13,14). We were able to detect other similarities between the regions of the sequence reported here and oligonucleotides described in the previous literature, but the homologies were frequently incomplete and are not located in Figure 3. These findings conform to previous reports that changes in single nucleotides are frequent among the genomes of related retroviruses (13,14,31,32).

iii.) The boundaries of "c".

The boundaries of "c" have in the past been defined operationally by deletions that affect uncertain portions of src (13,27,31,32). Here we define the domain as it is usually represented on formal maps of the ASV genome (1,12): bounded on the right by the poly(A) at the 3' terminus of the ASV genome, and on the left by the termination of src. The former is situated at position 936 in Figure 3. Can we locate the latter? Termination codons are scattered throughout two of the three reading frames in Figure 3, but the third frame (designated #1) is open from position 1 at the left of the sequence to position 460, where it is interrupted by an amber codon. We suggest that this

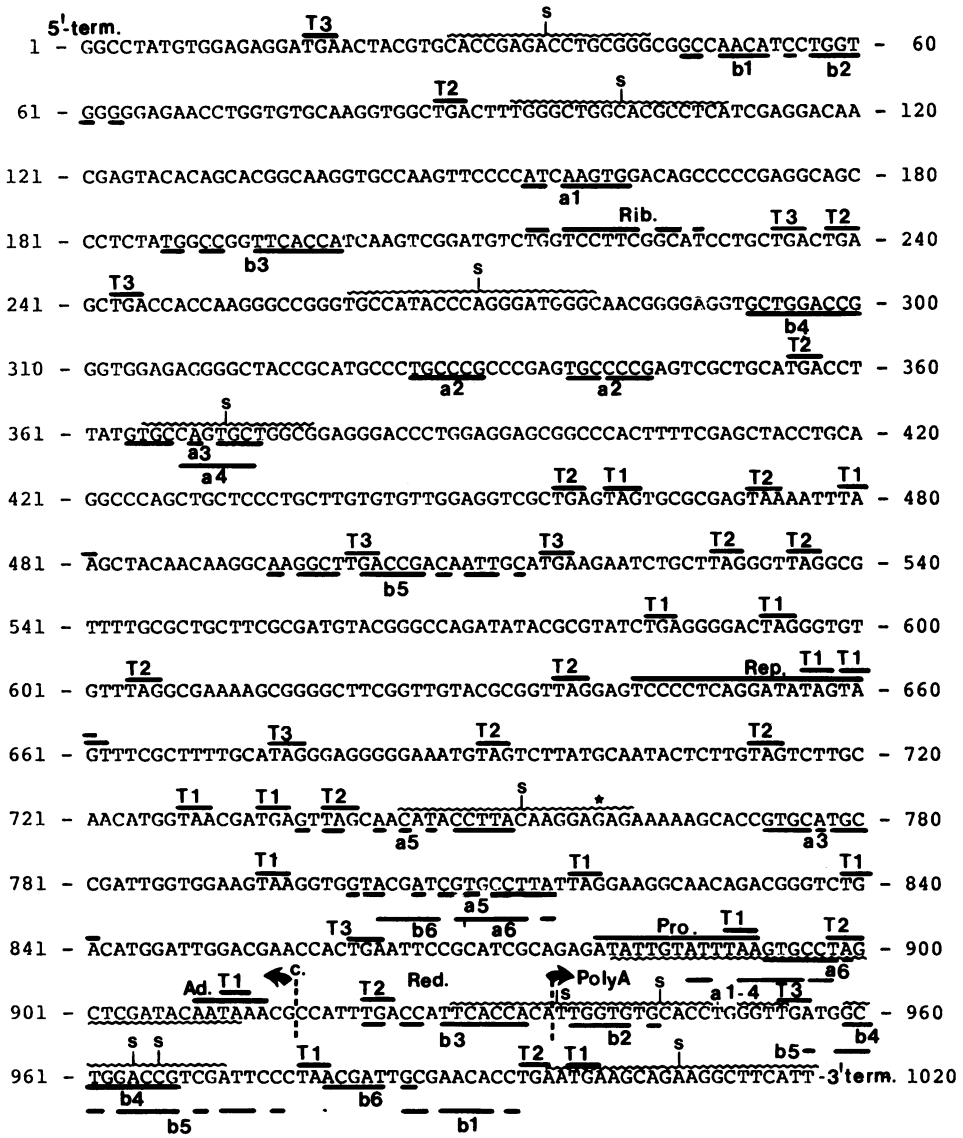

Figure 3. Nucleotide sequences in 3' and 5' domains of the ASV genome.

The illustration joins sequences at the 3' and 5' ends of the ASV genome as they appear in DNA transcribed from the genome by reverse transcriptase. The 3' terminus of viral RNA (i.e., the site of polyadenylation) is at position 936 (denoted by an arrow labelled polyA), the 5' terminus at position 916 (denoted by c, for "capping site"); the overlap between the 3' and 5' terminal sequences (posi-

tions 917-936, denoted Red.) represents the terminal redundancy of the viral genome, only one copy of which appears in the DNA synthesized by reverse transcriptase (R.S., manuscript in preparation). The sequences were derived as follows: positions 1-868, from the present study; positions 869-936, from a separate analysis of cloned ASV DNA that will be described elsewhere (R.S., manuscript in preparation); and positions 937-1020, from previous analyses of DNA transcribed from the 5' end of the genome of the Prague-C strain of ASV (30). Additional symbols: i.) T1, T2 and T3, termination codons in three reading frames; ii.) s, regions of dyad symmetry; iii.) Rib., a sequence that is complementary to the 3' end of 18S ribosomal RNA; iv.) Rep., a sequence repeated to the 5' side of src; v.) Pro., a possible component of the promoter for viral RNA synthesis; vi.) underlinings marked "a" denote oligonucleotides repeated within the 3' domain of the sequence; numbers identify specific oligonucleotides; vii.) underlinings marked "b" denote oligonucleotides found in both the 3' and 5' domains of the sequence; numbers identify specific oligonucleotides; viii.) Ad., the sequence alleged to signal processing and/or polyadenylation of mRNAs (positions 908-914); ix.) the underlining of positions 870-912 demarcates a region that can fold into a stable hairpin structure to include the alleged promoter sequence; and x.) the * at position 759 denotes an ambiguity in the sequence (G or T).

codon is likely to mark the 3' terminus of src; our provisional sequence of the entire src gene sustains this suggestion (work in progress). We conclude that the "c" region of SR-A ASV is probably 473 nucleotides in length. This conclusion is in accord with previous estimates obtained by mapping deletion mutants with restriction endonucleases (27), whereas analysis of heteroduplexes by electron microscopy had suggested a somewhat larger value (33).

iv.) Can the "c" region be translated?

The distribution of termination codons within the "c" region precludes continuous translation from any of the reading frames. This conclusion is based on the assumption that termination at the codon UGA - which is the predominant terminator codon in the sequence - is not suppressed at an appreciable frequency; recent findings have raised the possibility of such suppression in eukaryotic cells but have not assessed its prevalence (34; and B. Cordell et al, manuscript submitted). If translation is circumscribed by UGA codons, three regions within "c" could nevertheless give rise to polypeptides of appreciable size: (a) Positions 515-586; the polypeptide would consist of 23 amino acids with the following sequence: met lys asn leu leu arg val arg arg phe ala leu leu arg asp val arg ala arg tyr thr arg ile . (b) Positions 558-677; the polypeptide would consist of 39 amino acids

with the following sequence: met tyr gly pro asp ile arg val ser glu gly thr arg val cys leu gly glu lys arg gly phe gly cys thr arg leu gly val pro ser gly tyr ser ser phe ala phe ala. (c) Positions 777-863; the polypeptide would consist of 28 amino acids with the following sequence: met pro ile gly gly ser lys val val arg ser cys leu ile arg lys ala thr asp gly ser asp met asp trp thr asn his. The first of these possible polypeptides would be notably hydrophobic and basic, the composition of the others would be less distinctive.

It is of course possible that RNA splicing could rearrange the sequence of "c" so as to facilitate translation into a protein of reasonable size, but we and others have failed to detect a suitable mRNA in infected cells (35,36). We therefore conclude that "c" is unlikely to encode a protein larger than 39 amino acids unless UGA codons within the sequence are suppressed in eukaryotic cells. Previous proposals that translation from "c" might account for leukemogenesis are probably incorrect. It remains possible, however, that the domain gives rise to at least one of the small polypeptides described above; the factors that stimulate growth of eukaryotic cells include proteins of this size (37).

v.) Polyadenylation of viral RNA.

The sequence AAUAAA appears near the 3' terminus of most, if not all, eukarytic mRNAs and is thought to be the signal for some or all of the events that generate the polyadenylated 3' terminus of mature mRNA (38,39). The sequence of "c" contains this alleged signal at positions 909-914, or 23-28 residues from the 3' terminus of the viral genome as represented in Figure 3. This location is similar to that found in other eukaryotic mRNAs (38-42).

vi.) A signal for the initiation of viral RNA synthesis?

A portion of "c" is contained within the 330bp terminal redundancy that brackets the integrated provirus of ASV and allegedly harbors the promoter for viral RNA synthesis (10,11). We have isolated the terminal redundancy and analysed its nucleotide sequence; the results of these studies will be reported and discussed in full elsewhere (R.S., manuscript in preparation). Here we explore the relationship of these findings to the composition of "c".

In Figure 3, we have aligned the nucleotide sequences as they occur in the terminal redundancy, with the 3' end of "c" fused to the 5' end of the viral genome (10,11,17). According to this scheme, viral

RNA synthesis might begin at the 5' terminus of the genome proper (position 917) in response to signals possibly contained within the 3' domain of the terminal redundancy. We have therefore searched the sequence of "c" for oligonucleotides implicated previously in the initiation of mRNA synthesis by eukaryotic cells (38-42). A sequence bearing features of the canonical eukaryotic "promoter site" occurs at positions 880 to 893 in Figure 3, 24 nucleotides from the alleged site of initiation of RNA synthesis. The candidate signal for initiation can be included in a large hair-pin structure that spans positions 881 to 912 (see Figure 3). Several elements of dyad symmetry (denoted s in Figure 3) are located downstream from the sequence. These findings are all in accord with previous descriptions of sites where the synthesis of eukaryotic mRNA is initiated and encourage us to examine more rigorously the possibility that the promoter for ASV RNA synthesis resides in the terminal redundancy of viral DNA.

vii.) Secondary structure in "c".

Computer-assisted analysis of the sequence reported here revealed ca. 200 possibilities for the formation of hairpin duplexes containing eight or more base-paired nucleotides (not illustrated). We have no means at present by which to discern whether any of these structures actually exist in the native RNA, or whether any are of functional significance. We also found a number of regions that display appreciable dyad symmetry; Figure 3 illustrates ten of the more extensive examples.

viii.) A repeated nucleotide sequence that flanks src.

By comparison of the sequence presented here with our unpublished data, we have identified a sequence of at least twenty nucleotides (positions 643 to 662) that is repeated directly and most likely precisely in an apparently untranslated region to the 5' side of src. Moreover, the precise repeats are each extended in both directions by adjacent, partially repeated (90%) sequences composed of ca. 20 nucleotides. Hence, the repeated domains include at least 60 nucleotides. We tentatively propose that this repeated sequence might mediate deletion of src by homologous recombination within the ASV genome, or by some other mechanism. This proposal could account for the remarkable frequency with which src is deleted during the propagation of most strains of ASV (12).

ix.) Homology with 18S ribosomal RNA.

It is thought that complementary interactions between mRNA and 18S ribosomal RNA may figure in the synthesis of proteins in eukaryotic cells (43). We therefore searched the nucleotide sequence in Figure 3 for relationships to 18S ribosomal RNA and found that the sequence at positions 215 through 228 is complementary at 11 out of 14 nucleotides to the sequence at the 3' terminus of 18S ribosomal RNA. The significance of this finding is presently moot because we have no evidence that ribosome binding occurs anywhere within the domain of "c".

x.) Homologies with RNA.

A number of tRNAs from the host cell can base-pair with the genome of ASV (44,45). The significance and location of this base-pairing has been established in only one instance - $tRNA^{Trp}$, which is situated near the 5' terminus of the viral genome and serves as primer for reverse transcriptase (6). (Available data do not exclude the possiblility that $tRNA^{Trp}$ also binds to the genome at other sites.) We searched the sequence in Figure 3 for relationships to $tRNA^{Trp}$ and discovered non-random homologies with the sequence of $tRNA^{Trp}$, its inversion, the complement of $tRNA^{Trp}$, and its inversion; these homologies are illustrated in Figure 4. We also searched published sequences of the untranslated regions of globin mRNA and SV40 DNA for similar relationships to $tRNA^{Trp}$ and found none. Other investigators, however, have reported partial homologies with tRNAs in the gene for 23S ribosomal RNA of bacteria (46,47), in sequences that bracket the gene for cytochrome II oxidase in mitochondrial DNA of mice (38), and in the genome of polyoma virus (48).

xi.) Homologies with the 5' end of the ASV genome.

The genome of ASV contains a direct terminal redundancy (7,8). This fact prompted us to search for further evidence of relatedness between the 3' and 5' domains of viral RNA. Although our search was constrained by the limited amount of data available for the 5' end of the viral genome, we nevertheless found six regions of appreciable homology between the 5' and 3' domains (designated "b" in Figure 3). The significance of these findings is presently moot.

xii.) Repeated nucleotide sequences within the 3' domain.

The sequence from position 1 to 917 in Figure 3 contains six oligonucleotides of various sizes that are each partially repeated within the sequence (denoted by "a" in the figure). The repetitions were

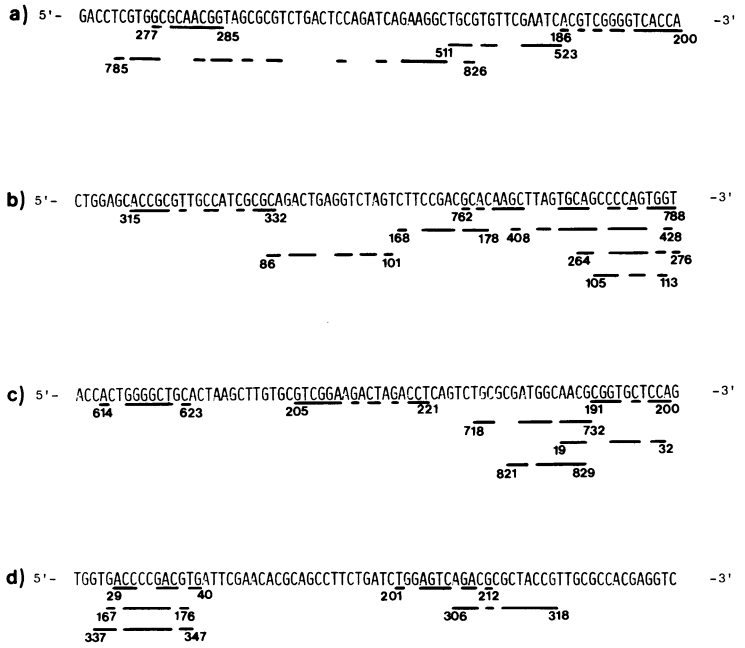

Figure 4. Homologies between the 3' domain of the ASV genome and $tRNA^{Trp}$.

Underlinings denote sequences that are also present in the sequence illustrated in Figure 3; numbers denote the positions in Figure 3. a) The sequence of chicken $tRNA^{Trp}$, rendered in deoxynucleotides; b) the complement of the sequence in a; c) inversion of the sequence in a; d) inversion of the sequence in b.

found both within "c" and in the portion of the sequence that we believe to be the 3' domain of src.

Concluding remarks

It appears unlikely that "c" encodes a protein, but the sequence may nevertheless serve essential functions in viral replication: previous recombination analyses indicated that the ability of avian retroviruses to replicate to high titers segregated with a particular form of "c" (16), and the data presented here implicate "c" in several crucial portions of the viral life cycle. To date, three forms of "c" have been identified in avian retroviruses: one whose prototype is carried by ASV and related viruses; a second whose prototype is in the genome of the endogenous avian retrovirus, RAV-0 (16,49); and a third,

recently identified in the genome of avian myeloblastosis virus and one of its helper viruses (unpublished results of T. Gonda and J.M.B.). It may prove instructive to compare the nucleotide sequences of these three forms of "c".

The avian leukosis viruses readily induce lymphomas in birds, yet apparently do not have a genetic locus devoted specifically to oncogenesis (1,12). Moreover, deletion of src from the genome of ASV deprives the virus of its ability to induce sarcomas, but gives rise to a leukosis virus (12,50). These findings have been attributed in the past to the possibility that "c" might encode an oncogenic protein; the present report calls this explanation into serious doubt. It remains possible, however, that "c" participates in leukemogenesis. We have argued elsewhere that the promoter for retrovirus RNA synthesis may, on occasion, induce transcription of otherwise silent cellular genes (in the manner of the "floating promoter" effect of some tranposable elements in bacteria (51)), and that induction of cellular genes in this manner could contribute to both leukemogenesis and cocarcinogenesis by retroviruses that do not possess specific oncogenes (52). If this suggestion is correct, and if the composition of "c" bears on the relative efficiency of retrovirus promoters, then some forms of "c" may be associated with leukemogenesis and others not. Again, comparative studies of "c" in the genomes of leukemogenic and non-leukemogenic avian retroviruses may be informative.

ACKNOWLEDGEMENTS

This research was supported by grants from the National Cancer Institute and the American Cancer Society. The authors thank Barbara Baker, Paul Luciw, Greg Payne and Richard Parker for gifts of cloned DNA, Hugo Martinez for assistance in the use of his computer program, R. Friedrich for assistance with statistical computations, J. Messing for M13mp2 and E. coli JM101, and D. Schwartz for personal communications of data prior to publication. We also thank B. Cook for helping us with the manuscript.

REFERENCES

1. Bishop, J.M. (1978). Ann. Rev. Biochem. 47, 35-88.
2. Furuichi, Y., Shatkin, A.J., Stavnezer, E. and Bishop, J.M. (1975). Nature 257, 618-620.
3. Wang, L.H. and Duesberg, P. (1974). J. Virol. 14, 1515-1529.
4. Haseltine, W.A. and Baltimore, D. (1976) in ICN-UCLA Symposia on Molecular and Cellular Biology, D. Baltimore, A.S. Huang, and C.F. Fox, Eds. Vol. 4, 175-214 (Academic Press, New York).

5. Varmus, H.E., Vogt, P.K. and Bishop, J.M. (1973). Proc. Nat. Acad. Sci. U.S.A. 70, 3067-3071.
6. Taylor, J.M. (1977). Biochim. Biophys. Acta 473, 57-72.
7. Schwartz, D.E., Zamecnik, P.C. and Weith, H.L. (1977). Natl. Acad. Sci. U.S.A. 74, 994-998.
8. Haseltine, W.A., Maxam, A.M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 989-993.
9. Coffin, J.M. and Haseltine, W.A. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 1909-1912.
10. Hughes, S.H., Vogt, P.K., Shank, P.R., Spector, D., Kung, H.-J., Breitman, M.L., Bishop, J.M. and Varmus, H.E. (1978). Cell 15, 1397-1410.
11. Sabran, J.L., Hsu, T.W., Yeater, C., Kaji, A., Mason, W.S. and Taylor, J.M. (1979). J. Virol. 29, 170-178.
12. Vogt, P.K. (1977) in Comprehensive Virology, H. Fraenkel-Conrat and R.R. Wagner, Eds. Vol. 9, 341-456 (Plenum Press).
13. Wang, L.H., Duesberg, P., Beemon, K. and Vogt, P.K. (1975). J. Virol. 16, 1051-1070.
14. Coffin, J.M. and Billeter, M.A. (1976). J. Mol. Biol. 100, 293-318.
15. Tsichlis, P.N. and Coffin, J.M. (1980). J. Virol. 33, 238-249.
16. Tsichlis, P.N. and Coffin, J.M. (1979). Proc. Nat. Acad. Sci. U.S.A. 76, 3001-3005.
17. Purchio, A.F., Erikson, E. and Erikson, R.L. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 4661-4665.
18. Crittenden, L.B., Hayward, W.S., Hanafusa, H. and Fadly, A.M. (1980). J. Virol. 33, 915-918.
19. DeLorbe, W.J., Luciw, P., Varmus, H.E. and Bishop, J.M. (1980). Submitted to J. Biol. Chem.
20. Gronenborn, B. and Messing, J. (1978). Nature 272, 375-377.
21. Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977). Proc. Nat. Acad. Sci. U.S.A. 71, 3612-3616.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
23. Maxam, A.M. and Gilbert, W. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 560-564.
24. Yamamoto, T., Jay, G. and Pastan, I. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 176-180.
25. Cordell, B., Bell, G., Tischer. E., De Noto, E.M., Ullrich, A., Pictet, R., Rutter, W.J. and Goodman, H.M. (1979). Cell 18, 533-543.
26. Friedrich, R., Kung, H.-J., Baker, B., Varmus, H.E., Goodman, H.M. and Bishop, J.M. (1977). Virology 79, 198-215.
27. Shank, P.R., Hughes, S.H., Kung, H.-J., Majors, J.E., Quintrell, N., Guntaka, R.V., Bishop, J.M. and Varmus, H.E. (1978). Cell 15, 1383-1395.
28. Taylor, J.M. (1977). Biochem. Biophys. Acta 473, 57-51.
29. Southern, E.M. (1975). J. Mol. Biol. 98, 503-517.
30. Shine, J., Czernilofsky, A.P., Friedrich, R., Bishop, J.M. and Goodman, H.M. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 1473-1477.
31. Wang, L.-H., Duesberg, P.H., Kawai, S. and Hanafusa, H. (1976). Proc. Nat. Acad. Sci. U.S.A. 73, 447-451.
32. Wang, L.-H., Duesberg, P., Mellon, P. and Vogt, P.K. (1976). Proc. Nat. Acad. Sci. U.S.A. 73, 1073-1077.
33. Junghans, R.P., Hu, S., Knight, C.A. and Davidson, N. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 477-481.
34. Geller, A.I. and Rich, A. (1980). Nature 283, 41-46.
35. Weiss, S.R., Varmus, H.E. and Bishop, J.M. (1977). Cell 12, 983-992.
36. Hayward, W.S. (1977). J. Virol. 241, 47-63.
37. Gospodarowicz, D. and Moran, J.S. (1976). Ann. Rev. Biochem. 45, 531-558.

38. Barrell, , B.C., Bakier, A.T. and Drouin, J. (1979). Nature 282, 189-194.
39. Honjo, T., Obata, M., Yamawaki-Kataoka, Y., Kataoka, T., Kwakami, T., Takahashi, N. and Madd, Y. (1979). Cell 18, 559-568.
40. Soeda, E., Arrand, J.R., Smolar, N. and Griffin, B.E. (1979). Cell 17, 357-370.
41. Arrand, J.R., Soeda, E., Walsh, J.E., Smolar, N. and Griffin, B.E. (1980). J. Virol. 33, 606-618.
42. Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979). Nature 278, 428-434.
43. Hagenbuchle, O., Santer, M., Steitz, J.A. and Mans, R.J. (1978). Cell 13, 551-563.
44. Sawyer, R.C. and Dahlberg, J.E. (1973). J. Virol. 12, 1226-1237.
45. Taylor, J.M., Cordell, B., Rohde, W., Goodman, H.M. and Bishop, J.M. (1975). Virology 65, 248-259.
46. Dahlberg, J.E., Kintner, C. and Lund, E. (1978). Proc. Nat. Acad. Sci. U.S.A. 75, 1071-1075.
47. Brosius, J., Dull, T.J. and Noller, H.F. (1980). Proc. Nat. Acad. Sci. U.S.A. 77, 201-204.
48. Soeda, E., Arrand, J.R., Smolar, N., Walsh, J.E. and Griffin, B.E. (1980). Nature 283, 445-453.
49. Coffin, T.M., Champion, M. and Chabot, F. (1978). J. Virol. 28, 972-991.
50. Biggs, P.M., Milne, B.S., Graf, T. and Bauer, H. (1973). J. Gen. Virol. 18, 399-403.
51. Starlinger, P. and Saedler, H. (1977) in Current Topics Microbiol. Immunol. 75, 111-140.
52. Bishop, J.M., Courtneidge, S.A., Levinson, A.D., Oppermann, H., Quintrell, N., Sheiness, D.K., Weiss, S.R. and Varmus, H.E. (1979). Cold Spring Harbor Symposium, Vol. 44, in press.