JOURNAL OF VIROLOGY, Mar. 1992, p. 1726–1736
0022-538X/92/031726-11$02.00/0

Vol. 66, No. 3

ART-CH, a New Chicken Retroviruslike Element

ANDREI V. GUDKOV,†* ELENA A. KOMAROVA, MIKHAIL A. NIKIFOROV, AND TATIANA E. ZAITSEVSKAYA‡

Laboratory of Molecular Genetics, Institute of Carcinogenesis, Cancer Research Center, Kashirskoye Shosse 24, Moscow 115478, Russia

Received 18 June 1991/Accepted 16 December 1991

A 3′ region of a previously unknown retroviruslike element named ART-CH (avian retrotransposon from chicken genome) was obtained in the course of polymerase chain reaction-mediated cloning of avian leukosis virus long terminal repeats (LTRs) from DNAs of infected chicken cells. About 50 copies of ART-CH are present in the genome of chickens of different breeds. ART-CH is not found in DNA of quails, ducks, turkeys, or several other birds tested. The ART-CH element is about 3 kb in size, including 388 bp LTRs. The major class of ART-CH-specific RNA, also 3 kb in size, is detected in various organs of chickens. An ART-CH polypurine tract, a $tRNA^{Trp}$-binding site, regions around the TATA box and polyadenylation signal, and the beginning of the putative *gag* gene strongly resemble the corresponding regions of avian leukosis viruses and EAV, the two described classes of chicken retroviruses. An open reading frame capable of encoding a polypeptide with a putative transmembrane domain is located upstream of the right ART-CH LTR. This sequence, as well as the U3 and U5 regions of the ART-CH LTR, has no obvious similarities with the corresponding parts of other known vertebrate retroviruses and retrotransposons. A short sequence upstream of the right LTR of ART-CH is very similar to sequences which flank the 3′ ends of the oncogenes v-*src*, v-*myc*, v-*fps*, and v-*crk* in four different recombinant avian retroviruses and which are absent from the genomes of other studied avian retroviruses. Thus, ART-CH is a new endogenous chicken provirus that may participate in the formation of recombinant oncogenic retroviruses.

Retroviruslike elements are widely distributed in eukaryotic genomes. They have been found in the DNA of all well-studied organisms, including yeasts (18), insects (26), and birds and mammals (9, 28, 40). Many of these elements seem to originate from germ line infection and remain capable of being expressed and taking part in recombination processes, thereby causing genetic variability either of somatic and germinal cells or of other retroviruses (33–36).

Only two classes of retroviruslike elements in the chicken genome have previously been described: *ev* loci and EAV elements. The *ev* loci were among the first endogenous proviruses characterized; some of them encode infectious endogenous retroviruses closely related to avian leukosis viruses (ALVs) (Rous-associated virus-0 [RAV-0], for example), while others are defective (3, 16, 17, 19, 20). EAVs have been recently found in chicken DNA by hybridization under nonstringent conditions with ALV-specific probes (8, 11, 12). They are typical proviruses distantly related to ALVs (5). Both of these retrovirus families putatively originated as a result of separate germ line infections. EAV appears in all *Gallus* species, consistent with germ line infection prior to specification, while the *ev* loci are specific for chickens and hence seem to be acquired later (15, 27).

Here we describe a new chicken retroviruslike element, which we named ART-CH, that was found by analysis of an unexpected polymerase chain reaction (PCR) product possessing some features of retroviral long terminal repeats (LTRs). This product was synthesized with primers corresponding to the conserved sequences around the U3 region of the ALV LTR and with ev^- chicken DNA as the template. It was used as a hybridization probe for isolation of homologous sequences from a chicken genomic library. Analysis of the cloned DNA fragments showed that they contained a new class of transcribed chicken-specific, repetitive retroviruslike elements that have short regions of similarity with the ALV genome.

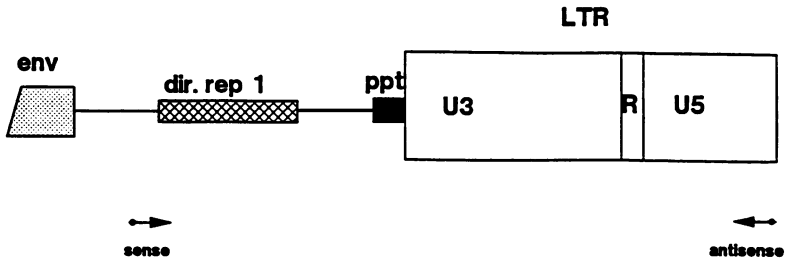
MATERIALS AND METHODS

Cells and viruses. Cultures of chicken embryo fibroblasts were prepared from trypsinized 9-day-old embryos of ev^- Brown Leghorn chickens bred and described previously (8). They were infected by different strains of ALV [RAV-50, RAV-1, Rous sarcoma virus (RSV) B77, myeloblastosis-associated virus MAV-1, and MAV-2 (N)] which were obtained as supernatants of culture fluids from virus-infected fibroblasts and stored in liquid nitrogen until use. Forty-eight to seventy-two hours after infection, cells were used for DNA isolation. The retroviruses used were a kind gift from Jan Svoboda.

DNA and RNA isolation, Southern and Northern (RNA) blot hybridization, and screening of the genomic library. These procedures were performed according to standard protocols (24). A genomic library of partially *Eco*RI-digested Brown Leghorn chicken DNA cloned in the lambda phage vector EMBL3 was kindly provided by A. Tikhonenko.

PCR. PCR was carried out with 1 μg of DNA isolated from infected and noninfected chicken embryo fibroblasts, 1 U of *Taq* polymerase (Ferment, Vilnius, Lithuania), and reaction kits as recommended by Perkin-Elmer/Cetus in a final volume of 50 μl. Two minutes of denaturation at 94°C was followed by two cycles which each included 1 min of denaturation at 94°C, 2 min of primer annealing at 47°C, and 2 min of extension at 72°C. This was followed by 30 similar cycles in which annealing was done for 2 min at 57°C. PCR

* Corresponding author.
† Present address: Department of Genetics, University of Illinois at Chicago, 808 S. Wood St., Chicago, IL 60612.
‡ Present address: Fred Hutchinson Cancer Research Center, Seattle, WA 98104.




```
5'                          3'
                ----------------> direct repeat 1
CT TGCGAATCGGGTTGTAACGG GGCAAGGC    RAV-0
CT TGCGAATCGGGTTGTAACGG G-CAAGGC    RAV-1
CT TGCGAATCGGGTTGTAACGG G-CAAGGC    RAV-2
CT TGCGAATCGGGcTGTAACGG GGCAAGGC    AMV
Cc TGCGAATCGGGcTGTA-CGG GcCA-Gat    PR-C-RSV
Cc TGCGAgTCGGGTTGTAACGG G-CAcatg    SR-A-RSV
   <------------------>
   sense primer sequence
```

Sense primer: 5'-GAATCGATTGCGAATCGGGTTGTAACGG-3'
Anti-sense primer: 5'-AGATCGATCTGCTTCATTCAGGTGTTCG-3'
Cla I

FIG. 1. PCR primers used for isolation of ALV LTRs and their positions on an ALV provirus. Alignment of sequences located at the beginning of direct repeat 1 (dir. rep 1) regions of different avian retroviruses is shown (in the case of RSVs, sequences shown correspond to the left copies of direct repeat 1). A 20-nucleotide conserved region was chosen for the sense primer. AMV, avian myeloblastosis virus; ppt, polypurine tract.

primers (amplimers) were synthesized with an Applied Biosystems DNA synthesizer (model 380A). The antisense amplimer was composed of a *Cla*I site followed by residues 90 to 71 of the genome of RSV strain PR-C (30) (antisense strand). This sequence corresponds to the 3′ end of the U5 region and is identical in all of the available sequences of ALV LTRs. The sense primer corresponds to the conserved sequence located closely upstream of the direct repeat 1 sequence of ALV genomes (Fig. 1).

Cloning of PCR products. PCR products were isolated from agarose gels with a GeneClean kit (Bio 101), digested with *Cla*I (100 U of the enzyme per μg of DNA), and cloned into *Cla*I-digested and alkaline phosphatase-treated vector plasmid pBluescript (SK+).

Sequencing of DNA. DNA sequencing was done with a Sequenase kit (U.S. Biochemicals). Universal and reverse 17-mer oligonucleotides and the oligonucleotides derived from already determined sequences were used as primers in sequencing reactions.

Nucleotide sequence accession numbers. The GenBank accession numbers for the ART-CH sequences and flanking regions presented here are M85057 and M85058.

RESULTS

Isolation of an LTR-like PCR product. The primary aim of this work was to gather a collection of LTRs of different ALVs by PCR. Since we were most interested in the U3 region, two conserved sequences surrounding U3 were chosen for primer synthesis. The left (sense) primer included the beginning of the so-called direct repeat sequence (Fig. 1), which is duplicated in oncogene v-*src*-containing avian sarcoma virus genomes (10, 30). A single copy of this highly conserved region is present in the genomes of nontransforming strains of ALV (4, 23). The right (antisense) primer corresponds to the 3′ end of the U5 region of the ALV LTR. Both primers were synthesized with *Cla*I restriction sites at their 5′ ends for cloning of PCR products (Fig. 1).

DNA from ev^- chicken embryo fibroblasts (8) infected by different avian retroviruses was used as the template in PCRs. The genome of ev^- chickens is free of *ev* loci, endogenous proviruses that are closely related to ALV (3, 16, 19, 20). The expected 400-bp PCR product is efficiently amplified by the chosen pair of primers with ev^+ DNA as the template (Fig. 2). The other known class of chicken endogenous proviruses, the EAVs, has no significant sequence homology with these primers (5). Hence, we could expect that ALV proviruses would be the only templates amplified by PCR under the conditions used.

We planned to use one universal pair of primers for the isolation of LTRs of different ALVs which could vary in their sequences even in highly conserved regions. To this end, we chose a comparatively low temperature of annealing (47°C) for the initial rounds of PCR. These PCR conditions made it possible to synthesize the expected products corresponding to LTRs of exogenous or endogenous (in the case of ev^+ DNA) viruses (Fig. 2). However, in most of the reactions, an additional 570-bp product was synthesized (Fig. 2). This product was more abundant in reactions using template DNA from noninfected ev^- cells, probably because ALV proviruses effectively competed for the primers. The 570-bp product did not appear if annealing was performed at 55°C. It was isolated from an agarose gel and cloned after *Cla*I digestion into the vector plasmid pBluescript. Inserts of

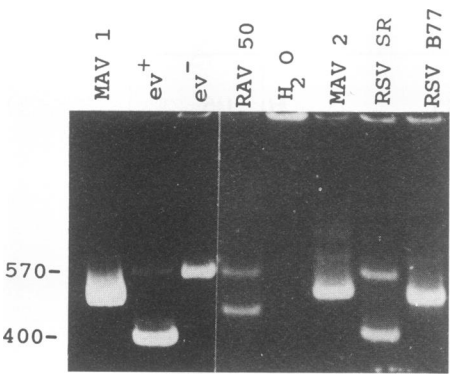


FIG. 2. PCR products synthesized by using ALV LTR primers (see Materials and Methods and Fig. 1) on different templates. DNA templates are indicated at the top of the figure. MAV 1, RAV 50, MAV 2, RSV SR, and RSV B77 correspond to DNAs from ev^- chicken embryo fibroblasts infected by the indicated viruses; ev^- represents DNA from control noninfected cells; ev^+ represents DNA of a Brown Leghorn chicken carrying the *ev-3* locus; H_2O represents a control reaction with no template. PCR products were separated on a 12% acrylamide gel and stained by ethidium bromide. Sizes of products are indicated at the left in base pairs.

two independently derived plasmids were sequenced from both universal and reverse M13 sequencing primers.

Analysis of this sequence showed that it had a number of short regions of strong homology with the functional sequences of the ALV genome (Fig. 3). One of these regions closely resembles the polypurine tract of ALV, and another resembles its polyadenylation signal together with the first half of the R region. Sixteen base pairs upstream of the putative polyadenylation signal, there is a sequence similar to a TATA box. The region between nucleotides 70 and 130 is similar to the ALV sequence located upstream of the right LTR. These similarities, as well as homologous arrangements of probable functional regions, strongly supported the idea that the fragment cloned is part of a right LTR (LTR X) of an unknown element.

Cloning of ART-CH element from chicken DNA. Using the cloned PCR product as a hybridization probe, we have isolated a number of recombinant phages from a chicken genomic library obtained from a partial *Eco*RI digest of ev^+ Brown Leghorn chicken DNA. *Eco*RI fragments of phage DNA which hybridized to the LTR X probe were partially sequenced either from an internal sequencing primer or from the universal primer after cloning in pBluescript. The schematic structure of the isolated phages and a partial sequence are presented in Fig. 4 and 5.

All three phages contained sequences almost identical to LTR X. Phages 5 and 15 contained one copy of LTR X, while phage 14 contained two copies of the LTR that were separated by a 2.8-kb sequence. The sequence between LTRs is present in the other two phages (phages 5 and 15) between the LTR X sequences and phage arms (Fig. 4). In both cases, the *Eco*RI site within this 2.8-kb region had been used for recombinant phage formation. Two base pairs downstream of the left LTR is an 18-bp region that is identical to the $tRNA^{Trp}$ primer-binding site of ALVs (phages 5 and 14). Immediately upstream of the right LTR is a polypurine tract-like sequence (phages 14 and 15). Hence, it

```
"LTR X"

   left primer
        10          20          30          40          50          60
ATCGATCGAA TCGGGTTGTA ACGGGCAAGC TTGGCAGGCA CGGAAAAGAA TTCCCTGTTG

        70          80          90         100         110         120
CTCTGATGAT TGCTTAAGAA CTGTAGTAGA AAATAGTAGG AATAGTGTGC TGAAATATAT

       130         140         150         160         170         180
TTAGGATTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGT TAGGTGTGTG TGTAAGTAGT

       190         200         210         220         230         240
ATTTAGCTTA GGGAGGGGGA GATGTTGTAG TAGGCGTCTT GCGGGGGCAC GGGATGTACG
          poly-purine tract ?
       250         260         270         280         290         300
GGACAGGCCT CTCCCTAAAC ATAGAGAGAT AGTGCTATCG TGCTGACCTT GTTGCAGAGA

       310         320         330         340         350         360
AAACAGGAGA AGAAGAAGGA TGATAAAAGA ATGTGGAAAC GGCCAAATAA GGCACAATGT

       370         380         390         400         410         420
TATCTGGTGT GAACTAATCA GAGTGGGACA TGACAGCACG GTTATCTAGG TAAAAATGTA
                                                              "TATA-
       430         440         450         460         470         480
TATAAGCTGT GTTTAGTAGT GAATAAACGC CATTTGCCTC ACTTACTCCT GGGGTCTGGG
box?                          poly(A)-signal?
       490         500         510         520         530         540
TGAGCATCTG GCCCCGACCT GGTAAAGGGT CGGTTTCGCC CAGCAGTAAG CCCTACATGT

       550         560         570    576
GGACAGAGGA CGAACACCTG AATGAAGCAG ATCGAT
            right primer
```

FIG. 3. Primary structure of the 570-bp PCR product. Regions of similarity with ALV sequences are underlined. The putative polypurine tract, TATA box, and polyadenylation [poly(A)] signal are marked.

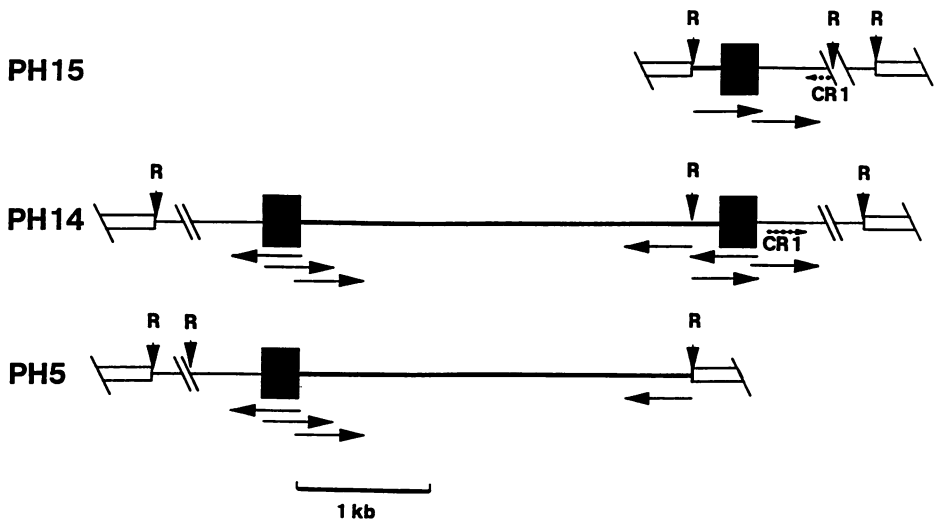


FIG. 4. Schematic structures of the recombinant lambda phage (PH) DNA fragments carrying sequences of LTR X. The scheme is based on the results of restriction mapping, cross-hybridization of DNA fragments, and partial sequencing. Arrows show the directions of sequencing from appropriate sequencing primers. LTRs and regions of identity are marked by thick lines. *Eco*RI restriction sites (R) and the positions of CR1 elements are indicated. Vector phage lambda EMBL3 arms are shown by open bars.

seems likely that phage 14 contains a complete retroviruslike element, while the two other phages (5 and 15) carry 3′ and 5′ parts of it, respectively. There is no similarity outside the elements in the three clones analyzed except for short regions downstream of the right LTR in phages 14 and 5. These regions correspond to the chicken repetitive element CR1 (37) and are present in different orientations (Fig. 4 and 5). Thus, the cloned retroviruslike elements belong to at least two different parts of the chicken genome, since they are surrounded by different sequences.

The element we found was named ART-CH (for avian retrotransposon from chicken genome). Sequence data for ART-CH and the results of comparisons with the known sequences of other avian retroviruses are shown in Fig. 5.

Like other retrotransposon LTRs, ART-CH LTRs start from TG, end with CA, and are framed by short inverted repeats 5 nucleotides long. Three-base-pair repeats flank ART-CH in phage clone 14, suggesting that duplication of the host sequence in the site of retrotransposon integration has occurred.

Although ART-CH significantly differs from other proviruses, strong homology was found in a variety of putative functional regions. For example, the polypurine tracts of ART-CH, ALV, and EAV are almost identical, as are the $tRNA^{Trp}$-binding sites. The homology between ART-CH and EAV extends 55 bp from the polypurine tract into the U3 region. The next conserved domain is located in the region of initiation and termination of transcription. In all three cases (ART-CH, ALV, and EAV), the TATA box is separated from the polyadenylation signal by a 16-bp sequence which is very similar in the ART-CH and EAV genomes. The first seven nucleotides of the R region are identical in all three proviruses; this is a strong indication that transcription of ART-CH is initiated at the same point as in ALV and EAV. We have not yet determined the position of the 3′ end of the ART-CH R region; it will be identified only after ART-CH RNA analysis.

There are no regions of pronounced similarity between ART-CH and ALV or EAV in the U5 region or in most of the U3 region of the LTR. Among the short sequences of similarity which were found in U5, an 11-bp sequence was identical to the 3′ half of the oligonucleotide used as the antisense primer in PCR. Homology with the sense primer is also incomplete (Fig. 5). The ART-CH LTR was apparently "caught" by these short regions, which formed duplexes with the primers only under nonstringent conditions of annealing (47°C).

The nontranslated region downstream of the 5′ ART-CH LTR contains a number of short sequences resembling different fragments of the corresponding parts of the EAV and ALV genomes. However, their order in ART-CH differs from that in EAV. Direct and inverted repeats are located in this region (Fig. 5). The 5′-nontranslated region is known to be involved in retroviral RNA packaging (2, 22, 32). We speculate that the conserved regions in this portion of the ART-CH genome may function in the same way.

Downstream of the 5′-nontranslated region, at the very end of the sequenced region, an open reading frame of the putative ART-CH *gag* gene begins. The predicted amino acid sequence encoded by this region has obvious similarity with the N-terminal parts of the amino acid sequences encoded by the ALV and EAV *gag* genes (Fig. 5).

The nontranslated region near the 3′ ART-CH LTR has sequences (from bp −155 to −87) similar to the comparable parts of the ALV provirus. The most remarkable portion of this region is located between nucleotides −90 and −25 of the ART-CH genome. It is almost identical to a sequence which is located downstream of the following oncogenes: (i) v-*src* in RSV strains Prague, Schmidt-Ruppin, B77, and 29 but not in the Bryan high-titer strain (13, 23, 30); (ii) v-*myc* in MH2 virus (38); (iii) v-*fps* in avian sarcoma virus Fujinami and PRCII (6, 31); and (iv) v-*crk* in avian sarcoma virus CT10 (25) (Fig. 6). In the case of RSV, this region was named F3 (30, 39). This sequence is repeated in U3 of the ART-CH LTR and is not found in the cognate cellular proto-oncogenes.

The region of homology between ART-CH and some recombinant acutely transforming avian sarcoma-leukemia viruses (ASLVs) is also partly homologous to direct repeat 1, the conserved sequence located in the 3′-nontranslated

FIG. 5. Comparison of ART-CH LTR and flanking regions from phage 14 (Fig. 4) with endogenous and exogenous avian retroviral LTRs. (A) Sequences of LTRs and their flanking regions of the ART-CH element from phage clone 14. Numbering starts from the first nucleotide of the U3 region of the LTR. I represents boundaries of the sequences which are indicated by designations at the ends of the dashed arrows. Alignments between similar regions of ART-CH and ALV, EAV, *Drosophila* retrotransposon 412, or CR1 are shown. The positions of characteristic sequences {polypurine tract; U3, U5, and R regions; $tRNA^{Trp}$-binding site, TATA box, nuclear factor 1 (NF-1)-binding site; polyadenylation [poly(A)] signal, and PCR primer-binding sites}, as well as the amino acid sequence encoded by the ART-CH 3′-end open reading frame, are indicated. Above the main ART-CH sequence obtained from phage 14, nucleotide residues which vary between ART-CHs in phages 14 and 5 are shown. Regions of homology and functional regions are not marked in the 3′ LTR, since no differences were found between LTRs in ART-CH from phage 14. (B) Schematic alignment of the ART-CH LTR with short flanking regions to those of EAV and RAV-0. The open parallelograms represent regions of 70% or more homology over 15 or more bases. ppt, polypurine tract; pbs, primer-binding site.

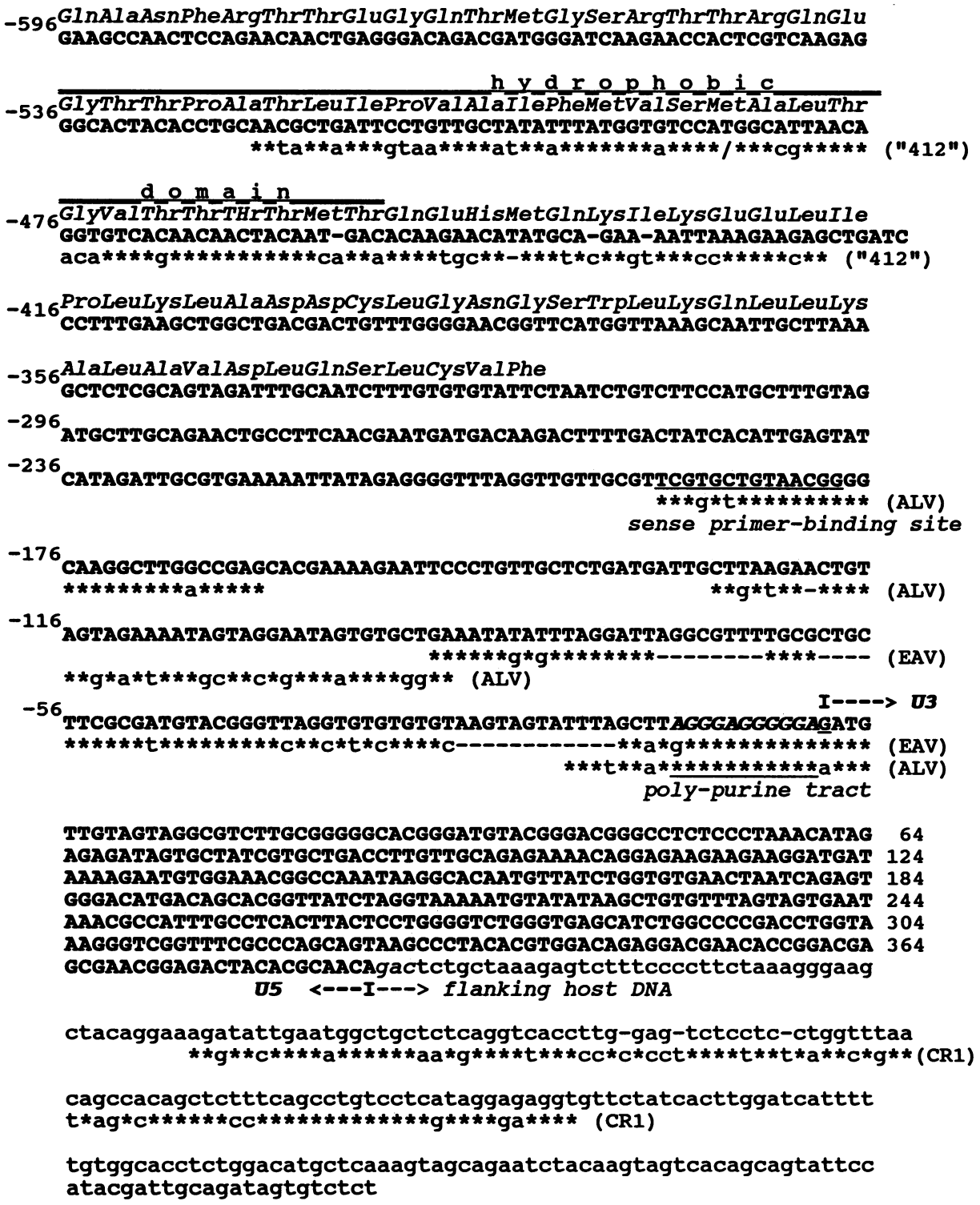


FIG. 5—*Continued.*

region of all known ALV genomes (Fig. 6). It is involved in viral RNA encapsidation, probably forming a part of the packaging signal (2, 32). One copy of direct repeat 1 is present in all of the studied ASLV genomes except that of RSV, where it is duplicated, surrounding the v-*src* oncogene. Thus, the four representatives of transforming recombinant ASLV mentioned above carry both the "classical" type of direct repeat 1 and its variant homologous to ART-CH. This observation suggests that ART-CH or closely related elements took part in the recombination events which led to the incorporation of oncogenes into ALV genomes.

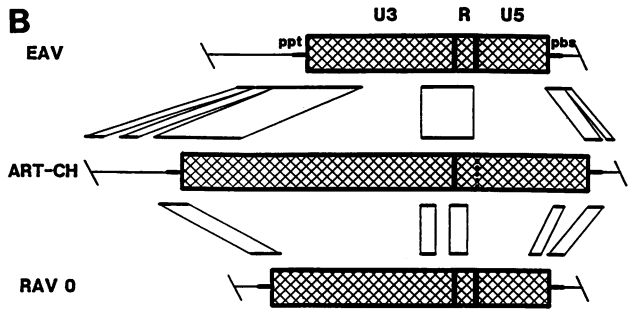


FIG. 5—*Continued.*

Inside the sequenced region upstream of the right LTR (from bp −320 to the beginning of the available sequence, bp −710), there is an open reading frame, a stretch encoding 25 hydrophobic amino acids followed by a highly hydrophilic region (Fig. 5). This sequence might encode the C terminus of an ART-CH transmembrane protein (perhaps *env*). No similarities were found between this putative coding region and genomes of other vertebrate viruses or retrotransposons. However, there is a region of clear homology with a fragment of *Drosophila* retrotransposon 412 (41) (Fig. 5).

The ART-CH family in the chicken genome. We used the ART-CH LTR as a probe for Southern blot hybridization of DNA from different species of birds (Fig. 7). A large number of bands were detected in chicken DNA digested by *Eco*RI or *Bam*HI after hybridization under stringent conditions. No hybridization was detected with DNAs of other birds tested. The sets of restriction fragments which hybridize with the ART-CH LTR probe vary slightly in DNAs of different breeds of chickens (Fig. 7). However, these variations are not comparable with differences in the sets of *ev*-specific bands that can be detected by hybridization of these DNAs with an ALV-specific probe (3, 17, 19). Low-stringency hybridization allows ART-CH-related sequences in quail and turkey DNA to be detected (data not shown). This is not surprising, since it is known that the DNAs of both species contain ALV-related sequences which should hybridize with the ART-CH probe at low stringency.

The cloned ART-CH element contains a single *Eco*RI site. If this is true for the rest of the family, then the number of bands on Southern blots should be twice the copy number of ART-CH loci in the chicken genome. Thus, there are at least 25 to 50 copies of ART-CH in chicken DNA; different individual birds contain similar sets of ART-CH loci.

ART-CH seems to be a recently acquired component of the chicken genome, since we failed to find sequences able to hybridize under stringent conditions with the ART-CH LTR probe in DNAs of other bird species, even those closely related to chickens (see legend to Fig. 7). The low level of divergence among independently cloned representatives of the ART-CH family shown by direct sequencing is in agreement with this proposal: LTRs belonging to the same copy of

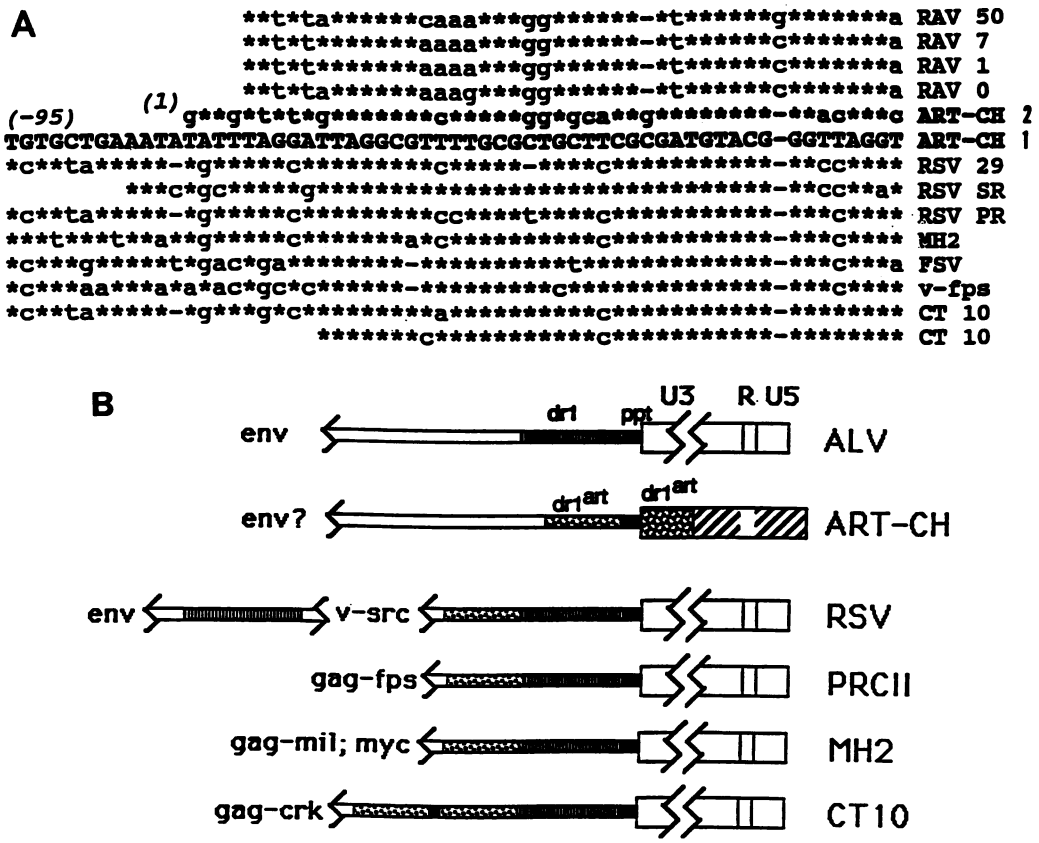


FIG. 6. ART-CH direct repeat 1 in other avian viruses. (A) Alignment of the direct repeat 1-like region of ART-CH with the sequences located closely downstream of four oncogenes in four different avian acutely transforming viruses (below the ART-CH sequence) and direct repeat 1 regions of ALVs (above the ART-CH sequence). FSV, Fujinami sarcoma virus. (B) Schematic structure of the 3′ regions of ART-CH and other chicken retroviruses (distributions and positions of direct repeat 1 [dr1] among different proviruses). ppt, polypurine tract.

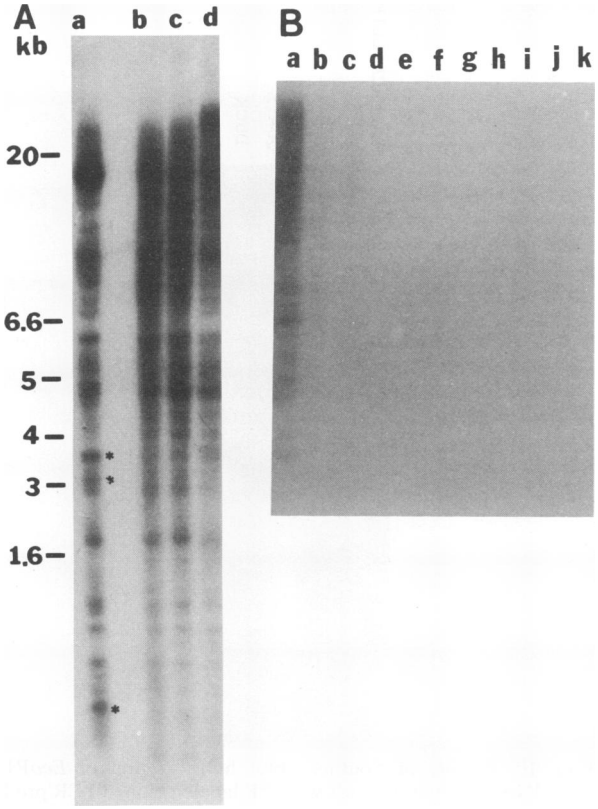


FIG. 7. Detection of ART-CH-specific restriction fragments in DNAs digested by *Eco*RI on Southern blots. Hybridization was performed with a ^{32}P-labeled LTR X probe (cloned 570-bp PCR product). Washing was done under stringent conditions (0.2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate], 68°C). (A) DNAs from different chicken embryos. Lanes: a, White Leghorn; b to d, Brown Leghorns. Some of the polymorphic fragments are marked with asterisks. (B) DNAs from chicken embryos (lane a) and the following birds: quail (lane b), turkey (lane c), duck (lane d), bullfinch (lane e), shrike (lane f), flycatcher (lane g), Bohemian waxwing (lane h), nutcracker (lane i), thrush (lane j), and warbler (lane k).

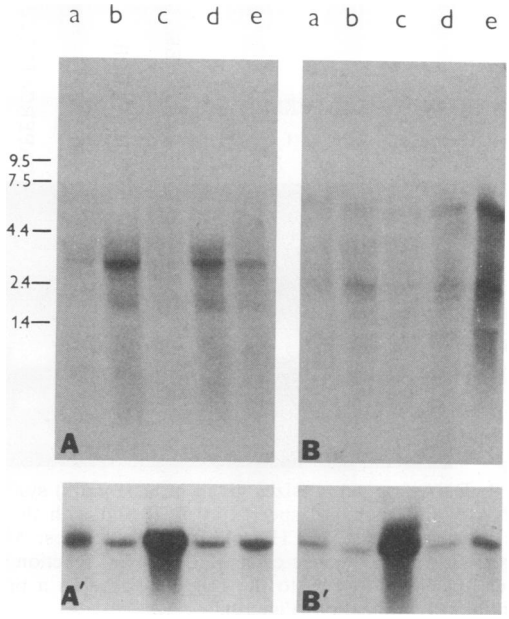


FIG. 8. Endogenous provirus-specific transcripts in RNA from different organs of an 18-day-old White Leghorn chicken embryo. Shown are results of Northern blot hybridization of total RNA from the following organs: muscles (lanes a), liver (lanes b), brain (lanes c), heart (lanes d), and lungs (lanes e). Hybridization was performed with a ^{32}P-labeled ART-CH probe (cloned 570-bp PCR product [LTR X]) (A) or ALV probe (R-COS plasmid) (B). (A′ and B′) Results of rehybridization of blots A and B with a tubulin probe (as expected, brain RNA contains more tubulin transcripts than other RNAs). Sizes are shown in kilobases at the left.

ART-CH are identical, while LTRs derived from independent clones differ in only two nucleotides (see legend to Fig. 5).

ART-CH transcription. We looked for ART-CH-specific RNA transcripts in different organs of 18-day-old chicken embryos. Results of Northern blot hybridization are shown in Fig. 8. For comparison, a parallel Northern blot was hybridized with an ALV probe which detects RNA transcripts of endogenous proviruses of the *ev* class. This probe hybridizes with two major transcripts of about 7.5 and 3 kb. These transcripts correspond to full-size and spliced *env*-specific RNAs of the *ev-3* locus present in the genomic DNA of the embryo. At this stage of development, *ev* loci are transcribed predominantly in lungs (unpublished observation).

The ART-CH probe detects RNA transcripts in all organs tested except the brain. ART-CH RNA is more abundant in the liver and heart than in skeletal muscle and lungs. A major transcript of 3 kb is detected. This transcript corresponds in size to the expected full-length RNA of the ART-CH element. There are also traces of smaller ART-CH transcripts which might either be transcribed from defective elements or arise as the result of splicing of full-size RNA.

Are there ART elements in other birds? The two primers used for the isolation of ART-CH seem to represent very conservative regions of avian retroviruses. We used them in PCRs with DNAs of different bird species and found that in some of the reactions (Fig. 9), products were synthesized. These products were isolated, cloned, and hybridized with each other and with DNAs of different birds.

There was no cross-hybridization between different products under stringent conditions, except for the pair of products that was synthesized in a PCR using capercaillie DNA (data not shown).

In Southern blot hybridization, each of the PCR products detected multiple bands in the particular DNA which was used for its synthesis but not in DNA from other birds. The exception was the capercaillie-specific product, which detected some sequences in duck DNA (Fig. 10). Hence, the pair of primers used for ART-CH isolation amplifies species-specific repetitive sequences from genomes of several birds which might represent fragments of other retroviruslike elements.

DISCUSSION

We used the PCR to isolate LTRs of different ALVs from DNA of ALV-infected chicken embryonic fibroblasts. Unexpectedly, an additional 570-bp product was synthesized. This band was seen even in control reactions with DNA of noninfected fibroblasts of ev^- chickens selected for the absence of RAV-0-related endogenous proviruses in their

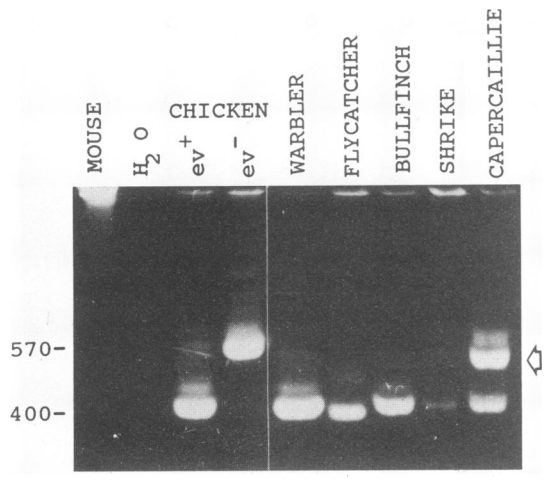


FIG. 9. DNA fragments (sizes given in base pairs) synthesized with DNA of different bird species by PCR and with the pair of primers used for the ART-CH isolation. Other lanes: MOUSE, DNA from BALB/3T3 mouse cells; H_2O, control reaction with no template. The arrow points to the fragment used as a probe for Southern blot hybridization (Fig. 10).

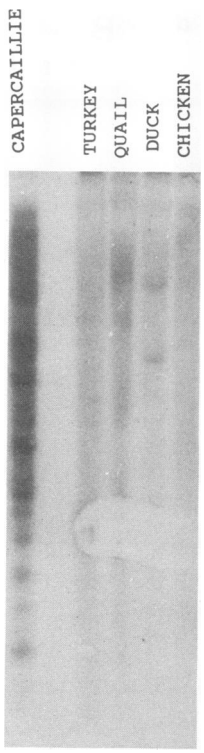


FIG. 10. Results of Southern blot hybridization of *Eco*RI-digested DNAs of different birds with ^{32}P-labeled cloned PCR product synthesized with capercaillie DNA. Conditions of washing: 1× SSC, 60°C.

genomes. This fragment was found to be a part of a novel retrovirus-like element which we named ART-CH.

ART-CH represents the third family of chicken retrovirus-like elements, in addition to the *ev* loci (3, 17, 19) and EAV proviruses (5, 12). Like other endogenous chicken retroviruses (15, 27), ART-CH was found to be a recently acquired component of the chicken genome, since it is not found in the genomes of other bird species, including such closely related species as turkeys and quails. LTRs of the element from phage 14 were found to be identical, also indicating a recent origin for this particular copy of ART-CH. Moreover, very few differences were found in independent clones of ART-CH LTRs, showing a very low level of divergence among members of the family. Despite a similar evolutionary age, ART-CH elements are much more abundant and stable components of the chicken genome than *ev* loci: chickens of different breeds contain about 25 to 50 copies of ART-CH, most of which are present in the DNA of every individual chicken.

The same pair of primers with which we isolated ART-CH was used for PCR with DNA of other bird species. Although the PCR-derived products have not yet been characterized, they all represent species-specific repetitive sequences and might also belong to unknown retroviruslike elements. This may be the fortuitous result of a choice of primers which encompass some extremely conserved retroviral sequences widely spread among avian retrotransposons.

We have sequenced the ART-CH LTRs and sequences surrounding them. Like all known retroviral LTRs (7, 21), the ART-CH LTR starts with the dinucleotide TG, ends with CA, and is framed by short inverted repeats. The U3 region contains 247 bp; thus, it is longer than the U3 regions of the majority of ALVs but shorter than the U3 region of avian myeloblastosis virus (285 bp). The 5′ quarter of the ART-CH U3 region, as well as its 3′ end downstream of the TATA box, is similar to that of EAV (Fig. 5). The rest of the ART-CH U3 region is a unique sequence which may contain enhancer elements. The close similarity of sequences representing the beginning of the R regions of ART-CH, ALV, and EAV LTRs enables us to predict the starting point of ART-CH transcription with a high probability. However, the 3′ end of the R region cannot yet be established. The length of the R+U5 portion of the ART-CH LTR is 141 bp, which is 40 bp longer than in all other studied avian retroviruses, including EAVs. The ART-CH, EAV, and ALV U5 regions contain a number of short stretches of homology. Since the function of the U5 part of the LTR is currently unknown, the significance of these similarities cannot be estimated.

Two sequences which play important roles in reverse transcription are located close to the LTRs: a tRNA primer-binding site (point of initiation of reverse transcription) and a polypurine tract (the site where synthesis of the plus strand of proviral DNA is initiated). ART-CH has a primer-binding site for $tRNA^{Trp}$, as do ALVs and EAVs. The ART-CH polypurine tract is identical to the same region in EAVs and differs in only one nucleotide from that of ALVs. The identity of these sequences may reflect a similarity of the enzymes which carry out proviral DNA synthesis in avian retroviruses. If ART-CH encodes a reverse transcriptase, we could expect it to be similar to ALV polymerases. Surprisingly, however, the cloned ART-CH element did not show any significant hybridization with the ALV *pol* probe (data not shown).

The 5′-nontranslated region and probably the 3′-nontranslated region in ALV RNA contain regions involved in RNA packaging (2, 22, 32). Clear regions of similarity exist in 5′-nontranslated regions of ART-CH, ALVs, and EAVs, indicating indirectly that they could form pseudotypes. A particular conserved sequence involved in RNA packaging and named direct repeat 1 is located in the 3′-nontranslated region of ALVs. This 100-bp-long fragment, which is present

in one copy in ALV genomes, is duplicated around the v-*src* oncogene in RSV (30). In EAVs, direct repeat 1-related sequences are located inside the U3 region (5). ART-CH has two fragments resembling direct repeat 1 with both locations (Fig. 6); one is upstream of the right LTR, and the other is in the U3 region, forming an imperfect direct repeat.

At least some ART-CH loci are capable of being expressed, since ART-CH-specific RNA is detected in organs of chicken embryos. There is one major size class of ART-CH RNA. This size (about 3 kb) is equal to the length of the cloned elements. This RNA is much shorter than any other known chicken provirus and can hardly encode all of the proteins which are usually encoded by a retrovirus genome. It is not clear whether the ART-CH RNA is spliced.

Although the ART-CH sequence is incomplete, it is clear that it may encode some proteins; the open reading frame of a *gag*-related protein is located at the 5′ end of the element. An open reading frame found near the 3′ end possibly encodes a protein with a hydrophobic domain including 29 amino acids. No similarity between this part of the ART-CH and any sequenced vertebrate retroelement was found. Surprisingly, *Drosophila* retrotransposon 412 has a sequence similar to a short sequence of the 3′-coding part of ART-CH (Fig. 5). The significance of this observation is not yet clear.

One of the most intriguing observations concerning ART-CH is a striking similarity between the ART-CH variant of direct repeat 1 and the sequence of unknown origin which follows the oncogenes v-*src* in RSV (the F3 region) (39), v-*myc* in MH2 (38), v-*fps* in avian sarcoma virus PRCII (6, 31), and v-*crk* in avian sarcoma virus CT10 (25) (Fig. 6). Since it does not originate from either the proto-oncogene or the helper virus, we speculate that ART-CH or another element which contains this sequence was involved in the recombination events resulting in the formation of acutely transforming viruses. The fact that different recombinant viruses with different oncogenes have the same ART-CH-related fragment is an argument for the existence of a common genetic mechanism for incorporation of cellular genes in ALV genomes, a mechanism which probably involves recombination with endogenous chicken retrovirus-like elements and subsequent recombination with exogenous ALV. It is unlikely that ART-CH replication is similar to that of the replication-competent retroviruses; the short length and presence of one major transcript indicate a unique strategy of ART-CH expression and replication. ART-CH might be an ALV-associated element which is spread, like murine VL30 retrotransposons, by packaging its RNA into virions of helper viruses (1, 29). It was found that VL30 had taken part in the recombination events which finally resulted in the formation of the acutely transforming Kirsten and Harvey murine sarcoma viruses (14). Endogenous elements in these processes may facilitate oncogene transduction, serving as donors of packaging signals or other sequences involved in homologous recombination with exogenous retroviruses. Complete sequencing of ART-CH and analysis of its expression and interaction with ASLVs will help to test these ideas.

ACKNOWLEDGMENTS

We thank Maxine Linial for helpful comments on the manuscript, Carolyn Zelnick and Kevin Noonan for critical reading and help in manuscript preparation, and Alexander Kazarov for help in some of the experiments.

REFERENCES

1. **Adams, S. E., P. D. Rathjen, C. A. Stanway, S. M. Fulton, M. H. Malim, W. Wilson, J. Ogden, L. King, S. M. Kingsman, and A. J. Kingsman.** 1988. Complete nucleotide sequence of a mouse VL30 retro-element. Mol. Cell. Biol. **8**:2989–2998.
2. **Aronoff, R., and M. Linial.** 1991. Specificity of retroviral RNA packaging. J. Virol. **65**:71–80.
3. **Astrin, S. M.** 1978. Endogenous viral genes of White Leghorn chickens: common site of residence and sites associated with specific phenotypes of viral gene expression. Proc. Natl. Acad. Sci. USA **75**:5941–5945.
4. **Bizub, D., R. A. Katz, and A. M. Skalka.** 1984. Nucleotide sequence of noncoding regions in Rous sarcoma virus-2: comparisons delineate conserved regions important in replication and oncogenesis. J. Virol. **49**:557–565.
5. **Boyce-Jacino, M. T., R. Resnick, and A. J. Faras.** 1989. Structural and functional characterization of the unusually short long terminal repeats and their adjacent regions of a novel endogenous avian retrovirus. Virology **173**:157–166.
6. **Carlberg, K., M. E. Chamberlin, and K. Beemon.** 1984. The avian sarcoma virus PRCII lacks 1020 nucleotides of the *fps* transforming gene. Virology **135**:157–167.
7. **Chen, H. R., and W. C. Barker.** 1984. Nucleotide sequence of the retroviral long terminal repeats and their adjacent regions. Nucleic Acids Res. **12**:1767–1778.
8. **Chernov, M. V., A. T. Tikhonenko, T. E. Zaitsevskaya, R. A. Denisova, I. B. Obukh, and A. V. Gudkov.** 1987. The genome of chickens free of endogenous proviruses contains sequences distantly related to Rous sarcoma and RAV-related viruses. Mol. Genet. Mikrobiol. Virusol. **3**:6–11.
9. **Coffin, J.** 1985. Endogenous retroviruses, p. 261–368. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
10. **Coffin, J.** 1985. Structure of the retrovirus genome, p. 17–74. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
11. **Dunwiddie, C. T., and A. J. Faras.** 1985. Presence of retrovirus reverse transcriptase-related gene sequences in avian cells lacking endogenous avian leukosis viruses. Proc. Natl. Acad. Sci. USA **82**:5097–5101.
12. **Dunwiddie, C. T., R. Resnick, M. Boyce-Jacino, J. N. Alegre, and A. J. Faras.** 1986. Molecular cloning and characterization of *gag*-, *pol*-, and *env*-related gene sequences in the ev^- chicken. J. Virol. **59**:669–675.
13. **Dutta, A., L.-H. Wang, T. Hanafusa, and H. Hanafusa.** 1985. Partial nucleotide sequence of Rous sarcoma virus-29 provides evidence that the original Rous sarcoma virus was replication defective. J. Virol. **55**:728–735.
14. **Ellis, R. W., D. Defeo, T. Y. Shih, M. A. Gonda, H. A. Young, N. Tsuchida, D. R. Lowy, and E. M. Scolnick.** 1981. The p21 src genes of Harvey and Kirsten sarcoma viruses originate from divergent members of a family of normal vertebrate genes. Nature (London) **292**:506–511.
15. **Frisby, D. P., R. A. Weiss, M. Roussel, and D. Stehelin.** 1979. The distribution of endogenous chicken retrovirus sequences in the DNA of galliform birds does not coincide with avian phylogenetic relationships. Cell **17**:623–634.
16. **Gudkov, A. V., E. Korec, M. V. Chernov, A. T. Tikhonenko, I. B. Obukh, and I. Hlozanek.** 1986. Genetic structure of the endogenous proviruses and expression of the gag gene in Brown Leghorn chickens. Folia Biol. (Prague) **32**:65–77.
17. **Gudkov, A. V., I. B. Obukh, S. M. Serov, and B. S. Naroditsky.** 1981. Variety of endogenous proviruses in the genomes of chickens of different breeds. J. Gen. Virol. **57**:85–94.
18. **Hauber, J., P. Nelbock-Hochstetter, and H. Feldmann.** 1985. Nucleotide sequence and characteristics of Ty element from yeast. Nucleic Acids Res. **13**:2745–2758.
19. **Hughes, S. H., K. Toyoshima, J. M. Bishop, and H. E. Varmus.** 1981. Organization of the endogenous proviruses of chickens: implication for origin and expression. Virology **108**:189–207.
20. **Humphries, E. H., M. L. Danhof, and I. Hlozanek.** 1984. Characterization of endogenous viral loci in five lines of White Leghorn chickens. Virology **135**:125–138.
21. **Ju, G., and B. R. Cullen.** 1986. The role of avian retroviral LTRs

in the regulation of gene expression and viral replication. Adv. Virus Res. **30**:179–223.
22. **Katz, R. A., R. W. Terry, and A. M. Skalka.** 1986. A conserved *cis*-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. J. Virol. **59**:163–167.
23. **Lerner, T. L., and H. Hanafusa.** 1984. DNA sequence of the Bryan high-titer strain of Rous sarcoma virus: extent of *env* deletion and possible genealogical relationship with other viral strains. J. Virol. **49**:549–556.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Mayer, B. J., M. Hamaguchi, and H. Hanafusa.** 1988. A novel viral oncogene with structural similarity to phospholipase C. Nature (London) **332**:272–275.
26. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5**:1630–1638.
27. **Resnick, R. M., M. T. Boyce-Jacino, Q. Fu, and A. J. Faras.** 1990. Rhylogenetic distribution of the novel avian endogenous provirus family EAV-0. J. Virol. **64**:4640–4653.
28. **Rovigatti, U. G., and S. M. Astrin.** 1983. Avian endogenous viral genes. Curr. Top. Microbiol. Immunol. **103**:1–21.
29. **Scolnick, E. M., W. C. Vass, R. S. Howk, and P. H. Duesberg.** 1979. Defective retrovirus-like 30S RNA species of rat and mouse cells are infectious if packaged by type C helper virus. J. Virol. **29**:964–972.
30. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
31. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**:787–795.
32. **Sorge, J., W. Ricci, and S. H. Hughes.** 1983. *cis*-acting RNA packaging locus in the 115-nucleotide direct repeat of Rous sarcoma virus. J. Virol. **48**:667–675.
33. **Soriano, P., T. Gridley, and R. Jaenisch.** 1987. Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov34 locus leads to early embryonic death. Genes Dev. **1**:366–375.
34. **Spence, S. E., D. J. Gilbert, D. A. Swing, N. G. Copeland, and M. A. Jenkins.** 1989. Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic *Srev* lines of mice. Mol. Cell. Biol. **9**:177–184.
35. **Stoye, J. P., and J. M. Coffin.** 1987. The four classes of endogenous murine leukemia virus: structural relationships and potential for recombination. J. Virol. **61**:2659–2669.
36. **Stoye, J. P., S. Fenner, G. E. Greenoak, C. Moran, and J. M. Coffin.** 1988. Role of endogenous retroviruses as mutagens: the hairless mutation of mice. Cell **54**:383–391.
37. **Stumph, W. E., C. P. Hodgson, M. J. Tsai, and B. W. O'Malley.** 1984. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. Proc. Natl. Acad. Sci. USA **81**:6667–6671.
38. **Sutrave, P., H. W. Jansen, K. Bister, and U. R. Rapp.** 1984. 3′-terminal region of avian carcinoma virus MH2 shares sequence elements with avian sarcoma viruses Y73 and SR-A. J. Virol. **52**:703–705.
39. **Takeya, T., and H. Hanafusa.** 1983. Structure and sequence of the cellular gene homologous to the RSV src gene and the mechanism for generating the transforming virus. Cell **32**:881–890.
40. **Tereba, A.** 1983. Endogenous retrovirus loci in chickens and mice. Curr. Top. Microbiol. Immunol. **107**:29–50.
41. **Yuki, S., S. Inouye, S. Ishimaru, and K. Saigo.** 1986. Nucleotide sequence characterization of Drosophila retrotransposon, 412. Eur. J. Biochem. **158**:403–410.