MOLECULAR AND CELLULAR BIOLOGY, Aug. 1989, p. 3563–3566
0270-7306/89/083563-04$02.00/0

Vol. 9, No. 8

Evidence that Chicken CR1 Elements Represent a Novel Family of Retroposons

ROSANE SILVA† AND JOHN B. E. BURCH*

Institute for Cancer Research, Fox Chase Cancer Center, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111

Received 13 March 1989/Accepted 9 May 1989

We report the first precise delineation of a chicken CR1 element and show that it is flanked by a 6-base-pair target site duplication that occurred when this repetitive element transposed. The 3′ end of this CR1 element is defined by an 8-base-pair imperfect direct repeat, and we infer that this sequence represents the 3′ end of all intact CR1 elements. In contrast, the 5′ ends are not unique, and we argue that this variation existed at the time each element transposed. We also provide evidence that CR1 elements transposed into preferred target sites. CR1 elements therefore appear to represent a novel class of passive retroposons.

O'Malley and co-workers were the first to describe and characterize the family of chicken middle repetitive sequences, which they termed CR1, for chicken repeat 1 (14, 16). Approximately 7,000 to 20,000 copies of CR1 sequences are present in the haploid chicken genome (6, 16), and it has been a puzzle as to how these elements dispersed so efficiently throughout the genome, especially considering that processed pseudogenes are exceptionally rare in avian as compared with mammalian genomes (1, 3, 13, 19, 20). This study was undertaken to further our understanding of this process and was prompted by the realization that the chicken VTGIII vitellogenin gene that we cloned (11a) contained the ancestral preintegration site of a previously characterized CR1 element (17). A sequence comparison between the second-intron regions of the VTGIII gene and VTGIII pseudogene (ΨVGTIII) revealed that the CR1 element within ΨVTGIII is precisely 836 base pairs (bp) long (Fig. 1). The fact that the CR1 sequence is flanked by direct repeats of a 6-bp sequence (CATTTC; boxed in Fig. 1) that appears only once at the preintegration site indicates that this CR1 element transposed into its present location.

Previous attempts to define the 3′ ends of CR1 elements relied on sequence comparisons between different CR1 elements. These studies indicated that the individual elements have similar 3′ ends, but the positions of these ends could not be determined unambiguously (6, 10, 14, 15, 17). For example, it was unclear whether the 3′ ends terminate (i) where the various CR1 sequences abruptly deviate from high to modest homology or (ii) approximately 20 bp further downstream, where the sequences completely deviate from any apparent consensus (15).

For the CR1 sequence within ΨVTGIII, our data demonstrate that the 3′ end maps precisely to the end of what was previously shown by O'Malley and co-workers to be the highly conserved region. This end is characterized by a direct repeat of the octamer sequence NATTCTGT (arrows in Fig. 1). All but one of the published CR1 sequences have a similar octamer direct repeat at what we infer to be their 3′ ends. Indeed, a comparison of all of the published CR1 sequences shows a consensus of NATTCTRT (where R is a purine) for the octamer. There is a strong preference for the first octamer to begin with a C and for the second octamer to begin with a G (Fig. 2). For a number of CR1 elements, the tandem octamers are flanked by short sequences that can also be found duplicated immediately upstream of the presumptive 5′ ends of these elements (Fig. 3), as would be expected if these other CR1 sequences also transposed as such. For example, CR1OVc appears to be flanked by direct repeats of NTAAC (15), CR1OVa by direct repeats of GATCAT (15), and CR1APOVLDLII by direct repeats of CAATTC (6). The short direct repeats flanking CR1 elements appear to vary in size, as is true of other families of nonviral retroposons (20).

The fact that the last six nucleotides (TTCTGT) of the octamer repeat are present at the preintegration site in the VTGIII gene (open arrow in Fig. 1) indicates that a limited degree of sequence homology may have contributed to the precise choice of target site for this CR1 insertion. On the assumption that the 3′ ends of all CR1 elements are indeed defined by the tandem octamers, it follows that target site preference may be a general feature of CR1 transpositions since, as noted above, a 20-bp window of modest sequence homology is apparent just downstream of the octamer repeat for the various CR1 elements (10). Preferred target sites have also been noted for other transposable elements (5, 11, 20).

In contrast to the conserved 3′ ends, each CR1 element has a distinct truncated 5′ end. The fact that direct repeats are still apparent flanking a number of these elements (Fig. 3) argues that these variable 5′ ends were created before or during the independent integration events. It is highly unlikely that this variation could be attributable to deletions which occurred after transposition. This situation is qualitatively similar to what has been described for other families of repetitive elements such as the human LINE-1 (long interspersed repeated sequence) elements (12). Their variable sizes almost certainly derive from the incomplete reverse transcription of an mRNA species expressed in embryonal cells (20). This so-called master LINE-1 mRNA contains two open reading frames (ORFs), one of which appears to encode a reverse transcriptase that presumably facilitates the transposition of this class of elements (8). Although CR1 elements are significantly smaller in size, the CR1 element discussed here also contains an ORF which begins before the left breakpoint and is capable of coding for a protein larger than 24 kilodaltons (Fig. 1). However, a search of the data base failed to provide any clues as to the possible identity of the putative protein encoded by this ORF (6). Thus, trans-

* Corresponding author.
† On leave from Instituto de Biofisica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil.

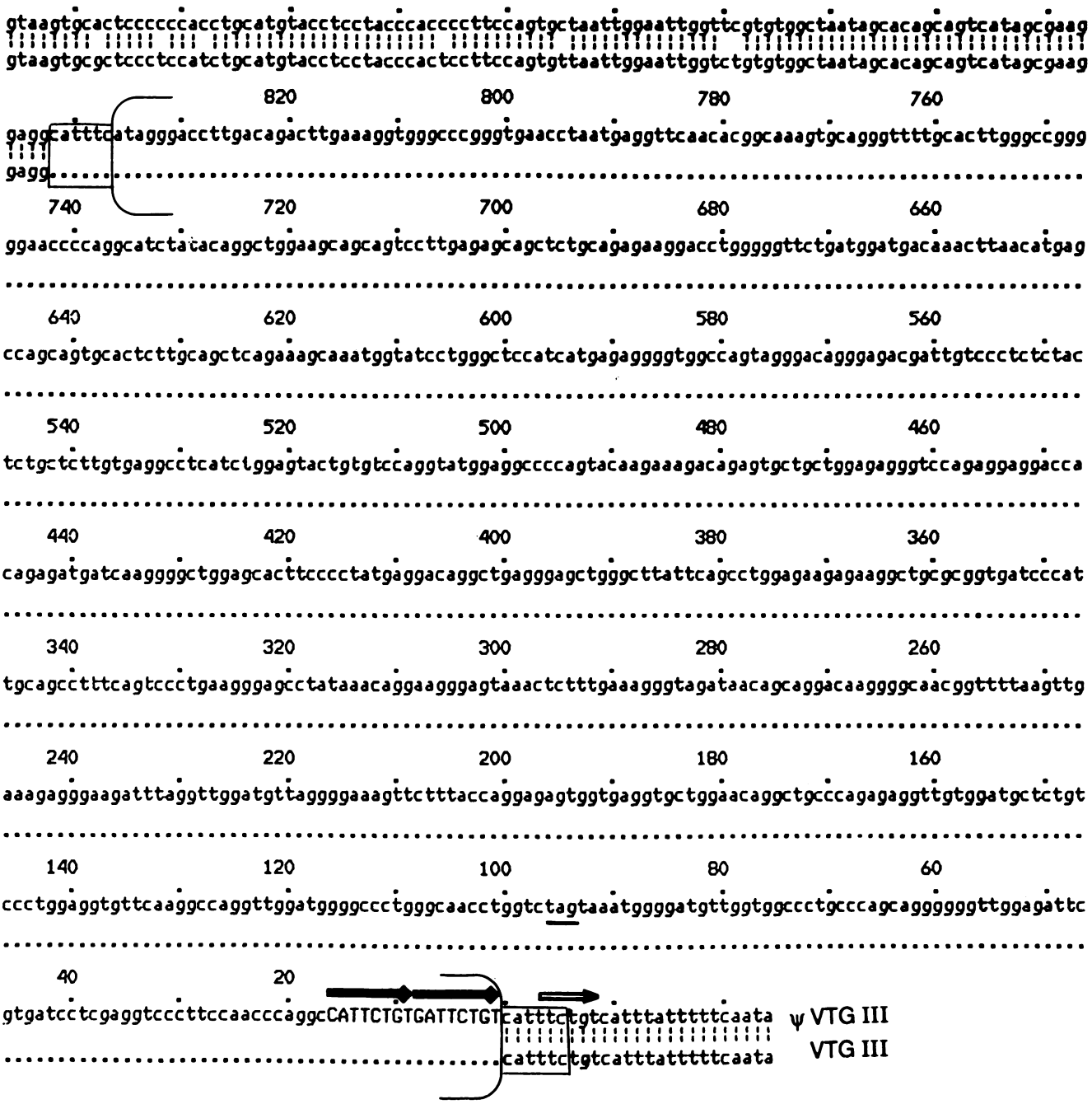

FIG. 1. Precise delineation of a CR1 sequence and identification of its ancestral preintegration site. The VTGIII second intron was sequenced on both strands and compared with the previously reported sequence for the remnant second intron of a VTGIII pseudogene (17). The CR1 sequence (bracketed) is numbered relative to the conserved 3′ end (see text). The ancestral preintegration site (CATTTC) that was duplicated to yield the direct repeats which now flank the CR1 element in the pseudogene intron is boxed. Solid arrows and capital letters mark the octamer sequence (NATTCTGT) that defines the 3′ end of the CR1 element. Note that this CR1 sequence contains an ORF that begins at the left end (position 836) and continues to the stop codon (TAG) underlined at position 96. The open arrow indicates a sequence that is present at the preintegration site and is identical to the last six nucleotides of the octamer sequence that defines the 3′ end of the CR1 element.

position of CR1 sequences may depend on a reverse transcriptase encoded elsewhere.

The oligo(A) tracts at the 3′ ends of transposed LINE-1 elements implicate the poly(A) tail of the master LINE-1 mRNA as the priming site for the reverse transcription of this family of repetitive elements (12). The use of poly(A) tracts for priming reverse transcriptase has been inferred for many classes of retroposons that derive from poly(A)-containing mRNA (20); even some classes of poly(A)-deficient transcripts are claimed to be reverse transcribed by this mechanism after first being aberrantly polyadenylated (4). The 3′ ends of the CR1 elements clearly do not conform to this paradigm, however.

How then can we account for the precisely defined 3′ ends

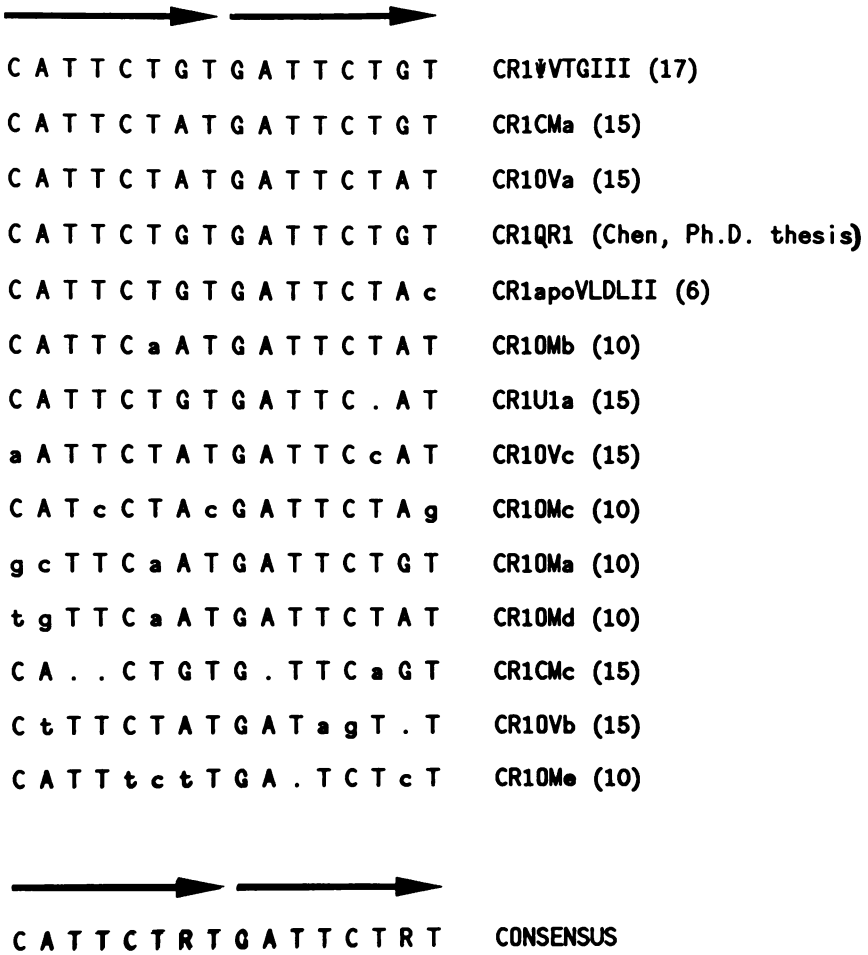

FIG. 2. Comparison of the 3′ tandem imperfect octamers for various CR1 elements. Sequences for the 3′ ends of CR1 elements (taken from the indicated references) were aligned relative to the octamer repeat that defines the 3′ end of the CR1 element from ΨVTGIII (see Fig. 1). Capital letters indicate a match and lowercase letters indicate a mismatch to the consensus sequence shown at the bottom. Point deletions are indicated by dots.

of CR1 elements? A comparison of CR1 sequences with mammalian U3 RNA pseudogenes may provide some useful insights. In an elegant series of experiments, Bernstein and colleagues showed that the simple addition of reverse transcriptase to full-length U3 RNA generated discrete cDNAs that were identical to the processed pseudogenes present in genomic DNA (2). Transcripts from a hypothetical master CR1 gene may also be exceptional in being able to fold in such a way as to allow precise priming to occur at the octamer direct tandem repeats which define the 3′ ends of these elements. Alternatively, sequences immediately downstream of the octamer repeat may serve as a precise primer-binding site for a heterologous RNA species such as occurs with retroviral replication (18). Obviously, the identification of such a primer-binding site requires that either the hypothetical master CR1 gene or its transcript be analyzed, since this information is not retained on the transposed elements.

In conclusion, we have presented evidence that CR1 elements represent a novel class of retroposons which share features of two distinct classes of retroposons. We speculate (Fig. 3) that CR1 retroposons may derive from the reverse transcription of a master CR1 transcript that is larger than the transposed elements at both the 5′ and 3′ ends. Expression of this transcript must occur early in development or in germ cells to account for the fact that the transpositions have been fixed in the germ line of chickens. Obviously, CR1 transpositions may also occur in somatic cells, as has recently been documented for human LINE-1 elements (7). Processed pseudogenes are exceptionally rare in avian species relative to mammalian species (1, 3, 13, 19, 20). The precise 3′ ends of CR1 elements suggest that a specific means for priming reverse transcriptase may have contributed to the fact that CR1 elements appear to be exceptions to this rule. The importance of this novel 3′ end is also implied by the fact that it is precisely conserved for a quail CR1 homolog (P.-J. Chen, Ph.D. thesis, University of Pennsylvania, Philadelphia, 1986). The fact that a nuclear factor binds immediately upstream of this conserved 3′ end (9) raises the intriguing possibility that such a factor may be involved in the transposition of CR1 elements. It is tempting to speculate that protein-protein interactions involving this factor might even be involved in directing the insertion of this conserved end into preferred target sites.

We thank Dorene Davis for excellent technical assistance and the Secretarial Services at the Institute for Cancer Research for help in preparing the manuscript. We also thank Susan Astrin, Kathleen Conklin, Richard Katz, John Taylor, and Philip Tsichlis for helpful comments.

This work was supported by Public Health Service grants 35535-04 (to J.B.), CA-06927, and RR-05539 from the National Institutes of

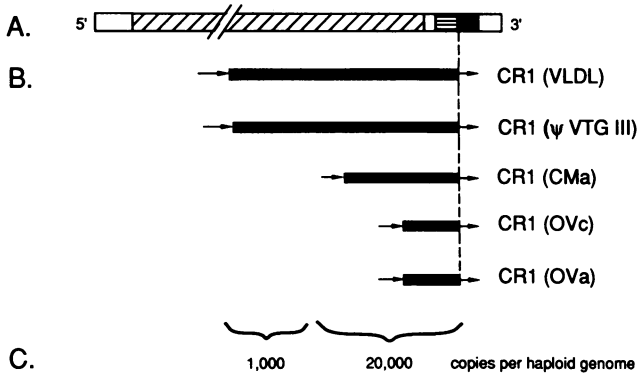

FIG. 3. Schematic representation of CR1 elements, similar to one used previously to depict human LINE-1 elements (12). (A) The hypothetical master CR1 transcript suggested to encode a protein expressed in embryonal or germ cells. Symbols: ▨, portion containing an ORF; ▤, 3′-terminal octamer direct repeat (NATTCTRT, where R is a purine); ■, predicted location of an efficient primer-binding site for reverse transcription. (B) Five representative CR1 elements aligned relative to the putative master CR1 transcript. The direct repeats that flank these elements are indicated (→). The relative sizes of these elements are drawn roughly to scale, with the VTGIII pseudogene (ΨVTGIII) CR1 sequence being precisely 836 bp (see Fig. 1). (C) Sequence complexities for two regions of CR1 sequences (6).

Health, a training grant from the Brazilian Government through Universidade Federal do Rio de Janeiro (to R.S.), and an appropriation from the Commonwealth of Pennsylvania.

LITERATURE CITED

1. **Alsip, G. R., and D. A. Konkel.** 1986. A processed chicken pseudogene (CPS1) related to the ras oncogene superfamily. Nucleic Acids Res. **14**:2123–2138.
2. **Bernstein, L. B., S. M. Mount, and A. M. Weiner.** 1983. Pseudogenes for human small nuclear FNA U3 appear to arise by integration of self-primed reverse transcripts of the RNA into new chromosomal sites. Cell **32**:461–472.
3. **Deininger, P. L., and G. R. Daniels.** 1986. The recent evolution of mammalian repetitive DNA elements. Trends Genet. **2**: 76–80.
4. **Eickbush, T. H., and B. Robins.** 1985. Bombyx mori 285 ribosomal genes contain insertion elements similar to the type I and II elements of Drosophila melanogaster. EMBO J. **4**: 2281–2285.
5. **Furano, A. V., C. C. Somerville, P. N. Tsichlis, and E. D'Ambrosio.** 1986. Target sites for the transposition of rat long interspersed repeated DNA elements (LINEs) are not random. Nucleic Acids Res. **14**:3717–3727.
6. **Haché, R. J. G., and R. G. Deeley.** 1988. Organization, sequence and nuclease hypersensitivity of repetitive elements flanking the chicken spoVLDLII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene. Nucleic Acids Res. **16**:97–113.
7. **Morse, B., P. G. Rotherg, V. J. South, J. M. Spandorfer, and S. M. Astrin.** 1988. Insertional mutagenesis of the *myc* locus by a LINE-1 sequence in a human breast carcinoma. Nature (London) **333**:87–90.
8. **Sakaki, Y., M. Hattori, A. Fujita, K. Yoshioka, S. Kuhara, and O. Takenaka.** 1986. The LINE-1 family of primates may encode a reverse transcriptase-like protein. Cold Spring Harbor Symp. Quant. Biol. **51**:465–469.
9. **Sanzo, M., B. Stevens, M. J. Tsai, and B. W. O'Malley.** 1984. Isolation of a protein fraction that binds preferentially to chicken middle repetitive DNA. Biochemistry **23**:6491–6498.
10. **Scott, M. J., M. J. Tsai, and B. W. O'Malley.** 1987. Deoxyribonuclease I sensitivity of the ovomucoid-ovoinhibitor gene complex in oviduct nuclei and relative location of CR1 repetitive sequences. Biochemistry **26**:6831–6840.
11. **Shih, C.-C., J. P. Stoye, and J. M. Coffin.** 1988. Highly preferred targets for retrovirus integration. Cell **53**:531–537.

11a. **Silva, R., A. H. Fischer, and J. B. E. Burch.** 1989. The major and minor chicken vitellogenin genes are each adjacent to partially deleted pseudogene copies of the other. Mol. Cell. Biol. **9**: 3557–3562.

12. **Skowronski, J., and M. F. Singer.** 1986. The abundant LINE-1 family of repeated DNA sequences in mammals: genes and pseudogenes. Cold Spring Harbor Symp. Quant. Biol. **51**: 457–464.
13. **Stein, J. P., R. P. Munjaal, L. Lagace, E. C. Lai, B. W. O'Malley, and A. R. Means.** 1983. Tissue-specific expression of a chicken calmodulin pseudogene lacking intervening sequences. Proc. Natl. Acad. Sci. USA **80**:6485–6489.
14. **Stumph, W. E., M. Baez, W. G. Beattie, M. J. Tsai, and B. W. O'Malley.** 1983. Characterization of deoxyribonucleic acid sequences at the 5′ and 3′ borders of the 100 kilobase pair ovalbumin gene domain. Biochemistry **22**:306–315.
15. **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.
16. **Stumph, W. E., P. Kristo, M. J. Tsai, and B. W. O'Malley.** 1981. A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats. Nucleic Acids Res. **9**:5383–5397.
17. **van het Schip, F., J. Samallo, F. Meijlink, M. Gruber, and G. A.** 1987. A new repetitive element of the CR1 family downstream of the chicken vitellogenin gene. Nucleic Acids Res. **15**:4193–4202.
18. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. Science **216**:812–820.
19. **Wagner, M.** 1986. A consideration of the origin of processed pseudogenes. Trends Genet. **2**:134–137.
20. **Weiner, A. M.** 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. Annu. Rev. Biochem. **55**:631–661.