MOLECULAR AND CELLULAR BIOLOGY, Aug. 1990, p. 4401–4405
0270-7306/90/084401-05$02.00/0

Vol. 10, No. 8

Human Proviral mRNAs Down Regulated in Choriocarcinoma Encode a Zinc Finger Protein Related to Krüppel

NOBUYUKI KATO,[1] KUNITADA SHIMOTOHNO,[1] DONALD VANLEEUWEN,[2] AND MAURICE COHEN[2]*

Virology Division, National Cancer Center Research Institute, Tsukiji, Chuo-ku, Tokyo, Japan,[1] and Department of Molecular Biology, D-93D, AP20/5, Abbott Laboratories, Abbott Park, Illinois 60064-2204[2]

Received 5 March 1990/Accepted 9 May 1990

RNA transcripts of the HERV-R (ERV3) human provirus that are abundant in placenta but absent in choriocarcinoma contain nonproviral genomic sequences at their 3′ ends. We report here the isolation of cDNA clones of these genomic sequences. The transcripts encode a Krüppel-related zinc finger protein consisting of a unique leader region and more than 12 28-amino-acid finger motifs.

We previously characterized a single-copy human provirus, HERV-R (ERV3), that is located on chromosome 7 and abundantly expressed in the placental chorionic villi throughout gestation. Although HERV-R transcripts are expressed in most other normal and malignant human tissues, they are absent from the six choriocarcinoma cell lines tested so far (12; M. Cohen, unpublished data). Gestational choriocarcinoma is a disease of the mother resulting from malignant transformation of the highly invasive placental trophoblast. Although extraembryonic syncytiotrophoblasts are normally shed into the maternal blood and accumulate in various organs (8), little is known regarding the timing or stage of differentiation during which transforming events occur.

There are three *env*-containing HERV-R mRNAs of 9, 7.3, and 3.5 kilobases (kb). Although the 3.5-kb mRNA contains only proviral sequences (4, 13), the 9- and 7.3-kb mRNAs are unusual, extending through the 3′ long terminal repeat to a splice donor site downstream from the provirus. These mRNAs contain an additional 5.5 and 3.8 kb, respectively, of human sequences (13).

To characterize the nonproviral portion of these mRNAs, we isolated a cDNA clone, Pr119, from a randomly primed human placental cDNA library (10) by using a hybridization probe, P3, from the genomic flanking region adjacent to the 3′ end of the provirus (Fig. 1a). Then, by using a fragment from Pr119, p4, we isolated an overlapping cDNA clone, Pr2-45, from the same library (Fig. 1a). The composite cDNA sequence (Fig. 1b) reveals a 1,281-base-pair open reading frame (ORF) beginning with two consensus Met initiator codons (16).

The ORF consists of two regions, an 84-amino-acid domain at the NH_2 terminus followed by a domain containing 12 28-amino-acid direct repeats of a zinc finger motif of the C_2H_2 class (7) (Fig. 2). The encoded protein would have a molecular weight of 50,056 and a calculated pI of 10.2. The motifs contain the consensus His-Cys link sequence connecting adjacent finger loops, which is characteristic of *Drosophila* Krüppel (*Kr*) (Fig. 2). *Kr* encodes a zinc finger protein which is structurally homologous to *Xenopus* transcription factor IIIA (17, 22). In *Drosophila melanogaster*, expression of the *Kr* gene is an essential part of normal embryonic development. The *Kr* gene product binds DNA in a sequence-specific manner and is thought to function as a transcription factor (21, 24). We named the gene *H-plk* for human provirus-linked Krüppel.

A search of the GenBank data base revealed that *H-plk* is very similar to a previously identified human Krüppel-related cDNA clone, HPF9 (M27879) (1). HPF9, which lacks upstream proviral sequences, begins near the middle of the *H-plk* 5′ noncoding exon. The two sequences are nearly identical throughout the remainder of this exon and the unique NH_2 terminus and first seven zinc finger motifs of the *H-plk* ORF. However, from that point the two sequences, while closely related, are clearly not identical. As was previously shown, the large HERV-R mRNAs differ only within their nonproviral sequences as a result of alternative processing in the region downstream from the provirus (13). Thus, *H-plk* and HPF9 may represent different, alternatively spliced mRNAs.

In comparisons with characterized genetic loci, *H-plk* is most closely related to the mouse Krüppel-related gene, *mKr-2*, which is expressed in neurons of embryonic and adult mice (3) (Fig. 2).

To determine the genomic organization of the HERV-R/*H-plk* locus, we probed Southern blots (23) of human, chimpanzee, and mouse genomic DNAs with *H-plk* *Eco*RI fragment p4 (Fig. 1a). This fragment contains the central portion of the *H-plk* zinc finger repeats, which extends from the middle of motif 4 to the middle of motif 8 (Fig. 2). When hybridized to DNAs of our *H-plk*-containing lambda clones, the fragment detected all of the *Eco*RI fragments which make up the ORF (data not shown). With this probe, more than 25 distinct bands were observed in human and chimpanzee DNAs. Longer autoradiographic exposure also revealed hybridization in mouse DNA (Fig. 3a). These results suggested that the zinc finger motifs of *H-plk* are part of a conserved, multigene family. From hybridization studies with a degenerate oligonucleotide probe for the consensus *Kr* His-Cys link region, Bellefroid et al. (1) estimated that human DNA actually contains about 300 *Kr*-related sequences.

In Northern (RNA) blots of placental RNA, the *H-plk* zinc finger probe, p4 (Fig. 1a), detected 9- and 7.3-kb mRNAs characteristic of HERV-R transcription (Fig. 3b). As expected for HERV-R-initiated mRNAs (12), this probe detected neither transcript in RNAs isolated from choriocarcinoma cells (Fig. 3b). The 4.2-kb transcript represents an *H-plk*-related mRNA, since its signal was significantly decreased by high-stringency washing, which is in contrast to the effect on the 9- and 7.3-kb signals (data not shown).

* Corresponding author.

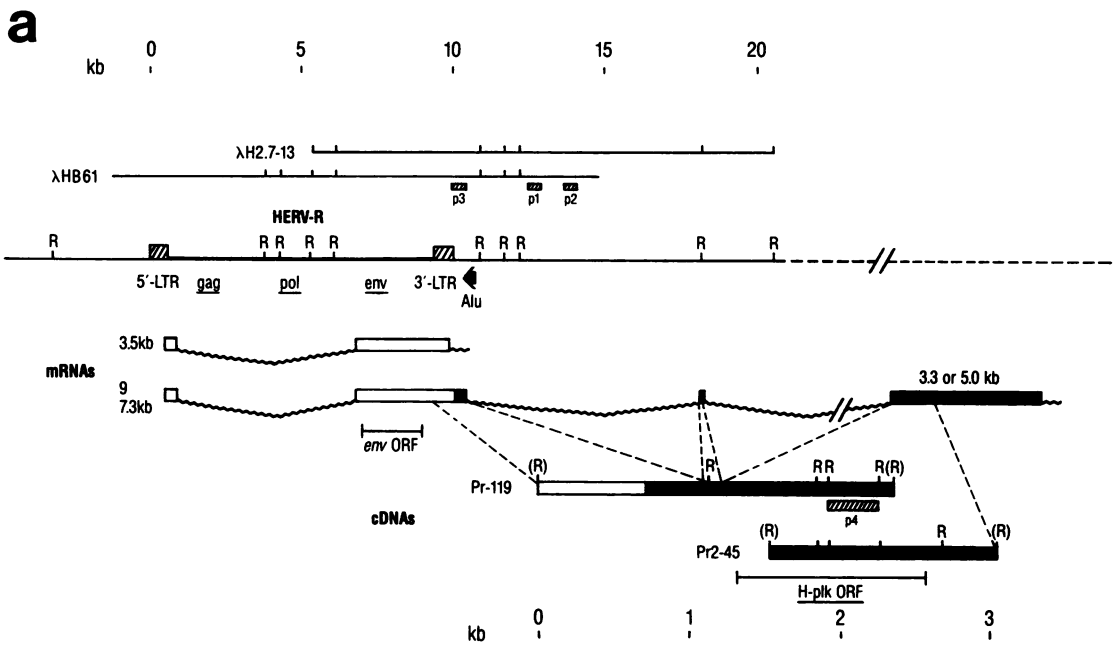

FIG. 1. (a) Isolation of *H-plk* cDNA and genomic clones (M33990). The following are indicated: HERV-R proviral regions (▭) and nonproviral regions of mRNAs and cDNA clones (■) (map is expanded for cDNA clones), hybridization probes (▨), and intervening sequences not present in mRNAs (∿∿). Genomic locations of cDNA sequences downstream from 5′ noncoding exon were not determined. R, *Eco*RI sites; (R), *Eco*RI sites generated as a result of cloning; ◀, *Alu* repeated element oriented in direction shown; LTR, long terminal repeat. (b) Composite nucleotide sequence of Pr-119 and Pr2-45 cDNA clones. Clones are identical in their approximately 0.82-kb overlap. Sequence begins in HERV-R *env* p15E (4). Domains: 3′ long terminal repeat, nucleotides 122 to 712; *Alu* repeat, nucleotides 929 to 1089; 5′ noncoding exon, nucleotides 1090 to 1222; *H-plk* ORF, nucleotides 1285 to 2565; Pr-119, nucleotides 1 to 2277. Although the sequences at the right terminus of Pr2-45 were not determined, the clone contains nucleotides 1462 to approximately 2949. ◆, Splice site.

From a HeLa cell partial *Eco*RI genomic library, we isolated genomic clone H2.7-13, which contains sequences downstream from the original HERV-R lambda clone (20), by using HERV-R-flanking fragments, p1 and p2, as probes (Fig. 1a). Sequence analysis revealed that a region approximately 8 kb downstream from the provirus is identical to a 133-base-pair sequence found in cDNA clone Pr-119 (Fig. 1b). This sequence, a 5′ noncoding exon, begins with the consensus splice acceptor sequence TTTTTGTGTTTTT CAG↓G and ends with the splice donor sequence G↓GTA

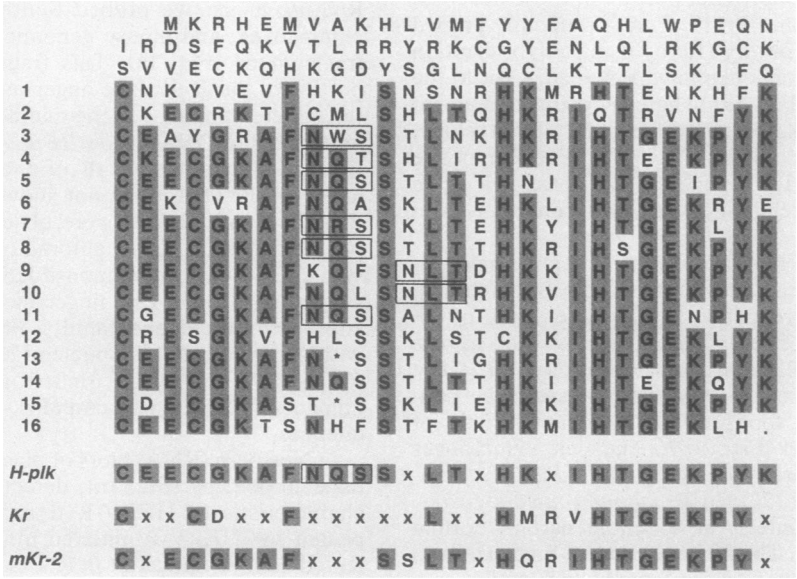

FIG. 2. Sequence of a protein with multiple zinc finger motifs, predicted by the *H-plk* ORF. Residues denoted by one-letter code (5) are shaded where they correspond to the consensus *H-plk* sequence. Consensus motifs were derived from sequences within the respective ORFs. Underlined Met residues denote consensus initiator codons (16). Boxes identify potential N-linked glycosylation sequences in ORF. Asterisks represent in-frame terminator codons. Consensus zinc finger motifs of *H-plk*, *Kr* (22), and *mKr-2* (3) are shown at the bottom.

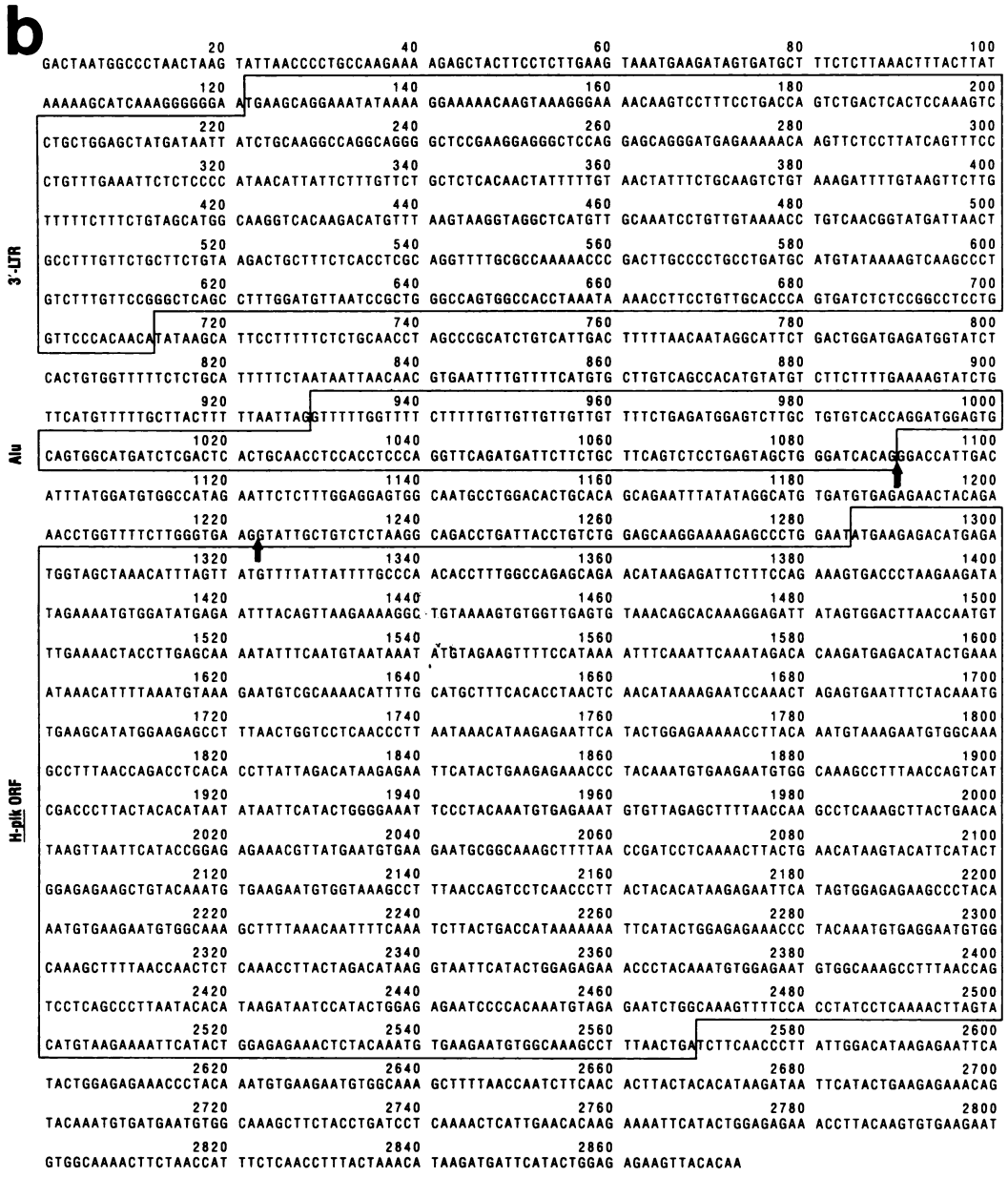

FIG. 1—*Continued.*

ACT. Furthermore, the exon is 61% identical to the 5′ noncoding region of *mKr-2* (3) (comparison not shown).

The splice donor site of the 9- and 7.3-kb HERV-R mRNAs was located precisely 377 base pairs downstream from the proviral 3′ long terminal repeat (Fig. 1b) (13). Sequencing of HERV-R in this region revealed that the splice donor site is located within the central region of an *Alu* repeated element oriented opposite to the direction of *H-plk* transcription (11). Alignment between the genomic *Alu* sequence and the cDNA clone indicates that the HERV-R 9- and 7.3-kb mRNAs contain the 3′ half of the *Alu* repeat (Fig. 1).

What evolutionary events may have given rise to a transcription unit containing both endogenous retrovirus and zinc finger protein sequences? Normal *H-plk* gene transcription in a primate ancestor of humans may have been disrupted by integration of the HERV-R provirus. The *H-plk*-coding exons were probably acquired in mRNAs initiating in the HERV-R 5′ long terminal repeat by subsequent evolution of splicing signals downstream from the provirus. The availability of a cryptic splice donor site in the *Alu* element may have been critical for the survival of the *H-plk* gene after HERV-R integration. Alternatively, the *Alu* element may have inserted after HERV-R integration, providing the means by which *H-plk* transcription could become associated with the provirus.

The importance of transcription regulators in transformation and malignancy is well established. c-*jun* and c-*fos* are examples of nuclear proto-oncogenes that contribute to transformation by positive activation (18). The role of other classes of nuclear proteins as transcription regulators in malignancy is still emerging. Expression of murine and human zinc finger genes *Evi-1* (19), *GLI* (14, 15), and *MOK-2* (6) is associated with transformation and neoplasia, and

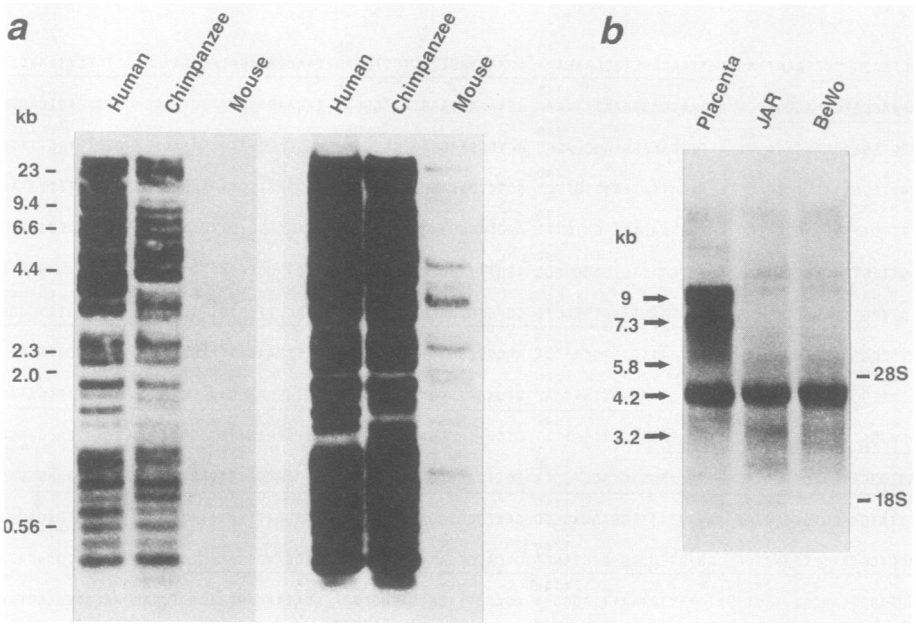

FIG. 3. (a) *H-plk* gene as a member of a multigene family. Cellular DNAs of normal human placenta, STLV-1-infected chimpanzee cell line chM 114 (25), and mouse NIH 3T3 cells were studied by Southern analysis (23). Autoradiographic exposure was fivefold longer on the right than on the left. (b) *H-plk* mRNAs expressed in human cells. RNAs isolated from human placenta and choriocarcinoma cell lines JAR and BeWo were analyzed by Northern hybridization with *H-plk* zinc finger probe, p4, (Fig. 1a).

recently a Krüppel-related zinc finger gene was located in a chromosomal region that is homozygously deleted in Wilms' tumor, suggesting that it may be the Wilms' tumor susceptibility locus (2, 9). We are studying the possible role of a highly expressed zinc finger gene, *H-plk*, during normal human development and the significance of its transcriptional down regulation in choriocarcinoma. Experiments to characterize human chromosome 7-choriocarcinoma microcell hybrids and to determine the cellular location and extent of *H-plk* protein expression are in progress.

We thank Y. Ikenoue for technical assistance and H. Tsujimoto, University of Tokyo, for the gift of DNA of chM 114 (25).

This work was supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Science and Culture, Japan.

LITERATURE CITED

1. **Bellefroid, E. J., P. J. Lecocq, A. Benhida, D. A. Poncelet, A. Belayew, and J. A. Martial.** 1989. The human genome contains hundreds of genes coding for finger proteins of the Krüppel type. DNA **8**:377–387.
2. **Call, K. M., T. Glaser, C. Y. Ito, A. J. Buckler, J. Pelletier, D. A. Haber, E. A. Rose, A. Kral, H. Yeger, W. H. Lewis, C. Jones, and D. E. Housman.** 1990. Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. Cell **60**:509–520.
3. **Chowdhury, K., G. Dressler, G. Breier, U. Deutsch, and P. Gruss.** 1988. The primary structure of the murine multifinger gene *mKr2* and its specific expression in developing and adult neurons. EMBO J. **7**:1345–1353.
4. **Cohen, M., M. Powers, C. O'Connell, and N. Kato.** 1985. Nucleotide sequence of the *env* gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology **147**:449–458.
5. **Dayhoff, M. O.** 1978. Survey of new data and computer methods of analysis, p. 1–8. *In* M. O. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
6. **Ernoult-Lange, M., M. Kress, and D. Hamer.** 1990. A gene that encodes a protein consisting solely of zinc finger domains is preferentially expressed in transformed mouse cells. Mol. Cell. Biol. **10**:418–421.
7. **Evans, R. M., and S. M. Hollenberg.** 1988. Zinc fingers: gilt by association. Cell **52**:1–3.
8. **Faulk, W. P., and J. A. McIntyre.** 1983. Immunological studies of human trophoblast: markers, subsets and functions. Immunol. Rev. **75**:139–175.
9. **Gessler, M., A. Poustka, W. Cavenee, R. L. Neve, S. H. Orkin, and G. A. P. Bruns.** 1990. Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. Nature (London) **343**:774–778.
10. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
11. **Kariya, Y., K. Kato, Y. Hayashizaki, S. Himeno, S. Tarui, and K. Matsubara.** 1987. Revision of consensus sequence of human *Alu* repeats—a review. Gene **53**:1–10.
12. **Kato, N., E. Larsson, and M. Cohen.** 1988. Absence of expression of a human endogenous retrovirus is correlated with choriocarcinoma. Int. J. Cancer **41**:380–385.
13. **Kato, N., S. Pfeifer-Ohlsson, M. Kato, E. Larsson, J. Rydnert, R. Ohlsson, and M. Cohen.** 1987. Tissue-specific expression of human provirus ERV3 mRNA in human placenta: two of the three ERV3 mRNAs contain human cellular sequences. J. Virol. **61**:2182–2191.
14. **Kinzler, K. W., S. H. Bigner, D. D. Bigner, J. M. Trent, M. L. Law, S. J. O'Brien, A. J. Wong, and B. Vogelstein.** 1987. Identification of an amplified, highly expressed gene in a human glioma. Science **236**:70–73.
15. **Kinzler, K. W., J. M. Ruppert, S. H. Bigner, and B. Vogelstein.** 1988. The GLI gene is a member of the Krüppel family of zinc finger proteins. Nature (London) **332**:371–374.
16. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
17. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
18. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins.

Science **245**:371–378.
19. **Morishita, K., D. S. Parker, M. L. Mucenski, N. A. Jenkins, N. G. Copeland, and J. N. Ihle.** 1988. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. Cell **54**:831–840.
20. **O'Connell, C., S. J. O'Brien, W. G. Nash, and M. Cohen.** 1984. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology **138**:225–235.
21. **Ollo, R., and T. Maniatis.** 1987. Drosophila Krüppel gene product produced in a baculovirus expression system is a nuclear phosphoprotein that binds to DNA. Proc. Natl. Acad. Sci. USA **84**:5700–5704.
22. **Rosenberg, U. B., C. Schröder, A. Preiss, A. Kienlin, S. Côté, I. Riede, and H. Jäckle.** 1986. Structural homology of the product of the Drosophila Krüppel gene with Xenopus transcription factor IIIA. Nature (London) **319**:336–339.
23. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
24. **Stanojevic, D., T. Hoey, and M. Levine.** 1989. Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krüppel in Drosophila. Nature (London) **341**:331–335.
25. **Tsujimoto, H., A. Komuro, K. Iijima, J. Miyamoto, K. Ishikawa, and M. Hayami.** 1985. Isolation of simian retroviruses closely related to human T-cell leukemia virus by establishment of lymphoid cell lines from various non-human primates. Int. J. Cancer **35**:377–384.