JOURNAL OF VIROLOGY, Sept. 1991, p. 4713–4719
0022-538X/91/094713-07$02.00/0

Vol. 65, No. 9

Molecular Cloning and Characterization of a Complete Chinese Hamster Provirus Related to Intracisternal A Particle Genomes

ANDREW J. DORNER,* FRANÇOISE BONNEVILLE, RONALD KRIZ, KERRY KELLEHER, KEVIN BEAN, AND RANDAL J. KAUFMAN

Genetics Institute, 87 Cambridge Park Drive, Cambridge, Massachusetts 02140

Received 19 March 1991/Accepted 4 June 1991

We report here the nucleotide sequence of a full-length Chinese hamster genomic proviral element, CHIAP34. CHIAP34 is 6,403 bp long with long terminal repeats of 311 bp at each end. The genetic organization of CHIAP34 was determined by comparison with intracisternal A particle (IAP) genetic elements from the mouse and Syrian hamster. Extensive homology at the nucleotide and deduced amino acid sequence levels was observed between CHIAP34 and the mouse and Syrian hamster IAP elements. CHIAP34 may represent a defective Chinese hamster IAP genetic element. The *gag* gene consists of 837 codons, of which 558 codons are in a single long open reading frame followed by several frameshifts. The *pol* gene begins with a −1 frameshift and consists of a long open reading frame of 753 codons followed by a short open reading frame of 103 codons. The putative *env* region contains multiple termination codons in all reading frames. CHIAP34 is representative of the predominant retroviral elements in the Chinese hamster ovary cell genome present at around 80 copies per haploid genome.

The genomes of most vertebrates are littered with remnants of past retroviral infections. In species of rodents, the diversity of integrated proviral elements is quite large. The endogenous retroviruses include type B, C, and A viral elements (18). Intracisternal type A particles (IAPs) are defective retroviruses which are encoded by a large family of structurally diverse proviral elements present in several hundred to a thousand copies per haploid rodent genome (6). IAPs assemble on and bud into the endoplasmic reticulum and are not released from the cell (6). Full-length IAP proviral genomes from both the mouse (MIA14) and the Syrian hamster (H18) have been cloned and sequenced (10, 12). Their genetic organization is similar to that of infectious retroviruses such as Moloney murine leukemia virus. The first long open reading frame (ORF) encodes the *gag* gene product. The *pol* ORF encoding reverse transcriptase is entered from the *gag* gene ORF by a −1 frameshift. In MIA14, the putative *gag* region is interrupted by one termination codon, while in H18, it is interrupted by several termination codons. Frameshifts also disrupt the *pol* coding regions of both MIA14 and H18. The *env* region is closed in all three reading frames by multiple termination codons. The IAP proviral elements are flanked at both ends by long terminal repeats (LTRs) which contain the appropriate regulatory elements present in functional LTRs.

IAPs are present in a variety of rodent cells including the mouse and Syrian hamster (9). While IAPs have not been observed in Chinese hamster cells, intracytoplasmic particles associated with kinetochores have been reported (5). Chinese hamster cells also can spontaneously produce an infectious endogenous type C retrovirus (8). Since Chinese hamster ovary (CHO) cells are now extensively used for production of human pharmaceuticals, we have begun studies of the structure and expression of retroviral elements present in the CHO cell genome.

A full-length provirus, CHIAP34, was molecularly cloned, and its complete nucleotide sequence was determined. CHIAP34 is 6,403 bp long with LTRs of 311 bp at each end. Imperfect long ORFs encode presumed *gag* and *pol* gene products with extensive similarity to those encoded by the mouse and Syrian hamster IAP genomes. CHIAP34 represents the first full-length endogenous proviral element isolated from the Chinese hamster genome and is being used to determine the genomic organization and cellular expression of the endogenous retroviruses of CHO cells. The extensive use of the CHO cell for mammalian somatic cell genetic studies (14) necessitates an understanding of the organization and fluidity of the endogenous retroviruses within these cells. With the primary role that recombinant protein production in CHO cells has assumed in biotechnology, information on the retrovirus content of these cells has become of increasing importance.

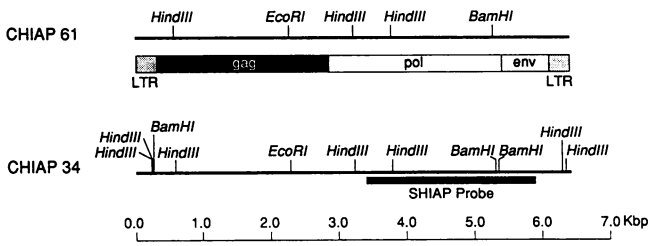

FIG. 1. Partial restriction maps of two Chinese hamster genomic retrovirus elements. CHIAP61 and CHIAP34 proviral elements are aligned based on shared restriction sites and cross hybridization of restriction fragments. The genetic organization of these elements is based on the nucleotide sequence of CHIAP34 as indicated. The position of the Syrian hamster IAP restriction fragment used as a probe is indicated as SHIAP probe.

* Corresponding author.

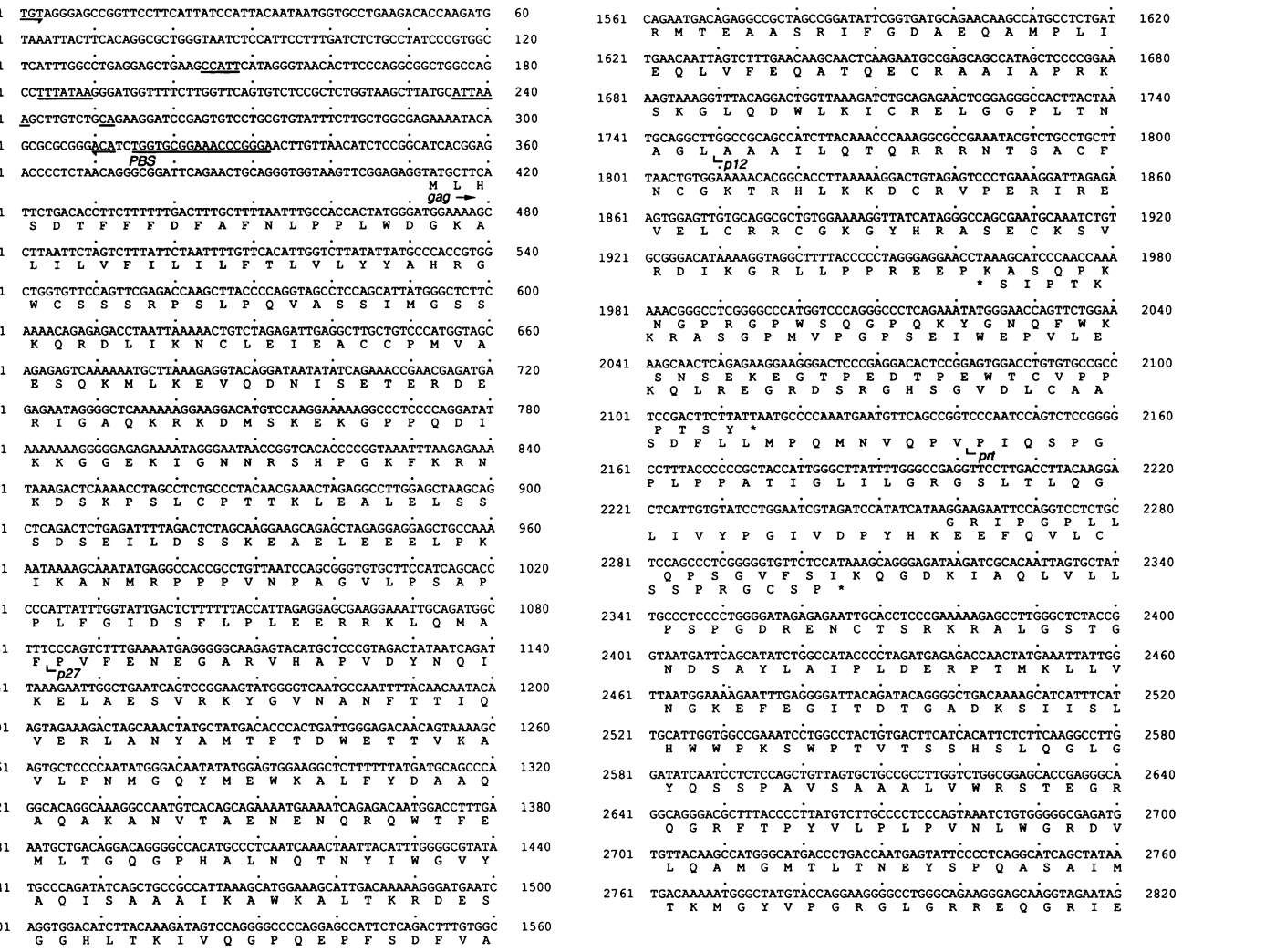

FIG. 2. Nucleotide sequence of CHIAP34, a full-length Chinese hamster genomic IAP proviral element. The DNA sequence of the coding strand is given. The deduced amino acid sequences of the *gag* and *pol* ORFs are placed under the nucleotide sequence. Stop codons are indicated by asterisks. Protein domains of the *gag* and *pol* gene products determined by amino acid and nucleotide sequence homology with the reported sequences of murine (10) and Syrian hamster (12) IAP elements are indicated. Probable regulatory elements in the LTRs are underlined and described in the text. Short inverted repeats at the ends of LTRs are underlined with half-arrows. RT, reverse transcriptase; PBS, primer-binding site.

MATERIALS AND METHODS

Genomic DNA cloning. CHO DUKX B11 cells deficient for dihydrofolate reductase were used in this study (20). A CHO cell genomic library in lambda EMBL3 was prepared by conventional techniques (15). CHO cell genomic DNA was subjected to partial *Sau*3A restriction enzyme digestion, ligated into the *Bam*HI site of lambda EMBL3, and packaged utilizing the EMBL3 vector kit (Stratagene). Restriction enzymes and ligase were supplied by New England BioLabs. Recombinant plaques were identified by hybridization at low stringency (2× SSC [1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate] at 50°C) with a 2.5-kbp *Hin*dIII DNA fragment derived from the Syrian hamster IAP clone H18, which encompasses most of the *pol* gene (12). Two clones, CHIAP34 and CHIAP61, were identified for further analysis. CHIAP34 was subjected to nucleotide sequence analysis.

DNA sequence analysis. The DNA sequence was determined for both strands of the 5- and 1.1-kbp *Bam*HI fragments of the CHIAP34 provirus by using a *Bal* 31 nuclease deletion series procedure (13) and dideoxy sequencing procedures (16). The DNA sequence of nucleotides 1 to 443 was determined by using synthetic oligonucleotide primers (19) and Sequenase version 2.0 (United States Biochemicals). The nucleotide sequence between the *Bam*HI sites at 5310 and 5380 was also determined by using synthetic oligonucleotide primers. Nucleotide sequences and translations were analyzed by using the GCG package of genetic analysis programs (4).

Southern blot DNA analysis. High-molecular-weight DNA was prepared, and 10 μg was digested, subjected to agarose gel electrophoresis, and blotted onto nitrocellulose (15). A 5-kbp *Bam*HI fragment derived from CHIAP34 was used as a probe under low-stringency hybridization conditions. Syrian hamster kidney (BHK-21) cells and Armenian ham-

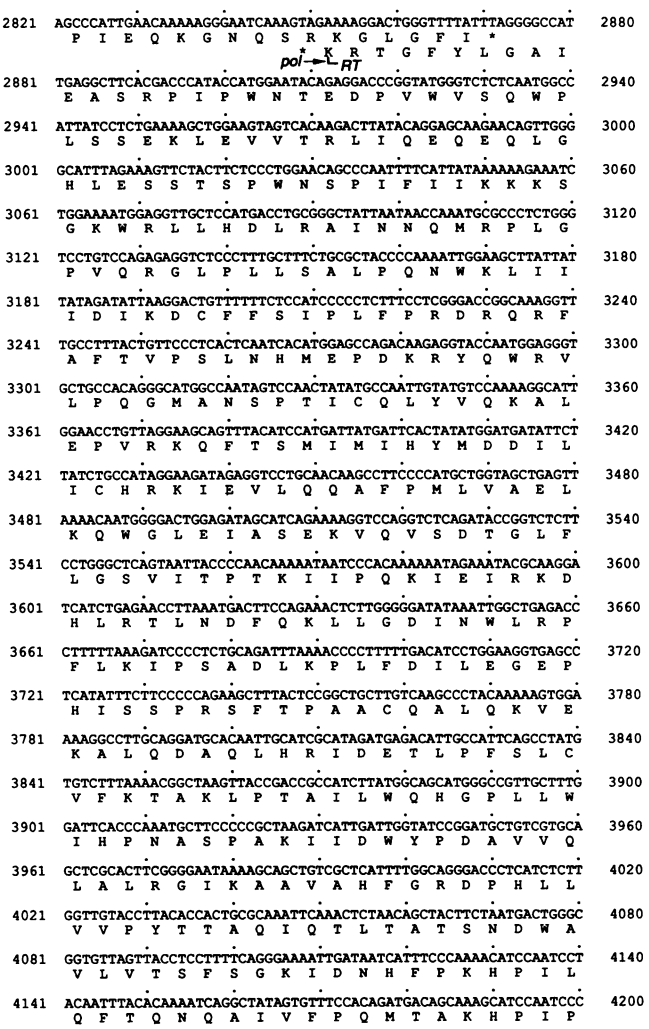

FIG. 2—*Continued.*

ster lung (AHL-1) cells were obtained from American Type Culture Collection.

RESULTS AND DISCUSSION

Molecular cloning of CHIAP sequences. A CHO cell genomic DNA library was screened at low stringency with a 2.5-kbp *Hin*dIII fragment derived from the SHIAP clone H18 (12). This fragment contains most of the *pol* gene. An initial screen of 2×10^5 plaques yielded 75 positively hybridizing plaques. A second screen of 25 plaques yielded 9 positive plaques, which were analyzed by restriction digestion and hybridization with the SHIAP *pol* probe. From these, three clones were identified which exhibited *Bam*HI fragments which hybridized strongly with the *pol* probe. Two provirus clones designated CHIAP34 and CHIAP61 were selected for further analysis. Restriction maps of these two clones are shown in Fig. 1. Based on Southern blot analysis (data not shown), CHIAP34 and CHIAP61 contain a common *Eco*RI site within the *gag* gene and a *Bam*HI site near the end of the *gag* gene. In addition, several *Hin*dIII sites were detected in common between the clones.

Nucleotide sequence of CHIAP34. The complete nucleotide sequence of the CHIAP34 provirus was determined. The

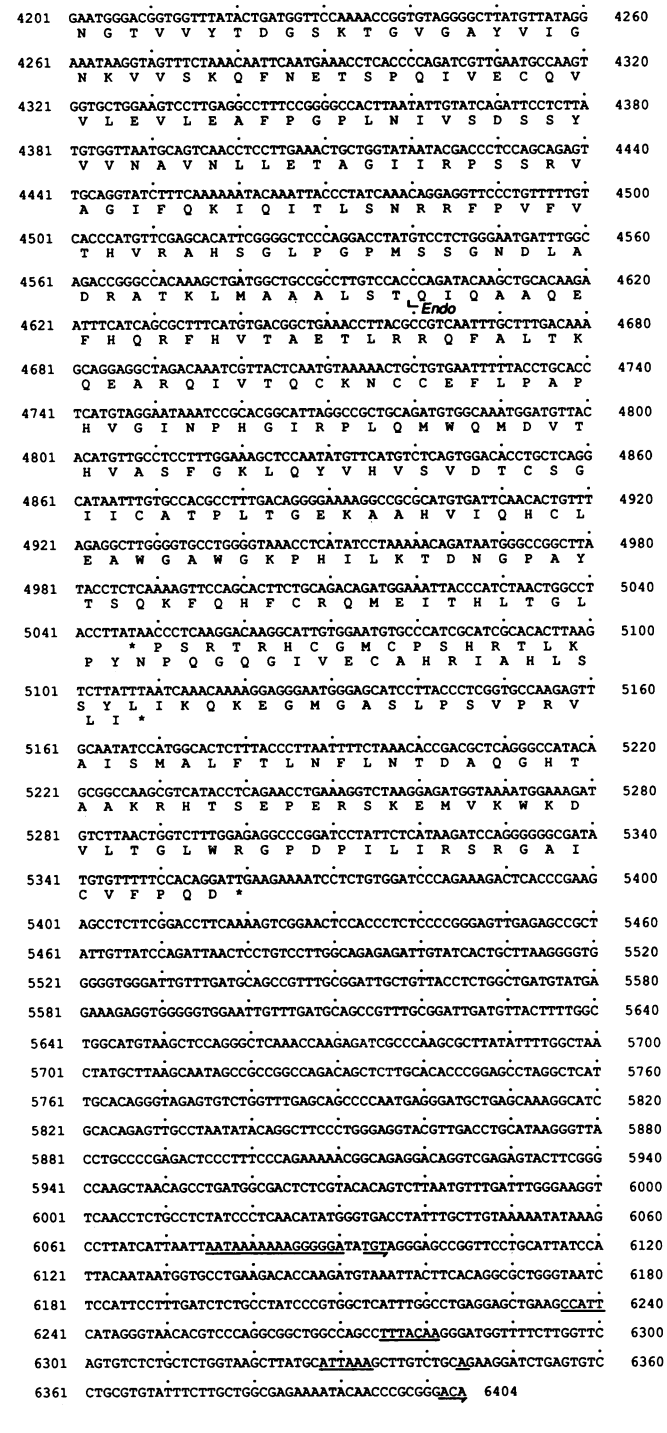

FIG. 2—*Continued.*

nucleotide and deduced amino acid sequences of CHIAP34 are shown in Fig. 2. CHIAP34 was 6,403 bp long, with LTRs of 311 bp present at both ends. The LTRs were 2.3% divergent from each other, suggesting an integration time of 1.76 million years based on 0.8 million years/1% divergence (2). Sequence analysis of flanking sequences revealed the presence of a 6-bp repeat of cellular DNA (GATGAT) at the ends of the provirus. On each LTR, common structural and

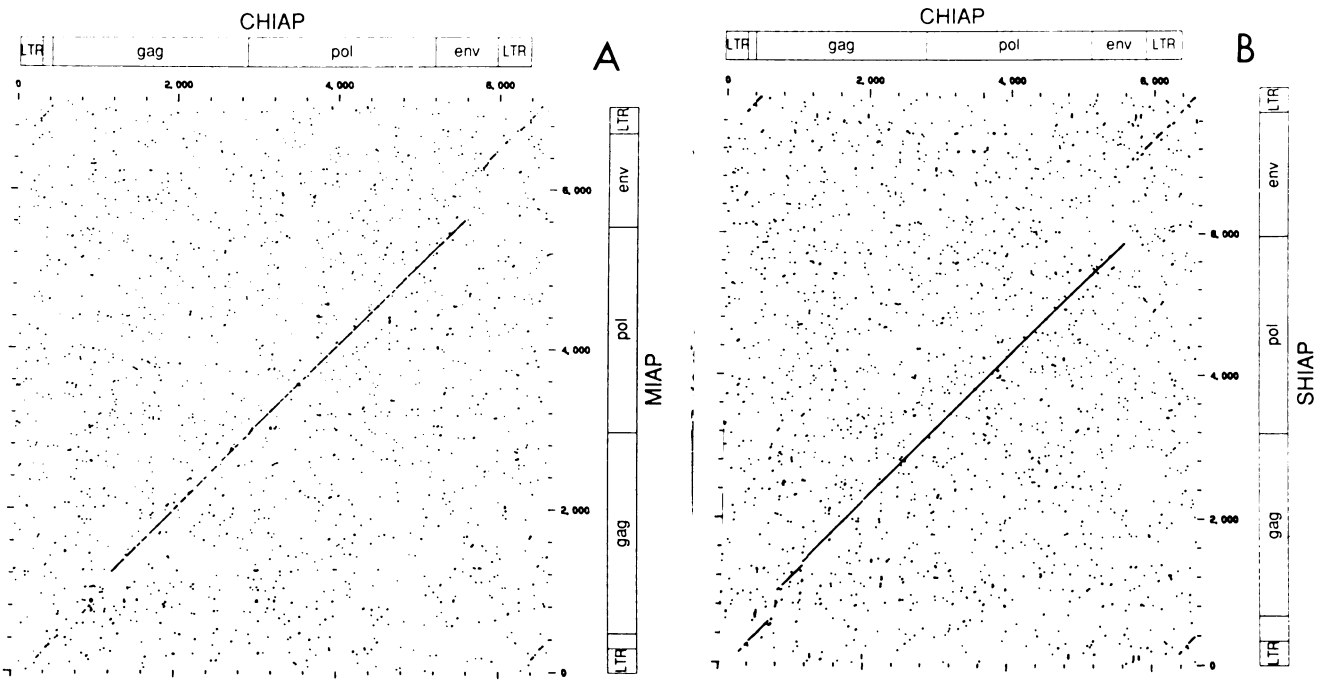

FIG. 3. Nucleotide sequence comparison of CHIAP34 and a murine IAP element, MIA14 (A), or a Syrian hamster IAP element, H18 (B). Dot-matrix analysis was generated by the GCG program package (4). Each dot represents a minimum homology of 14 nucleotides over a 21-nucleotide interval. The genomic organizations are indicated.

regulatory elements were present (3). The 5′ LTR contained a CAAT box (CCATT) located at base 143 that was followed by a consensus TATA box sequence (TTTATAA) at base 182. A probable polyadenylation signal (ATTAAA) was located at base 235 and was followed by the presumed polyadenylation site (CA) 9 bp downstream. Downstream of the 5′ LTR, the primer-binding site (PBS) was identified by its complementarity to 17 of the terminal 18 bases of mammalian phenylalanyl-tRNA. Phenylalanyl-tRNA has also been identified as the putative primer tRNA for Syrian hamster and murine IAP elements (11). Adjacent upstream to the 3′ LTR was a typical polypurine tract of 16 bp.

Nucleotide sequence homology with mouse and Syrian hamster IAP elements. Complete nucleotide sequences of mouse (MIA14) and Syrian hamster (H18) IAP-related proviral genomes have been reported (10, 12). A dot-matrix nucleotide sequence comparison between CHIAP34 and MIA14 is shown in Fig. 3A and one between CHIAP34 and H18 is shown in Fig. 3B. Substantial sequence similarity existed among the Chinese hamster and mouse and Syrian hamster elements at distinct regions of the genome. Little similarity existed within the 5′ portion of the genome including the 5′-terminal region of *gag*. However, beginning with sequences which encode protein sequence with homology to p27 and extending through the *pol* gene, extensive similarity was apparent among all three proviral sequences. Sequence similarity was reduced in the putative *env* gene and 3′-terminal sequences. It is evident from this analysis that the Syrian and Chinese hamster proviruses are more closely related to each other than either is to the mouse sequence.

Comparison of a partial Chinese hamster provirus sequence consisting of the 5′ LTR and partial *gag* gene sequences (17) with that of CHIAP34 revealed 94% nucleotide sequence homology. However, the absence of significant long ORFs in this clone precluded deduced amino acid comparison with CHIAP34. Several cDNA clones of retrovirus-related sequences expressed in CHO cells have been isolated which consist of partial and deleted sequences without intact *gag* and *pol* genes (1). Their relationship to CHIAP34 is unknown.

Genetic organization of CHIAP34. Translation of the CHIAP nucleotide sequence revealed the presence of several long ORFs (Fig. 4). To determine the genetic organization of CHIAP34, we compared the deduced amino acid sequences of these ORFs with the predicted *gag* and *pol* gene products of MIA14 and H18. By this analysis, ORFs 1

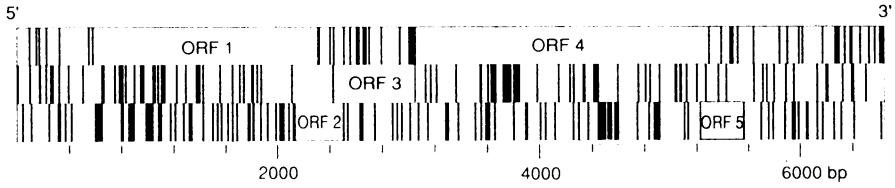

FIG. 4. ORFs of CHIAP34. Stop codons in each reading frame of the coding-strand sequence are indicated by vertical lines. ORFs encoding *gag* and *pol* protein domains are numbered. ORFs 1 to 3 compose the *gag* gene. ORFs 4 and 5 compose the *pol* gene.

FIG. 5. Comparison of deduced amino acid sequences of the *gag* gene products of CHIAP34 and MIA14. ORFs 1 to 3 were translated and combined together based on homology with the reported *gag* gene of MIA14 to generate a putative *gag* gene product for CHIAP34. Sequence comparison with the deduced amino acid sequence of the MIA14 *gag* gene product is shown. Lines between amino acids indicate a conserved amino acid residue, while dots indicate conservative changes. Periods within the amino acid sequence indicate gaps inserted for maximal alignment. The processed protein domains of p27, p12, and prt (protease) are indicated. The single-letter amino acid code is used.

FIG. 6. Comparison of deduced amino acid sequence of the *pol* gene products of CHIAP34 and MIA14. ORFs 4 and 5 were translated and combined together based on homology with the reported *pol* gene of MIA14 to generate a putative *pol* gene product for CHIAP34. Sequence comparison with the deduced amino acid sequence of MIA14 is shown. Lines between amino acids indicate a conserved amino acid residue, while dots indicate conservative changes. Periods within the amino acid sequence indicate gaps inserted for maximal alignment. The inferred junction between the reverse transcriptase and endonuclease domains is indicated by Endo. The single-letter amino acid code is used.

through 3 were determined to encode the *gag* gene. Beginning with a methionine codon at base 412, ORF 1 encoded 567 amino acids. As a result of frameshifts around bases 2058 and 2285, ORFs 2 and 3 also encoded *gag*-related amino acids. When these ORFs were translated together and compared with the deduced amino acid sequence of the MIA14 *gag* gene product, significant similarity (61% identity) was detected beginning at residue 224 within the putative p27 coding region (Fig. 5). This similarity extended to the end of the *gag* gene product. The amino-terminal 223 residues of CHIAP34 *gag* showed no similarity with the corresponding region of MIA14. This region of *gag* is highly divergent between different IAP genomes and shows no similarity with the *gag* genes of other retroviruses (6).

ORF 4 is entered by a −1 frameshift from the *gag* ORF 3 and encodes 752 amino acid residues of the *pol* gene before interruption by a stop codon. ORF 5 continues the *pol* amino acid sequence for another 91 residues before interruption by

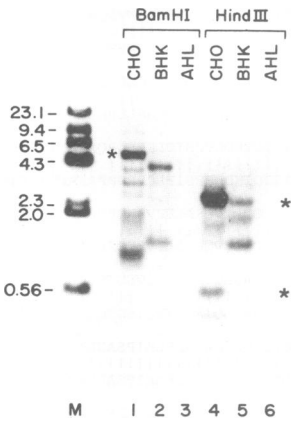

FIG. 7. DNA blot analysis of genomic organization of IAP-related sequences in hamster cell lines. High-molecular-weight DNA from Chinese hamster ovary (CHO) cells, Syrian hamster kidney (BHK) cells, and Armenian hamster lung (AHL) cells was digested with *Bam*HI (lanes 1 to 3) or *Hin*dIII (lanes 4 to 6). DNA was blotted onto nitrocellulose and hybridized with the 5-kbp *Bam*HI fragment from CHIAP34 which contains both *gag* and *pol* gene sequences. Fragments contained in CHIAP34 are indicated by asterisks (*). Numbers to the left indicate the sizes in kilobase pairs of *Hin*dIII-digested lambda DNA.

a stop codon at base 5358. Comparison of the translation product of these two ORFs with the *pol* gene product of MIA14 revealed extensive amino acid sequence similarity (78% identity) throughout both the reverse transcriptase and endonuclease domains (Fig. 6). On the basis of this amino acid similarity, the *pol* gene of CHIAP34 encodes a protein of 832 residues. This is close in size to the 867-residue *pol* gene product of MIA14. The extensive homology at the nucleotide and deduced amino acid sequence levels with Syrian hamster and murine IAP genomes suggests that the CHIAP34 provirus represents an endogenous IAP-like genome of the Chinese hamster.

The region corresponding to the *env* gene contained multiple stop codons in all three reading frames. Nucleotide sequence similarity with the *env* gene of H18 was only 67%. The greater degeneracy of the *env* gene compared with the *gag* and *pol* regions suggests that this provirus last moved in the genome without an intact *env* gene. The relative conservation of the *gag* and *pol* ORFs compared with the *env* gene suggests that these intact ORFs but not an intact *env* gene are necessary for transposition of the proviral elements.

Organization of CHIAP elements in the CHO genome. The organization of retrovirus sequences in the CHO cells was compared with that of cells from several different species of hamster. High-molecular-weight genomic DNA was isolated from Syrian hamster (BHK) cells, Armenian hamster (AHL) cells, and CHO cells and digested with *Bam*HI or *Hin*dIII. Southern blot analysis was performed with the CHIAP34 *Bam*HI fragment as a probe at low stringency. This fragment encompasses the *gag* gene and most of the *pol* gene of CHIAP34. Comparison of the *Bam*HI digestion pattern revealed a lack of common fragments between Chinese hamster and Syrian hamster sequences and no detectable hybridization with Armenian hamster sequences (Fig. 7, lanes 1 to 3). Similar comparison of the *Hin*dIII digestion pattern showed a fragment of approximately 2.5 kbp which appeared to comigrate in the Chinese hamster and Syrian hamster DNAs (Fig. 7, lanes 4 to 6). Additional Southern blot analysis demonstrated that the 5-kbp *Bam*HI fragment (indicated by an asterisk) and the 2.6- and 0.6-kbp *Hin*dIII fragments (indicated by asterisks) comigrated with fragments from CHIAP34 (data not shown), indicating that CHIAP34 is representative of a large number of similar members of a family of diverse but related sequences. Dot-blot analysis of copy number with the CHIAP34 *Bam*HI fragment as a probe indicated the presence of about 80 copies of hybridizable sequences per haploid genome (data not shown).

Comparison by Southern blot analysis of CHO cell DNA and Chinese hamster liver DNA revealed no differences in pattern or intensity of the retrovirus bands with the CHIAP34 *Bam*HI fragment as a probe (data not shown). The observation that the pattern and intensity of IAP elements in the Chinese hamster liver DNA and CHO cell DNA are the same indicated that no significant rearrangements or amplifications of these proviral sequences have occurred since CHO cells were derived from the Chinese hamster. Since the hybridization pattern of provirus sequences in the genomes of Chinese hamster cells and Syrian hamster cells did not exhibit extensive common components, the proviruses may have entered the Chinese hamster genome after the divergence of the Chinese hamster (*Cricetus griseus*) from the Syrian hamster (*Mesocricetus auratus*) 7.5 million years ago (2). The absence of hybridization with Armenian hamster cell DNA suggests that the proviral sequences are more distantly related or are absent.

While IAPs are common in most rodent cells, they have not been reported in CHO cells (6). Syrian hamster cells but not Chinese hamster cells can be induced to produce IAPs and IAP-related RNA by treatments which block DNA methylation such as 5-azacytidine (7). Treatment of CHO cells with 5-azacytidine or iododeoxyuridine did not induce expression of retrovirus-specific RNAs detectable by Northern (RNA) blot analysis with the CHIAP34 fragment as a probe, while similar treatments of BHK cells did induce expression of retrovirus-specific RNAs (data not shown). The inability to induce IAP expression in CHO cells suggests that the retroviral elements in the Chinese hamster genome are under transcriptional regulation which is different from that of the elements in other rodent cells.

ACKNOWLEDGMENTS

We thank M. Ono for providing fragments of SHIAP H18 for use as probes, K. Kuranda for assistance in copy number quantitation, D. Merberg for assistance with computer sequence analysis, J. Coffin for critical review of the manuscript, and E. Fritsch for support. We acknowledge R. Helmiss for computer-generated artwork.

REFERENCES

1. **Anderson, K. P., Y. S. Lie, M. Low, S. R. Williams, E. H. Fennie, T. P. Nguyen, and F. M. Wurm.** 1990. Presence and transcription of intracisternal A-particle-related sequences in CHO cells. J. Virol. **64**:2021–2032.
2. **Benveniste, R. E., R. Callahan, C. J. Sherr, V. Chapman, and G. J. Todaro.** 1977. Two distinct endogenous type C viruses isolated from the Asian rodent *Mus cervicolor*: conservation of virogene sequences in related rodent species. J. Virol. **21**:849–862.
3. **Coffin, J.** 1984. Structure of the retroviral genome, p. 261–370. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed., vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Devereaux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequencing analysis programs for the VAX.

Nucleic Acids Res. **12**:387–395.

5. **Heine, U. I., I. Margulies, A. E. Demsey, and R. G. Suskind.** 1979. Quantitative electron microscopy of intracytoplasmic type A particles at kinetochores of metaphase chromosomes isolated from Chinese hamster and murine cell lines. J. Gen. Virol. **45**:631–640.
6. **Kuff, E. L., and K. K. Lueders.** 1988. The intracisternal A-particle gene family: structure and functional aspects. Adv. Cancer Res. **51**:183–276.
7. **Lesser, J., J. Lasneret, M. Canivet, R. Emanoil-Revier, and J. Peries.** 1986. Simultaneous activation by 5-azacytidine of intracisternal R particles and intracisternal A particle related sequences in Syrian hamster cells. Virology **155**:249–256.
8. **Lieber, M. M., R. E. Benveniste, D. M. Livingston, and G. J. Todaro.** 1973. Mammalian cells in culture frequently release type C viruses. Science **182**:56–59.
9. **Lueders, K. K., and E. L. Kuff.** 1983. Comparison of the sequence organization of related retrovirus-like multigene families in three evolutionarily distant rodent families. Nucleic Acids Res. **11**:4391–4408.
10. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61**:3020–3029.
11. **Ono, M., and H. Ohishi.** 1983. Long terminal repeat sequences of intracisternal A particle genes in the Syrian hamster genome: identification of tRNA-phe as the putative primer tRNA. Nucleic Acids Res. **11**:7169–7179.
12. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55**:387–394.
13. **Poncz, M., D. Solowiejczyk, M. Ballantine, E. Schwartz, and S. Surrey.** 1982. "Non-random" DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA **79**:4298–4302.
14. **Puck, T. T.** 1985. Development of the Chinese hamster ovary (CHO) cell for use in somatic cell genetics, p. 37–64. *In* M. M. Gottesman (ed.), Molecular and cellular genetics. John Wiley & Sons, Inc., New York.
15. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
17. **Servaney, M., J. J. Kupiec, F. Galibert, J. Peries, and R. Emanoil-Ravier.** 1988. Nucleotide sequence of the Chinese hamster intracisternal A-particle genomic region corresponding to 5′ LTR-gag. Nucleic Acids Res. **16**:7725.
18. **Stoye, J., and J. Coffin.** 1985. Endogenous retroviruses, p. 357–404. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed., vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Strauss, E. C., J. A. Kobori, G. Siu, and L. E. Hood.** 1986. Specific-primer-directed DNA sequencing. Anal. Biochem. **154**: 353–360.
20. **Urlaub, G., and L. A. Chasin.** 1980. Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. Proc. Natl. Acad. Sci. USA **77**:4216–4220.