JOURNAL OF VIROLOGY, Aug. 1985, p. 387–394
0022-538X/85/080387-08$02.00/0
Copyright © 1985, American Society for Microbiology
Vol. 55, No. 2

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

MASAO ONO,[1]* HIROYUKI TOH,[2] TAKASHI MIYATA,[2] AND TOMOE AWAYA[1]

Department of Molecular Biology, School of Medicine, Kitasato University, Sagamihara-shi, Kanagawa 228,[1] and Department of Biology, Faculty of Science, Kyushu University, Fukuoka-shi, Fukuoka 812,[2] Japan

Received 8 February 1985/Accepted 19 April 1985

We determined the complete nucleotide sequence of the intracisternal A-particle gene, IAP-H18, cloned from the normal Syrian hamster liver DNA. IAP-H18 was 7,951 base pairs in length with two identical long terminal repeats of 376 base pairs at both ends. On the coding strand, imperfect open reading frames corresponding to *gag* and *pol* of the retrovirus genome were observed, whereas many stop codons were present in the region corresponding to *env*. The putative H18 *gag* gene (809 amino acids) had a sequence homologous to the N-terminal half of the mouse mammary tumor virus *gag* gene and locally to the Rous sarcoma virus *gag* gene. The putative H18 *pol* gene (900 residues) was homologous to the Rous sarcoma virus *pol* gene almost throughout the entire region. Two conserved regions among the retrovirus *pol* genes have been reported. One presumably corresponds to the DNA polymerase and the RNase H domain, and the other corresponds to the DNA endonuclease domain of the multifunctional protein *pol*. By the comparison of the deduced amino acid sequences of the putative endonuclease domain of six representative oncovirus genomes, a phylogenetic tree of the oncovirus genomes was constructed, and the intracisternal A-particle (type A) genome was found to be more closely related to the mouse mammary tumor virus (type B) and squirrel monkey retrovirus (type D) genomes.

Morphologically and biochemically, intracisternal A particles (IAPs) are retrovirus-like structures which are consistently observed in a variety of tumor cells and in early embryonic cells derived from normal rodents, such as mice, rats, and Syrian hamsters (12). Mouse IAPs contain a major *gag*-like protein of 73,000 daltons (8, 18), a magnesium-dependent reverse transcriptase (31), and a polyadenylated RNA molecule (IAP RNA) (18). DNA sequences complementary to the IAP RNAs (IAP genes) are interspersedly present on the rodent chromosomes (12) in several hundred to a thousand copies per haploid genome (11, 13, 16, 28), although the function of these genes in vivo is unknown.

Recently, the amplification of a particular subset of the IAP genes (25) and activation of a proto-oncogene, c-*mos*, due to the integration of an IAP gene into its coding region close to the N terminal (2), have been observed in mouse myeloma cells. Furthermore, the integration of an IAP gene into the intron of an actively transcribed immunoglobulin kappa light chain gene has been reported to cause inactivation of that gene (9). These findings indicate that a considerable number of IAP genes present in the rodent genome can act as endogenous insertion mutagens which may cause genetic diseases such as cancer.

There may be a close evolutionary relationship between the IAP and retrovirus genes, because the molecularly cloned IAP genes from two species of *Mus* and Syrian hamster were 6 to 8 kilobases in length (10, 15, 16, 28) with long terminal repeat (LTR) sequences of ca. 0.35 kilobases at both ends of the gene (2, 4, 9, 16, 17). These LTRs possessed all of the structural features commonly observed on the retrovirus LTR, although they show no marked sequence homology with other retrovirus LTRs. Furthermore, a presumed primer tRNA (phenylalanine tRNA) (17) for reverse transcription was different from known retrovirus primers. Recently, a significant homology of the deduced amino acid sequence in the *pol* region among the retrovirus genes has been reported (3, 29). However, due to the incompleteness of the total nucleotide sequencing, the internal structure of the IAP gene and a definite relationship between the IAP and retrovirus genes have yet to be clarified. Among the cloned IAP genes, those from two *Mus* species were fairly polymorphic (10, 15, 16), and a portion of the IAP gene was missing in some of them. Those from the Syrian hamster, however, were not markedly polymorphic, and no deletion was observed among them (28). Thus, we chose the Syrian hamster IAP gene H18 (28), which is a representative among the IAP genes we have isolated, for determining the complete nucleotide sequence.

The nucleotide sequence of IAP-H18 predicted a typical LTR-*gag*-*pol*-*env*-LTR structure, although many stop codons were present in the region corresponding to *env*. The computer-assisted comparison of the deduced amino acid sequences corresponding to the putative DNA endonuclease domain of the *pol* region showed IAP-H18, mouse mammary tumor virus (MMTV) (type B), and squirrel monkey retrovirus (SMRV) (type D) genomes to be closely related. Furthermore, based on the homology of the deduced amino acid sequences corresponding to that region, we were able to construct a phylogenetic tree of six representative members of the oncovirus subfamily.

MATERIALS AND METHODS

Clones and DNA sequencing analysis. The IAP gene clone IAP-H18 from the Syrian hamster has been previously described (28). DNA fragments were labeled either at 5′ ends with [γ-^{32}P]ATP (Amersham Corp., Arlington Heights, Ill.) and T4 polynucleotide kinase or at 3′ ends with [α-^{32}P] ddATP (Amersham Corp.) and terminal deoxynucleotidyl

* Corresponding author.

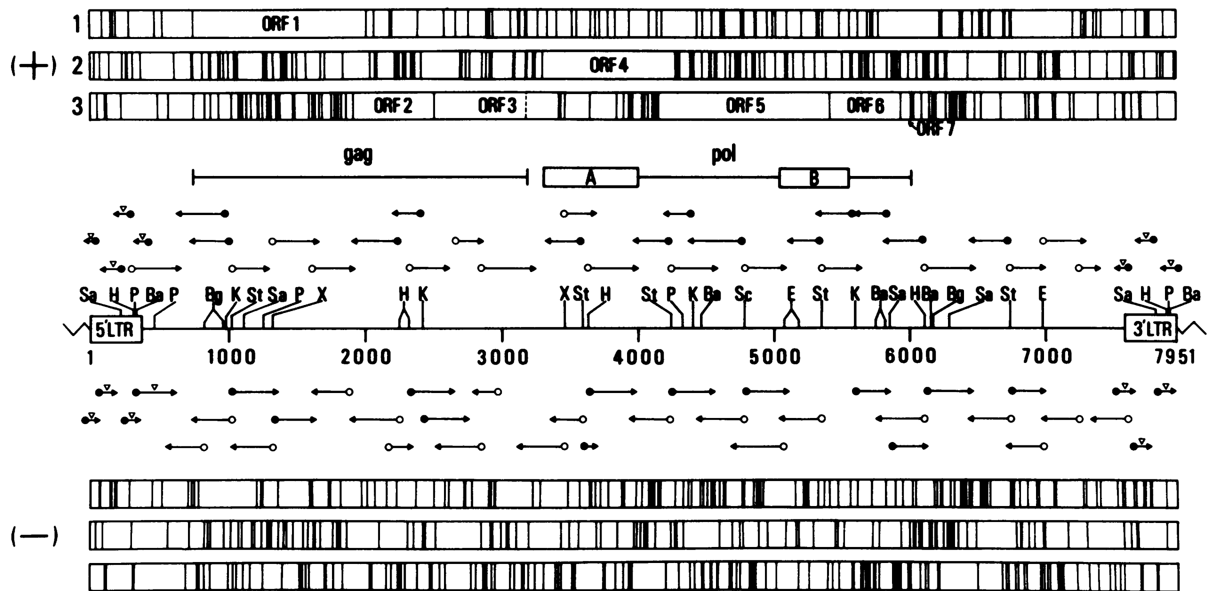

FIG. 1. Restriction map, sequencing strategy, and open reading frames of the Syrian hamster IAP gene, H18. The arrows indicate the extent and direction of the sequence determined for the coding strand (top; +) and the noncoding strain (bottom; −). Symbols: ○, labeled at the 5′ end; ●, labeled at the 3′ end; and ▽, determined previously (17). Stop codons in each phase of both strands are shown by the vertical lines. Putative *gag* and *pol* regions are shown, and two mutually conserved regions, A and B, in the retrovirus *pol* genes are boxed. Restriction enzyme abbreviations are as follows: Ba, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; Sa, *Sac*I; Sc, *Sca*I; St, *Stu*I; and X, *Xba*I.

transferase. The nucleotide sequence of the fragments was determined by the method of Maxam and Gilbert (14), and the sequencing strategy is shown in Fig. 1. For the comparisons of the deduced amino acid sequences by the homology matrix, a computer program was used to generate diagonal lines indicating segments of 30 residues long that show homology with a probability of occurrence by chance of less than 3×10^{-4} (29). With these matrices, the sequences were aligned manually or by the computer-assisted method of Sankoff (21). Amino acid sequence homology was calculated as the percentage of the mutually identical residues at the same position. In calculating homology, each gap was counted as one substitution, regardless of its length.

RESULTS

Structural features of IAP-H18. IAP-H18, a representative clone of the IAP genes present in multiple copies in the Syrian hamster genome (13, 28), was 7,951 base pairs (bp) in length with two identical LTRs of 376 bp at both ends (Fig. 2). On each LTR, many structural features commonly observed on the retrovirus LTR, such as the CAT box, TATAA box, and polyadenylation signal, were present (17). About 60% homology at the nucleotide sequence level was detected between H18 LTR and those of two *Mus* species (16), whereas these IAP LTRs had little sequence homology with either MMTV (6), Rous sarcoma virus (RSV) (22), Moloney murine leukemia virus (Mo-MuLV) (27), or human T-cell leukemia virus type I (HTLV-I) (24) LTRs. On the H18 LTR, four consecutive enhancer-like direct repeats of 23 bp as a standard unit were found (DR1 in Fig. 2) (17). Starting at position 611, eight consecutive and almost identical direct repeats of 14 bp with unknown functions were present (DR2 in Fig. 2) in the region corresponding to the leader sequence in the retrovirus genome. Except in the regions mentioned above, neither direct nor inverted repeats of more than 13 bp in length could be detected.

After translation of the H18 nucleotide sequence into the amino acid sequence, positions of the stop codons on each reading frame were identified (Fig. 1B). On the reading frames of the coding strand, several open reading frames (ORF1 to ORF7) were present in the region corresponding to *gag* and *pol* of the retrovirus genome, whereas many stop codons were found in the region corresponding to *env*. No open reading frame of more than 140 amino acid residues in length was detected on the reading frames of the noncoding strand.

IAP-H18 *gag* gene. To elucidate the internal structure of the H18 gene, deduced amino acid sequences of ORF1 to ORF7 were compared with known retrovirus *gag* or *pol* gene products by the computer-assisted graphical matrix method as previously described (29). Deduced amino acid sequences starting from ORF1 to ORF3 showed a meaningful resemblance to the RSV *gag* gene (22) in three regions (Fig. 3). The first region (N-p27) of ca. 90 residues was located at the N terminus of p27, a major structural protein of the RSV virion core; the second region (C-p27 + N-p12) of ca. 140 residues was positioned at the C terminus of p27 and the N terminus

FIG. 2. Nucleotide sequence of the Syrian hamster IAP gene, H18. The DNA sequence of the coding strand is given. LTRs are enclosed by brackets. The CAT box, TATAA box, and polyadenylation signal are boxed. A presumed primer-binding site (PBS) and short direct repeats (DR1 and DR2) are underlined. Nucleotides differing from a standard 14 bp of DR2 are shown by ▲. In the putative *gag* region, deduced amino acid sequences homologous to the RSV *gag* gene are presented in the box. In the putative *pol* region, deduced sequences of the two conserved regions, A and B, are presented in the box.

```
                                                        DR1
TGTGAGGAGCCGCCCTCGCTATCGCTATTGCCGTTAGAACATCGCGCTGACATCCACTGTCAGTTGGAGTTAACTGTCGCTGTCAGTTGGAGTTAACTGTCGCTGTCAGTTAGAGTTAAC  120
TGTCGCTGTCAGTTGGAGTTAACCGTTTAAACTGTGCCTCTCCCGTGACGTCATCTGGGCTCATCTCCAAACCACCAATCCCGCCTCTACACATCTCACTCGGAGCTCCTAGGCTTATAT  240
AAGGGGCTGGGTTTCTTAGCTTGGGGTCTCCCTCTAAGAAGCTCATCATCTATGTCTCAAGATCCATTAAAGCTTTACTCCAGAAGGATCCGAGTCTTCTCCGTCGTTCTTGCTCGCGAG  360
                   PBS
ACGGTAGCGCGGGACATCTGGTGCCGAAACCCGGGAACTCTTCAACATCGCCGGCGCCCGCGGAGACCCCTCGGAAAACGGGGCGGATTCAGAACTGCAGGGACGTAGGTTCAGAGAGGT  480
ATGCTTTATCCTGAACAACCCTTTTCTCGACTTTGGCCTTAGTTTGTCATCACCGTGGACGGGTCAGTCGAGGCTTTGGTCCTAGTCCTCTTTGTCCTTCTCCTTCATCGTCGCTCGTC  600
           DR2
CTGACAGAACGGCGCTGTCCATCTGGCGCTGTCCGTCTGGCGCTGTCCATCTGGCGCTGTCCATCTGGCGCTGTCCATCTGGTGCTGTCAGTCTGGTGCTGTCAGTCTGGTGCTGTCAGT  720
CTGGTAACGGAGTTGAGCGCGTTTTTTAGATAAGCTGTCAGCGCGATGGGTCGTCACAGTCAGTGGTTACAGCCTTACAGACTGTGCTAAAACAAAGAGATCTTAAAATTCCACCACGA  840
ACACTTCAAAATTTTATGAAGGAGGTAGATAGAGTGGCACCATGGTATGCTTGTTCCGGGTCCCTCACTCTGGCCTCATGGAACAAGCTAGGTAAAGAACTAGACAGGAAGCATGCACAA  960
GGAGATCTCTGCTTAGGTACCAAAGCCATTTGGAAGCTAGTTAAAAACTGCTTGGAGGATGAGGCCTCTCACCCTGCTATAATAGAAAGTCAGGGAACCTTAGAGGAACTTCAGGATAGC 1080
ATGTCAGAAACTGAACGGAGTGAGAGAATGGAGCTCGCAAAAGAAAACACATGTCTAAGAAAAAAGGCCCTCCCCAGGAAGTTAAGAAAGGGGAGAGAAAGAAGGGACTTATCATTCA 1200
CCCCTTAACAAATCTAAGAAAAAGAAAAAGCCAGAATCTAGTCAATACCCTGCACTAGAGCTAGAAGCCTTGGAGCTAGATAATTCTGATTCTCACACTCTGGATTCTAGCGAAGAAGGT 1320
GGTCTAGAGGAGGAGGTGGCACGGTATGAGGAAAAAAGATATCACCCTGATAGACACCGGCCACTGAAAACAAAAATGAATCTGAGGCCACCGCCTATAAATCCGGCGGGCTCACGCCCC 1440
TCAGCGCCTCCGCCATATGAGCTTCGGCTGAATACTGGTACTGACTCCTTTTACCTTTAGAGGAACGGAGAAAAATACAAATCGCTTTTCCAGTATTCGAGAATGCGCAAGGCGGAAGA 1560
                                                                                          ProValPheGluAsnAlaGluGlyGlyArg
                                                                                              p27
GTACATACCCCCGTAGATTACAATCAAATTAAAGAATTGGCCGAATCCGTTCGTAATTATGGGGTCAATGCTAACTTTACCACATTACAGGTGGAAAGGCTTGCTAATTTTGCCATGACT 1680
ValHisThrProValAspTyrAsnGlnIleLysGluLeuAlaGluSerValArgAsnTyrGlyValAsnAlaAsnPheThrThrLeuGlnValGluArgLeuAlaAsnPheAlaMetThr
CCCACTGACTGGCAAACGACAGTGAAAGCAGTGCTTCCCAACATGGGACAATACATGGAATGGAAGGCTCTCTGGTATGATGCAGCCCAGGCACAAGCTAGAGTTAATGCCACGGCTGAA 1800
ProThrAspTrpGlnThrThrValLysAlaValLeuProAsnMetGlyGlnTyrMetGluTrpLysAlaLeuTrpTyrAspAlaAlaGlnAlaGlnAlaArgValAsnAlaThrAla
GATGAGAATCAGAGGCAGTGGACATTTGAAATGCTGACAGGTCAGGGTCAACATGCCCTTAATCAAACCAACTACATCTGGGGCGTCTATGGTCAGGTGTCGGCCGCCGCTATTAAAGCA 1920
TGGAAGGCCCTCACAAAAAGGGGAGGTCCAGTGGACACCTTACAAAGATTGTCCAAGGACCCCAAGAGCCATTTTCAGACTTTGTGGCCAGAATGACAGAGGCCGCCAGTCGCATCTTTG 2040
                     GlyArgSerSerGlyHisLeuThrLysIleValGlnGlyProGlnGluProPheSerAspPheValAlaArgMetThrGluAlaAlaSerArgIlePhe
GTGATTCAGAACAAGCTATGCCCCTGATTGAGCAACTTGTCTATGAACAAGCAACCCAGGAATGTCGAGCGGCTATCGCCCCTCGCAAAAGCAAAGGTTTACAAGATTGGCTCAGAGTGT 2160
GlyAspSerGluGlnAlaMetProLeuIleGluGlnLeuValTyrGluGlnAlaThrGlnGluCysArgAlaAlaIleAlaProArgLysSerlysGlyLeuGlnAspTrpLeuArgVal
GCAGGGAGCTGGGGGGACCCCTAACCAATGCAGGACTAGCTGCCGCCATTTTACAAACGCACAGGTATAGAGATTTATCTAATCGAAAAGCTTGTTTTAATTGTGGGAGAATGGGGCACC 2280
CysArgGluLeuGlyGlyProLeuThrAsnAlaGlyLeuAlaAlaAlaIleLeuGlnThrHisArgTyrArgAspLeuSerAsnArgLysAlaCysPheAsnCysGlyArgMetGlyHis
                                         p27                                     p12
TTAAGAAAGATTGTCAAGCCCCTGAAAGAACAAGAGAATCGAAGCTTTGCTATCGTTGTGGTAAAGGTTATCATAGGGCTAGCGAATGCAGGTCAGTCCGGGATGTAAAAGGGAGGCTTC 2400
LeuLysLysAspCysGlnAlaProGluArgThrArgGluSerLysLeuCysTyrArgCysGlyLysGlyTyrHisArgAlaSerGluCysArg
TGCCCCCTATTGAAGAATCCAAGGTACCCCAATCAAAAAACGGGTCACGGGGCCCATGGTCCCAGGGCCCCAGAAATATGGAACCAATTTCAGAAAAGCAATCAAGAGGAAGTAGAGGT 2520
TTCCGAGGAACCTCCGGAATGGACTTGCGTGCCGCTCCGACTTCTTACTCATGCCCCAAATGAATGTTCAACCGGTCCCAGTCCATTCCCCTGGACCTTTGCCCCCTGCCACCATTGGGC 2640
TTATTCTGGGCCGTGGCTCCTTAACTTTACAGGGACTCATTATATATCCTGGGATTGTTGACCCATATCATAAAGAGGAGATTCAAGTTCTCTGTTCTAGCCCTCGGGGTGTATTTTCTA 2760
TAAAGCAGGGAGATAGAATTGCACAACTGGTGTTGCCGCCCTCCCTCGGGGATGGAGAAACCTACACTTTACAAAAAAGGGCTATGGGATCTTCTGGAAGTGATTCAGCCTATCTGGCTA 2880
TACCCCTAAATGAGAGACCTACAATAAAATTGTTAATTAATGGAAAAGAATTTGAAGGGATAATGGATTCGGGAGCAGACAAAAGCATTATTTCACTCCATTGGTGGCCGAAGTCTTGGC 3000
                                                         GlyIleMetAspSerGlyAlaAspLysSerIleIleSerLeuHisTrpTrpProLysSerTrp
CCACTGTTGTTTCATCTCATTCTCTACAAGGTCTTGGATACCAGTCCTCTCCTGCCATCAGTGCCTCAGCCTTAACCTGGCGGGATGCTGAAGGCAAACAGGGATGTTTTACCCCCTATG 3120
ProThrValValSerSerHisSerLeuGlnGlyLeuGlyTyrGlnSerSerProAlaIleSerAlaSerAlaLeuThrTrpArgAspAlaGluGlyLysGlnGlyCysPheThrProTyr
                                                                p15
TGTTGCCACTCCCTGTAAATTTATGGGGACGAGATGTGTTACAAGCCATGGGCATGACCCTAACTAATGAATACTCCCCCCAGGCATCAGCCATTATGGCAAAGATGGGGTATACAAATC 3240
ValLeuProLeuProValAsnLeuTrpGlyArgAspValLeuGlnAlaMetGlyMetThrLeuThrAsnGlu
GTAGAGGCTTGGGTAGGCAAGAACAAGGCAGAATAAAACCCATTACACAACACGGAAATCGGGGTAGAAAAGGACTGGGTTTTATTTGGGGGCCGTTGAGGCTTCACGACCCATACCATG 3360
                                                                                     LeuGlyAlaValGluAlaSerArgProIleProTrp
GAAAACAGAGGAGCCGTTATGGGTCTCTCAATGGCCTCTATCCTCTGAAAAATTAGAGGCTGTCACAAGATTAGTGCAAGAACAGGAACGGCTGGGGCATCTAGAGCCTTCTACCTCCCC 3480
 LysThrGluGluProLeuTrpValSerGlnTrpProLeuSerSerGluLysLeuGluAlaValThrArgLeuValGlnGluGlnGluArgLeuGlyHisLeuGluProSerThrSerPro
ATGGAATACACCAATTTTTGTTATTAAAAAGAAATCTGGGAAGTGGAGATTACTCCACGACCTGCGGGCCATTAACAATCAGATGCATCTTTTTGGCCCTGTTCAAAGAGGCCTTCCTTT 3600
 TrpAsnThrProIlePheValIleLysLysLysSerGlyLysTrpArgLeuLeuHisAspLeuArgAlaIleAsnAsnGlnMetHisLeuPheGlyProValGlnArgGlyLeuProLeu
GCTTTCTGCACTTCCTCAAGATTGGAAGCTTATTATTATAGATATTAAGGATTGTTTCTTCTCTATTCCACTTTACCCACGGGATAGACCAAGGTTTGCCTTCACTATCCCTTCTCTTAA 3720
 LeuSerAlaLeuProGlnAspTrpLysLeuIleIleIleAspIleLysAspCysPhePheSerIleProLeuTyrProArgAspArgProArgPheAlaPheThrIleProSerLeuAsn
A
TCATATGGAACCAGACAAGAGATTTCAGTGGAAGGTACTGCCGCAAGGCATGGCCAATAGCCCAACAATATGTCAGCTATATGTGCAGGAGGCTTTGGAGCCAATTAGGAAGCAATTTAC 3840
 HisMetGluProAspLysArgPheGlnTrpLysValLeuProGlnGlyMetAlaAsnSerProThrIleCysGlnLeuTyrValGlnGluAlaLeuGluProIleArgLysGlnPheThr
ATCTTTAATCGTTATCCATTATATGGACGATATTCTTATTTGTCACAAAGAGTTAGATGTTTTACAAAAGGCTTTTCCCATGCTGGTGGCTGAATTAAAACAGTGGGGATTAGAGATTGC 3960
 SerLeuIleValIleHisTyrMetAspAspIleLeuIleCysHisLysGluLeuAspValLeuGlnLysAlaPheProMetLeuValAlaGluLeuLysGlnTrpGlyLeuGluIleAla
CTCAGAAAAGGTTCAAATTGCAGATACAGGGCTCTTTTTGGGCTCAAAGATTACTCCCAAAAATATAGTTCCTCAGAAAATAGAAATTCGCAAAGACCACTTACAAACATTAAATGATTT 4080
 SerGluLysValGlnIleAlaAspThrGlyLeuPheLeuGlySerLysIle
TCAGAAACTTCTGGGAGATATAAATTGGTTGAGGCCCTTTTTAAAAATTCCATCTGCTGATCTAAAACCCCTCTTCGATCTCCTGGAGGGTGAGCCTCATATCTCCTCACCCAGAAAATT 4200
TACCCCTGCCGCCCACCGAGCCTTACAGATGGTGGAAGAGGCCTTACAGGAAGCACAAATTACAACGAATTGAGGAGTCACTACCTTTTAATCTGTGTGTTTTTAAGACAGCTCAATTGC 4320
CAACTGCAGTCCTGTGGCAACATGGACCATTACTTTGGATTCATCCAAATGCTTCCCCTGCTAAGATTATTGACTGGTACCCTGATGCTGTTGCACAGCCGCGCTCGCGAATAAAAGCAG 4440
CTGTTACACATTTTGGCAGGGATCCTGATTCCTTGATTGTGCCTTATACTGCTGCACAAGTTCAGACCTTGGCAGCCACATCTAGTGATTGGGCAGTTTTAGTTACTTCCTTTTCAGGAC 4560
AAATTGATAATCATTTTCCAAAGCATCCAATTTTACAGTTTGCCCTAAACCAGGCTATAGTGTTTCCACAAGTGACAGCTAAAGACCCACTCCCAGATGGGACTGTGGTGTACACTGATG 4680
GATCAAAAACTGGTCTGGGAGCTTATGTGGTCAAGGATAGGGTTATATCTAAACAATATAATGAAACTTCACCTCAAGTTGTAGAATGTCTGATAGTACTGGAGGTTCTTGAGGCTTTTC 4800
CAGGACCCCTCAATATTGTATCAGATTCCTCTTATGTGGTTAATGCAGTTAATCTCCTTGAGATAGCTGGAATAATAAGATCCTCCAGCAGAGTTGCTAACATCTTTCAGAAGATACAAG 4920
CTGCCTTGTTAAATAGGAGATTTCCTGTTTTCATTACCCATGTCCGAGCACATTCTGGACTCCCAGGTCCCATGTCCTTAGGGAATGACTTAGCGGACAAGGCTACAAAGCTGGTGGCCA 5040
CTGCCTTGTCTACCCATGCACAAGCAGCAAAAGAATTCCATAAGCGCTTTCATGTGACGGCTGAGACCTTGCGCCGTCGTTTTGCTCTATCCAGGAAAGAGGCTAGGGAAATTGTTACTC 5160
                       AlaAlaLysGluPheHisLysArgPheHisValThrAlaGluThrLeuArgArgArgPheAlaLeuSerArgLysGluAlaArgGluIleValThr
AATGCCAGAATTGCTGCGAATTCTTACCTACTCCCCACATGGGAATAAACCCACGTGGCATTAGACCATTGCAGATGTGGCAGATGGATGTTACTCATATTCCCTCTTTTGGAAGGCTTC 5280
GlnCysGlnAsnCysCysGluPheLeuProThrProHisMetGlyIleAsnProArgGlyIleArgProLeuGlnMetTrpGlnMetAspValThrHisIleProSerPheGlyArgLeu
B
AATATGTTCATGTTTCTGTTGACACCTGTTCTGGCGTAATGTTTGCCACACCCCTAACAGGGGAAAAGGCCTCATATGTGATTCAACATTGCCTTGAGGCTTGGAGTGCTTGGGGAAAAC 5400
GlnTyrValHisValSerValAspThrCysSerGlyValMetPheAlaThrProLeuThrGlyGluLysAlaSerTyrValIleGlnHisCysLeuGluAlaTrpSerAlaTrpGlyLys
CCAGAATTTAAAAAACAGACAATGGACCAGCATATACCTCTCAAAAATTTCGACAGTTTTGTCGGCAAATGGATGTTACCCATCTAACTGGCCTACCTTATAACCCTCAAGGACAGGGCA 5520
ProArgIle***LysThrAspAsnGlyProAlaTyrThrSerGlnLysPheArgGlnPheCysArgGlnMetAspValThrHisLeuThrGlyLeuProTyrAsnProGlnGlyGlnGly
TTGTGGAACGTGCCCATCGTACACTCAAATCATATCTGATTAAACAAAAGGGGAGTATTGAGGATGTTCTACCCTCGGTACCAAGAGTTGCTGTATCCATGGCACTTTTTACCCTTAATT 5640
IleValGluArgAlaHisArgThrLeuLysSerTyrLeuIleLysGlnLysGly
TCTTGAACTGTGATGCCCAAGGCCACACTGCGGCTGATCGCCACAGTTTAGAGCCTGATAGGCCAAAAGAAATGGTAAAATGGAAAGATGTCTTGACTGATCTCTGGAAAGGCCCGGATC 5760
CTATTCTTATAAGATCCAGGGGAGCTGTGTGTGTTTTTCCACAGGAGGAAGAAAACCCCCTGTGGATCCCTGAAAGACTCACTCGAAGAGCTCCTTCACAACTCCAGAATAAGAAGGATG 5880
GAGGCCAAGTGGGTGATGAAGACTCTCCTACTCGGGATGATGGTTCCTAGTATGATGGGAGAGTTACGGTGGGCCATCATGTCTACCTTCCCCCTACCGATGCCAGTGATGCAGAATGCC 6000
CAGGTGTTCCCTAGATTTTTCACAAATAATAGAGAGTTGGGGCTTGCCTTTTTGCCTTTAGATCCTCAGATACAGAGTTTTCAGGATGAGAGAACCTTGTCATTAGTAGGAAGCTTGTGC 6120
TTTACTGTGATTTCTTCATTAAAGGATAATTTTATTCATATGGATCCAGAAAGATCTAATGCTAATCCTTTTGAATTAATGGAAGCTGTACTGATCTTTCTCCTATGATAGTGAGAAGGG 6240
ACAGATTGGAAAGGGGCATTAGGCAACAGTCTGAGTGAGGGACAGAGCTCTGGATCATCTGATTGGTCCTCCTCCAATGTTAAGTATACCCCTTTTAAGAGAACTAATTATTCAGCTGCC 6360
CTTTATTTGGGTTGTGAGTAATGATACCTTAGGTAAGAATCAGAAAGAAATAAATTGTTCTGCTTAGAATTGTTTTTATGCATTATGCTGGGATGCTATTAAGCACCCTTTTGCTTTGGT 6480
AATGCGCATGCCTCGATTTATTCCAGCTCCTGTTGAGGCGCCAAGTAATTTGAGTTTGTTTAGGGAAAAGCGAGATTTTTGGGATCTCTGCTATTATTGTGGGCCTGATAGCCACCGCCG 6600
CGGTTACTGCTTCGGTTACCGCCTCGGCCCTTGCACTTTCTGCCACTGTGCAGACAACTCAAACCATTAATGAACTATAGGCCACTGTGACCATGGTCCTAGATAAACAGGCCACTGCCA 6720
ATTCTCAGATACAAGGAGGCCTGATGTTGGTAAACCAACAGATTAATTTGGTTCAGGAACAATTGGATATATTGTGGCAGATGGCCAAATTGGGATGTGAACAAAGTTGCCTGGCCTTTG 6840
TGTCACTTAGTACAATATGAGAATTTTACAAGAGCAGCTAACCTGTCAAAAGCTCTGTCTCAACATTTATTACAGAATTGGACTTTTGAGTTTGAGCAGACACTCCGTGAGTTGAGAATG 6960
GCGATCCTGCAAGTGAATTCGACGCATCTGGATCTTTCTCTGATGGAAGACCTCTCCCCCTGGATTTCATCTGCATTTTCTTATTTTAAGGAATGGGTGGGGGTAGGATCATTTAGCCTA 7080
GTTCTCTGCTGTGGAATAGTGTTGATGCTTTGGATGGTGTGTACGCTCAGAGCCCAAACAAAAGGAGAAAGTGGTGGTGGCACAAGCACTTGCTGCCTTAGAACACGGAGCATCCCCTGA 7200
CATTTGGTTAACCATGTTAAAGCAATAGGTCGCTGGCCTCCCAGCTCTTGCACCCCACGAGGCTAGTCTCATTGCACGGGATAGAGTGGCTAGGCTTCAGCAGCCCAAGGAGTTGCAAGG 7320
CTAAGCACTGCACAGGAGAGGTCTGCGGTATAACGACTTTCTCCTGGGAGATAAGTCATCTTGCATGAAGGTTCTATGTCATGATCTCCTTCCCCCAGAAAAAAGACATCGGACTGGTCA 7440
GGACTTCCTCTGGGGATAAGACCCTGGAGAGGAGCTTTGTACATGTTCCGAATTTGCACATGGGGATTTGACCTCTATCTCCACTCCAAAGTTGTGGGTCTCCTATTGCTTCAAAAATAA 7560
                                                                   DR1
ATGAATAGGGGGAGATGTGAGGAGCCGCCCTCGCTATCGCTATTGCCGTTAGAACATCGCGCTGACATCCACTGTCAGTTGGAGTTAACTGTCGCTGTCAGTTGGAGTTAACTGTCGCTG 7680
TCAGTTAGAGTTAACTGTCGCTGTCAGTTGGAGTTAACCGTTTAAACTCTGCCTCTCCCGTGACGTCATCTGGGCTGATCTGCAAACCACCAATCCCGGCTCTACACATCTCACTCGGAG 7800
CTCCTAGGCTTATATAAGGGGCTCGGTTTCTTAGCTTGGGGTCTCCCTCTAAGAACCTGATCATCTATGTCTCAAGATCCATTAAAGCTTTACTGCAGAAGGATCCGAGTGTTCTGCGTC 7920
GTTCTTGCTGGCGACACGCTAGCCCGGGACA                                                                                           7951
```

GAG

POL

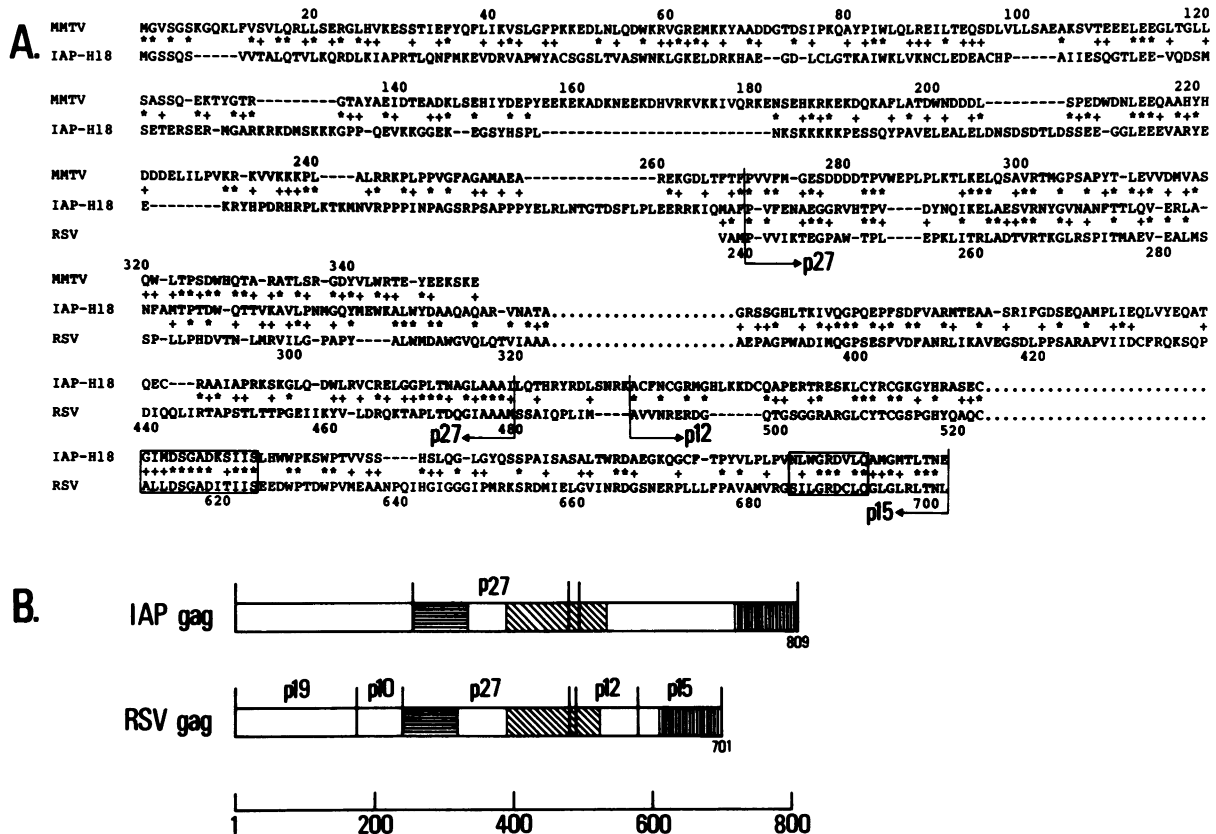

FIG. 3. (A) Alignment of the putative *gag* gene product of H18 with the MMTV and RSV *gag* gene. The deduced amino acid sequences of the putative H18 *gag* gene homologous to the MMTV or RSV *gag* gene are aligned. Amino acids are abbreviated as follows: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine, V, valine; W, tryptophan; and Y, tyrosine. Asterisks (*) indicate identical residues. Favored (conservative) amino acid substitutions are shown by (+). Favored substitutions are defined (23) as pairs of residues both belonging to the same group. The groups are as follows: A, G, P, S, and T; D, E, N, and Q; H, K, and R; I, L, M, and V; and F, W, and Y. Gap (−) is inserted to increase similarity. Two highly conserved stretches of sequence in the putative p15 of H18 and RSV as reported by Toh et al. (submitted for publication) are boxed. A highly divergent region is presented by (. . .). (B) Schematic representation of the putative H18 and RSV *gag* genes. Homologous regions are indicated by areas of similar shading.

of p12, a basic protein forming a complex with genome RNA. The third region (C-p15) was found at the C terminus of p15, which has a protease activity for the processing of polyprotein precursor into the mature form. In these regions, the amino acid sequence homology between the H18 and RSV *gag* gene product was calculated as 34% in N-p27, 28% in C-p27 + N-p12, and 32% in C-p15. The putative H18 p15 region had two highly conserved stretches of the sequence (Fig. 3A) commonly observed in retrovirus genomes such as Mo-MuLV, HTLV-I, and 17.6 (a copia-like mobile genetic element in *Drosophila* spp.) and cauliflower mosaic virus genomes. These may form the active site of acid protease-like activity (H. Toh, R. Kikuno, H. Hayashida, T. Miyata, and K. Saigo, submitted for publication).

Reported N-terminal 353 residues of the putative MMTV *gag* gene (7) could be aligned with the sequence deduced from ORF1 by the insertions of gaps into each sequence, whereas the N-terminal 236 residues of the RSV *gag* gene could not be aligned with the ORF1 sequence. In this region, the sequence homology between ORF1 sequence and MMTV *gag* gene product was 27%, whereas in the N-p27 region the homology was calculated to be 35%. ORF1 to ORF3 was thus concluded to be the H18 *gag* region.

IAP-H18 *pol* gene. The deduced amino acid sequence starting from ORF4 to ORF7 was found to be significantly homologous to the entire *pol* region of the RSV genome, so we tentatively designated this region as H18 *pol* (Fig. 1 and 2). The sequence alignment of this region is shown in Fig. 4. In addition, H18 *pol* had distant but detectable homology to the Mo-MuLV *pol* gene product in two regions. The first region (A) corresponds to 230 residues, starting from the N terminus of the RSV *pol* gene, whereas the other (B) corresponds to ca. 170 residues, starting from the 576th residue of the RSV *pol* gene. The observed sequence homology among these three *pol* gene products was 1.2 to 1.4 times higher in region A than B.

In region A, the estimated sequence homology between the H18 and RSV *pol* genes was 51%, and 70% of the residues were homologous when we included favored substitution for estimation. The sequence homology between the H18 and Mo-MuLV *pol* genes in region A was calculated as 30%, which is about the same (33%) as that between the RSV and Mo-MuLV *pol* genes in this region. The H18 *pol* gene was 37% homologous to the RSV *pol* gene in region B, where the significant sequence homology among the *pol* gene products of RSV, MMTV, and SMRV has been reported (see below). Assuming that the N terminus of the H18 *pol* gene starts at the same position as that of the RSV *pol* gene

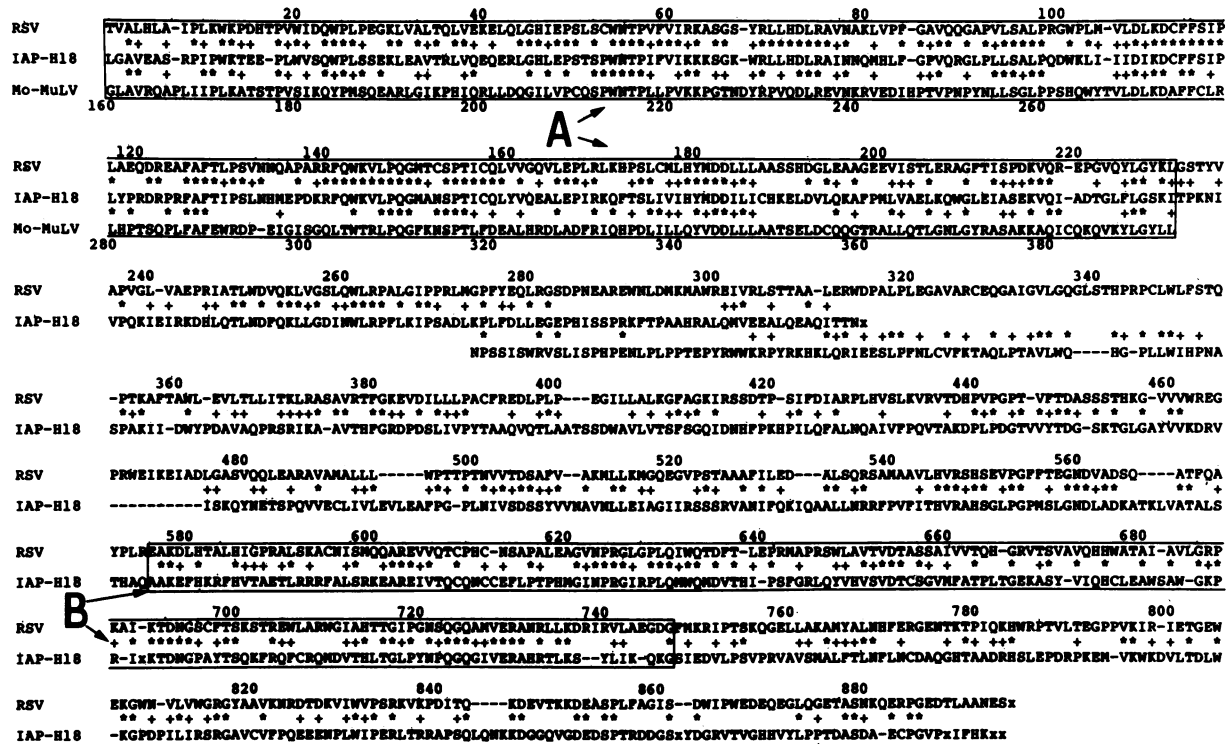

FIG. 4. Alignment of the putative *pol* gene product of H18 with RSV and Mo-MuLV *pol* genes. The putative H18 *pol* gene is aligned with the entire region of the RSV *pol* gene. Two conserved regions, A and B (boxed), between H18 and Mo-MuLV *pol* genes are also aligned. Definitions for amino acid abbreviations, (*), (+), and (−) are the same as described in the legend to Fig. 3. A stop codon is indicated as (×). The coding frame is shifting in the region starting at position 274 to 314 of the RSV *pol* gene.

and ends with Lys located just before the two stop codons (Fig. 4), the number of residues of the H18 *pol* gene was calculated as 900, a value about the same as the RSV *pol* gene (895 residues).

No significant homology of the deduced amino acid sequence in the *env* region could be detected between the H18 gene, including two open reading frames capable of encoding ca. 150 amino acid residues in the reading frame 1 on the coding strand, and MMTV, RSV, Mo-MuLV, or HTLV-I gene.

Close relationship among A-, B-, and D-type oncovirus genomes. By a comparison of the amino acid sequences corresponding to either the A or B region (Fig. 4) in which significant homology was observed in a variety of *pol* gene products (3, 29; Toh et al., submitted for publication), we tried to make a phylogenetic tree of the oncovirus subfamily. Although, as mentioned previously, the sequence homology in the A region was usually ca. 1.3 times higher than that in the B region, the available sequences of five representative oncovirus *pol* gene products were those of the B region, which is known as the putative DNA endonuclease domain of the *pol*. Therefore, we aligned these sequences as shown in Fig. 5. At 21 positions of 172 residues of the RSV *pol* gene, deduced amino acid residues from all six oncovirus genomes were identical, and at 70% of these positions, the residues of more than three oncogene products were identical.

From Fig. 5, the sequence homology was calculated (Table 1). A combination with the highest detectable homology (54%) was observed between MMTV and SMRV. The second highest homology was found in H18 versus MMTV and H18 versus SMRV, with ca. 50%, which was higher than any other combination including H18. We could thus conclude that, among the five representative oncovirus genomes, the IAP-H18 genome is closely related to both the MMTV and SMRV genomes in the putative endonuclease domain of *pol*.

Based on the sequence homology shown in Table 1, a phylogenetic tree of the oncovirus genomes was constructed (Fig. 6). Assuming the lower sequence homology to be due to earlier divergence from the common progenitor, mammalian type C oncovirus was considered to have first diverged from the progenitor of the oncovirus, then from the HTLV type, avian type C, and then from type A oncovirus in that order. Finally, types B and D diverged from each other.

DISCUSSION

In this study, we determined the entire nucleotide sequence of the representative clone, H18, of the Syrian hamster IAP genes and clarified the structure of the *gag* and *pol* regions of the retrovirus-like IAP gene.

The positions of both the N and C termini of the H18 *gag* gene have yet to be determined. Since the codon next to the C-terminal Glu was TAC (Tyr) in the putative H18 *gag* p15 and the stop codon TAG was positioned next to the C-terminal Leu in RSV p15, G→C conversion seems to give rise to the inconsistency of the C-terminal position between the H18 and RSV *gag* gene. Supposing that the N terminus of the H18 *gag* starts at the same position as the MMTV *gag* gene (Fig. 2 and 3) and ends with the C-terminal Glu of putative H18 p15, we calculated the number of residues of the H18 *gag* gene to be 809, which is 108 residues more than the RSV *gag* gene (701 residues).

The many stop codons present in the *env* region of H18

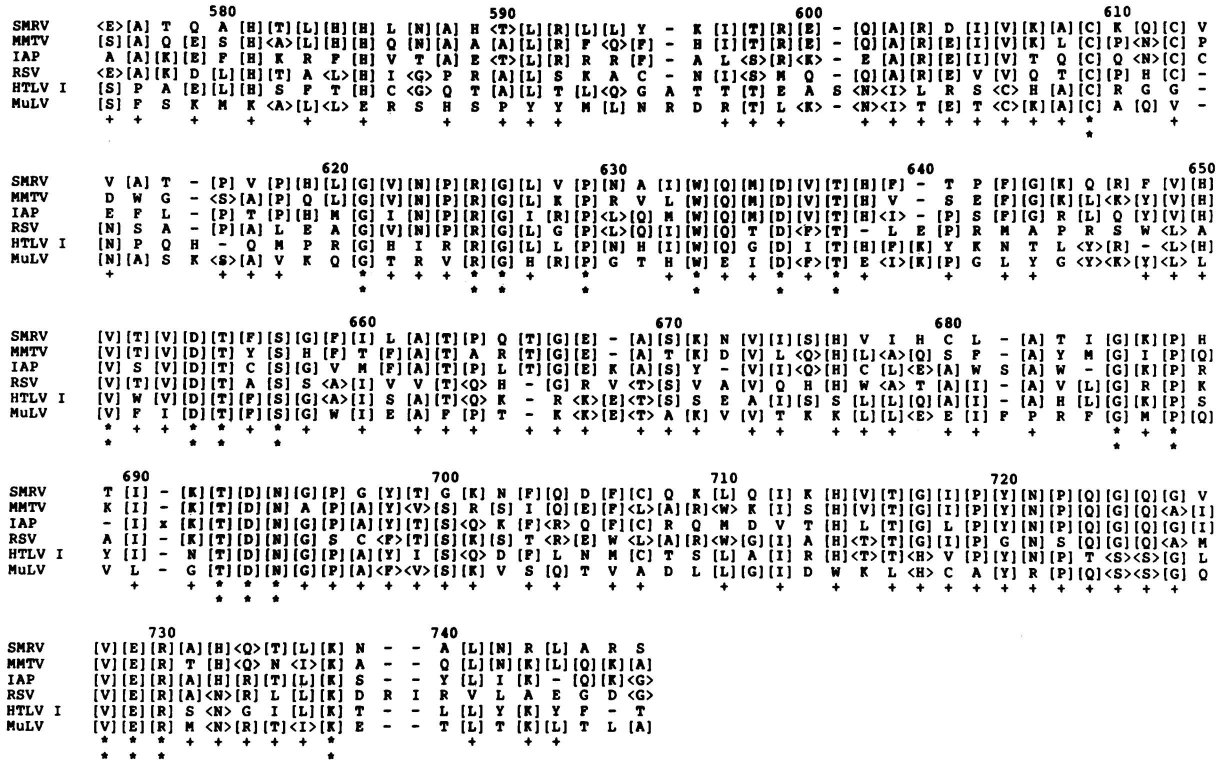

FIG. 5. Alignment of the putative DNA endonuclease domain of six representative oncovirus genomes. Deduced amino acid sequences of the *pol* regions corresponding to the region starting at position 576 to 747 of the RSV *pol* gene (total of 172 residues) are aligned. Identical residues at the same position are indicated as [ ] or ⟨ ⟩. Gap (−) is inserted to increase similarity. Also indicated are the position in which the residues of all six genes are identical (*), the position in which the residues of more than three genes are the same (+), and the stop codon (×).

must cause the production of prematurely terminated IAP *env* protein, even when the transcription and processing of the active *env* mRNA advance normally. This situation is also apparent in other IAP genes, since the presence of *env*-like proteins encoded by IAP RNA and associated with the IAPs has yet to be reported. Therefore, one probable reason for the inability of IAP genes to produce budding of the infectious virus particles through plasma membranes might be the lack of the *env* protein synthesis.

From the morphological, biochemical, and molecular biological standpoints, we can divide the oncovirus subfamily of the retrovirus family into six representative types, namely, A, B, avian C, mammalian C, D, and HTLV. The relationship among these members is still unclear. Recently, representative oncovirus genes have been cloned, and some of them have been sequenced. Based on filter hybridization experiments with cloned genes, Chiu et al. (3) found the sequence homology between the avian type C and type A, B, or D oncovirus genomes and determined the homologous region and extent of their homology. Instead of the IAP

TABLE 1. Amino acid sequence homology in the putative DNA endonuclease domain of the retrovirus *pol* genes[a]

Virus genome (type)	% Homology with:					
	SMRV	MMTV	IAP	RSV	HTLV-I	MuLV
SMRV (D)	100					
MMTV (B)	54	100				
IAP (A)	50	49	100			
RSV (avian C)	38	39	37	100		
HTLV-I	36	30	32	31	100	
MuLV (mammalian C)	27	28	25	23	32	100

[a] Sequence homology is indicated as a percentage of the mutually identical residues at the same position shown in Fig. 5. Each gap was counted as one substitution, regardless of its length.

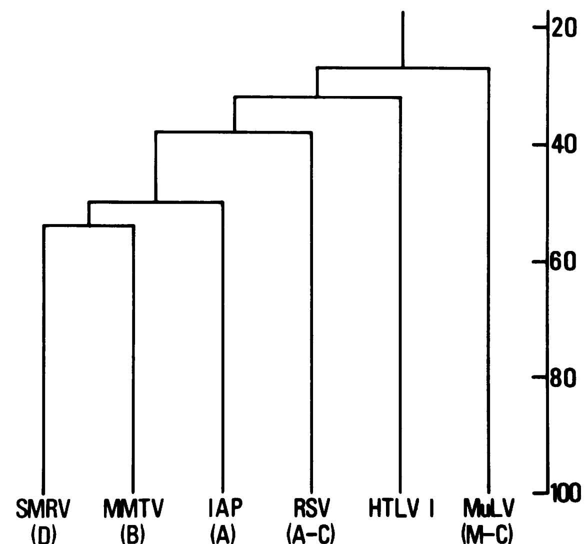

FIG. 6. Phylogenetic tree of oncovirus genomes. Abscissa represents the amino acid sequence homology indicated as a percentage.

gene, they used the M432 gene, assumed to be a recombinant between an IAP gene in the *Mus* genus and an unknown retrovirus gene (1). Their results, though qualitative, agree well with ours. To clarify the interrelationship of these members, attempts are presently being made to determine the homology among the cloned oncovirus genomes mainly at the deduced amino acid sequence level. A distant but significant homology of the deduced amino acid sequence in the *pol* region between RSV and Mo-MuLV was first described by Toh et al. (29). Later, Chiu et al. (3) reported the deduced amino acid sequence homology in the B region of *pol* among MMTV, SMRV, and RSV. Based on the deduced amino acid sequence homology in the *pol* B region, we were finally able to determine the interrelationship of the six representative oncovirus genomes.

The retrovirus *pol* gene product is a multifunctional protein (5, 30) possessing not only DNA polymerase activity but that of RNase H and DNA endonuclease as well. Thus, conserved regions such as A and B in *pol* are considered to form a domain having the same characteristic activity. The domain bearing both DNA polymerase and RNase H activity seems to be located in the A region, whereas the endonuclease domain possibly having a significant role in the integration of the retrovirus genome into the chromosome may quite possibly be in the B region. These intrinsic activities of the *pol* protein appear to be very essential for reverse transcription and integration of the retrovirus genome. Consequently, eucaryotic mobile genetic elements such as copia-like elements in *Drosophila* spp. and Ty elements in yeast cells, which were found to share similarities with the retroviral provirus genome (19, 26), have heen reported to possess the region having such activities. Recently, Saigo et al. (20) reported the complete nucleotide sequence of a copia-like 17.6 element and found three open reading frames presumably corresponding to *gag*, *pol*, and *env*. In the 17.6 putative *pol* gene, they found the deduced amino acid sequence homologous to both the A and B regions of the Mo-MuLV *pol* gene (20; Toh et al., submitted for publication). Since the copia-like element and IAP gene have been reported to share similar biological features (such as rare gene expression in normal tissue except for embryos and the inability to produce infectious particles budding from plasma membranes) and the usual gene expression in cultured or malignant cells was infrequently followed by translocation, the common structural and biological features between these two genes strongly suggest that the copia-like elements have been derived from an ancestor common to the retrovirus gene and formed by a mechanism similar to that of IAP genes accumulated in the rodent genomes.

ACKNOWLEDGMENTS

We thank K. Saigo and M. Kawakami for their valuable comments and encouragement throughout the course of this work.

This work was supported by a grant-in-aid from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED

1. **Callahan, R., E. L. Kuff, K. K. Lueders, and E. Birkenmeier.** 1981. Genetic relationship between the *Mus cervicolor* M432 retrovirus and the *Mus musculus* intracisternal type A particle. J. Virol. **40**:901–911.
2. **Canaani, E., O. Dreazen, A. Klar, G. Rechavi, D. Ram, J. B. Cohen, and D. Givol.** 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. U.S.A. **80**:7118–7122.
3. **Chiu, I.-M., R. Callahan, S. R. Tronick, J. Schlom, and S. A. Aaronson.** 1984. Major *pol* gene progenitors in the evolution of oncoviruses. Science **223**:364–370.
4. **Cole, M. D., M. Ono, and R. C. C. Huang.** 1981. Terminally redundant sequences in cellular intracisternal A-particle genes. J. Virol. **38**:680–687.
5. **Dickson, C., R. Eisenman, H. Fan, E. Hunter, and N. Teich.** 1982. Protein biosynthesis and assembly, p. 513–648. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
6. **Donehower, L. A., A. L. Huang, and G. L. Hager.** 1981. Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. J. Virol. **37**:226–238.
7. **Fasel, N., E. Buetti, J. Firzlaff, K. Pearson, and H. Diggelman.** 1983. Nucleotide sequence of the 5′ noncoding region and part of the *gag* gene of mouse mammary tumor virus: identification of the 5′ splicing site for subgenomic mRNAs. Nucleic Acids Res. **11**:6943–6955.
8. **Kuff, E. L., R. Callahan, and R. S. Howk.** 1980. Immunological relationship between the structural proteins of intracisternal A-particles of *Mus musculus* and the M432 retrovirus of *Mus cervicolor*. J. Virol. **33**:1211–1214.
9. **Kuff, E. L., A. Feenstra, K. Lueders, L. Smith, R. Hawley, N. Hozumi, and M. Shulman.** 1983. Intracisternal A-particle genes as movable elements in the mouse genome. Proc. Natl. Acad. Sci. U.S.A. **80**:1992–1996.
10. **Kuff, E. L., L. A. Smith, and K. K. Lueders.** 1981. Intracisternal A-particle genes in *Mus musculus*: a conserved family of retrovirus-like elements. Mol. Cell. Biol. **1**:216–227.
11. **Lueders, K. K., and E. L. Kuff.** 1977. Sequences associated with intracisternal A particles are reiterated in the mouse genome. Cell **12**:963–972.
12. **Lueders, K. K., and E. L. Kuff.** 1981. Sequence homologous to retrovirus-like genes of the mouse are present in multiple copies in the Syrian hamster genome. Nucleic Acids Res. **9**:5917–5930.
13. **Lueders, K. K., and E. L. Kuff.** 1983. Comparison of the sequence organization of related retrovirus-like multigene families in three evolutionarily distant rodent genomes. Nucleic Acid Res. **11**:4391–4408.
14. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **Ono, M., M. D. Cole, A. T. White, and R. C. C. Huang.** 1980. Sequence organization of cloned intracisternal A particle genes. Cell **21**:465–473.
16. **Ono, M., H. Kitasato, H. Ohishi, and Y. Motobayashi-Nakajima.** 1984. Molecular cloning and long terminal repeat sequences of intracisternal A-particle genes in *Mus caroli*. J. Virol. **50**:352–358.
17. **Ono, M., and H. Ohishi.** 1983. Long terminal repeat sequences of intracisternal A particle genes in the Syrian hamster genome: identification of the $tRNA^{Phe}$ as a putative primer tRNA. Nucleic Acids Res. **11**:7169–7179.
18. **Paterson, B. M., S. Segal, K. K. Lueders, and E. L. Kuff.** 1978. RNA associated with murine intracisternal type A particles codes for the main particle protein. J. Virol. **27**:118–126.
19. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in *Drosophila*, p. 329–361. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
20. **Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki.** 1984. Identification of the coding sequence for reverse transcriptase-like enzyme in a transposable genetic element, *17.6*, in *Drosophila melanogaster*. Nature (London) **312**:659–661.
21. **Sankoff, D.** 1972. Matching sequences under deletion/insertion constraints. Proc. Natl. Acad. Sci. U.S.A. **69**:4–6.
22. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
23. **Schwartz, R. M., and M. O. Dayhoff.** 1978. Matrices for detecting distant relationships, p. 353–358. *In* M. A. Dayhoff (ed.), Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biochemical Research Foundation, Washington, D.C.

24. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. U.S.A. **80:**3618–3622.
25. **Shen-Ong, G. L. C., and M. D. Cole.** 1984. Amplification of a specific set of intracisternal A-particle genes in a mouse plasmacytoma. J. Virol. **49:**171–177.
26. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. Nature (London) **302:**119–124.
27. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
28. **Suzuki, A., H. Kitasato, M. Kawakami, and M. Ono.** 1982. Molecular cloning of retrovirus-like genes present in multiple copies in the Syrian hamster genome. Nucleic Acids Res. **10:**5733–5746.
29. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305:**827–829.
30. **Varmus, H., and R. Swanstrom.** 1982. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Molecular biology of tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Wilson, S. H., E. W. Bohn, A. Matsukage, K. K. Lueders, and E. L. Kuff.** 1974. Studies on the relationship between deoxyribonucleic acid polymerase activity and intracisternal A-type particles in mouse myeloma. Biochemistry **13:**1087–1094.