Characterization of a mouse interferon gene locus I. Isolation of a cluster of four α interferon genes

K.A.Kelley and P.M.Pitha*

Oncology Center 1-109, Johns Hopkins University School of Medicine, 600 North Wolfe St.,
Baltimore, MD 21205, USA

Received 20 November 1984; Revised and Accepted 9 January 1985

ABSTRACT

A BALB/c mouse genomic library was screened with a murine interferon α_2 (MuIFN- α_2) cDNA coding region fragment. Eight clones were isolated which contain different mouse chromosomal segments related to the MuIFN- α_2 probe and a 28 kilobase (kb) region of mouse genomic DNA containing four different MuIFN- α genes (α_1 , α_4 , α_5 and α_6) was identified and characterized; an intergenic 1000 nucleotide long conserved sequence was found to be associated with three of these four α genes, indicating that this α -IFN gene cluster evolved through tandem duplications. Sequence analysis revealed the absence of a polyadenylation site in the 3' untranslated region of MuIFN- α_1 , and showed that one of the genes (α_4) contains an internal deletion of 5 amino acids in the coding region.

INTRODUCTION

In recent years, most attention has focused on the human interferon system and the availability of the cloned interferon genes has contributed greatly to our understanding of their molecular organization in human DNA. Thus it was shown that the virus-induced, acid stable type I human interferons (HuIFNs) are coded for by a multiple α gene family containing at least 12 closely related, but distinct genes and by the single copy β_1 interferon gene. Type I HuIFNs belong to a rare family of nonspliced genes, and are localized on the short arm of human chromosome 9; however, their exact position on this arm is not known as of yet. Some of the interferon genes are closely linked to each other, while the remaining α -IFNs and the β_1 interferon gene are not. Based on a significant homology between the HuIFN- α and β_1 genes, it has been suggested that these genes arose from a common ancestor by gene duplication (1).

Our knowledge about the mouse α and β IFN genes is much more limited. In the past year, we have cloned a mouse α_2 cDNA by reverse transcription of mRNA isolated from mouse C243 cells induced with Newcastle disease virus (NDV) to synthesize MuIFNs. A MuIFN- β_1 cDNA was recently isolated (2), while the cloning of a segment of genomic DNA containing MuIFN- α_1 and the isolation of a MuIFN- α_2 cDNA was also reported (3). Using these specific probes, it was shown that similarly as in humans, the MuIFN- α genes represent a family of at least 10-15 genes while the MuIFN- β_1 gene is present in the mouse genome in a single copy. Using mouse/hamster somatic cell hybrids and the MuIFN- α and β_1 cDNAs as specific probes, we have localized both the α and β MuIFN genes to chromosome 4 (4, and Kelley and Pitha, in preparation), indicating that in mouse and man, the type I interferon genes are clustered together on the same chromosome. Analysis of genomic DNAs from BALB/c and C57/BL mice have revealed EcoRI and HindIII restriction fragment length polymorphisms in several of the MuIFN- α s. Using recombinant and congenic strains of mice generated between these two inbred strains, we were able to show that the MuIFN- α gene family is linked to the minor-histocompatibility locus H-15 on chromosome 4 (5).

Previously, incomplete cDNAs for two other putative interferon genes (β_2 and β_3) were identified (6). These two genes did not show an extensive DNA homology with the MuIFN- β_1 gene and are not localized on chromosome 4 (unpublished results); further characterization is needed to confirm that these cDNAs truly represent additional β interferon genes.

In order to characterize in detail the MuIFN- α gene locus, we screened, in the present study, a BALB/c mouse genomic library with the MuIFN- α_2 cDNA and identified a 28 kb segment of mouse DNA containing four MuIFN- α genes. The characterization of this α gene cluster, and the nucleotide and amino acid sequence of these four genes are presented.

MATERIALS AND METHODS

Construction and screening of mouse genomic library

The mouse genomic library (provided by P. Leder) was

constructed by cloning size-selected partial MboI fragments from BALB/c mouse embryo DNA into the BamHI site of a λ Charon 28 (stock 2221) vector (7). Approximately 600,000 plaques were grown on E.coli LE392 and screened by the method of Benton and Davis (8) with a 292 bp BglII/HincII fragment from the coding region of a MuIFN- α_2 cDNA (4) which was nick-translated according to Ribgy, et al (9). The phage DNAs were hybridized to ^{32}P -labeled probe as described (4).

Purification of phage λ DNA, plasmid DNA, and restriction endonuclease mapping

After three rounds of plaque purification, DNA from positive plaques was isolated by the method of Williams and Blattner (10). Small scale preparations of phage λ DNA were used for restriction endonuclease analysis. HpaII restriction fragments from the genomic clones were ligated into the ClaI site of pBR322 with T4 DNA ligase (Bethesda Research Laboratories) and the subclone DNAs were purified by a modification of the procedure of Holmes and Quigley (11).

Restriction endonuclease maps were constructed by multiple digest analysis of phage and subclone DNAs. The restricted DNAs were electrophoresed through 1.0% horizontal agarose gels, transferred to nitrocellulose membranes as described by Southern (12) and hybridized to the MuIFN- α_2 BglII/HincII probe as described for the screening of the genomic library. The restriction digest map of one clone, pKK10, was generated by the partial digest method of Smith and Birnstiel (13).

DNA sequence analysis

5' end labeled fragments for sequencing were generated as follows. 10 μg of plasmid DNA were digested with the appropriate restriction enzyme and treated for 60 min. at 65°C with 200 units of bacterial alkaline phosphatase (Bethesda Research Laboratories) in 100 μl of 50mM Tris-HCl, pH8.5. The dephosphorylated DNA was purified by phenol/chloroform extraction followed by ethanol precipitation, and labeled with 10 units of T4 kinase (Bethesda Research Laboratories) according to Maxam and Gilbert (14). After purification by phenol/chloroform extraction and ethanol precipitation, the

labeled DNA was digested with a second restriction enzyme in order to obtain fragments labeled at only one 5' end and electrophoresed through 1.0% low melting point agarose (Bethesda Research Laboratories). The desired DNA fragments were visualized by ethidium bromide staining and purified by melting gel slices at 65°C, followed by phenol/chloroform extraction and ethanol precipitation. Sequencing reactions were performed according to Maxam and Gilbert (14). The reactions were electrophoresed through 8% and 20% polyacrylamide, 7M urea gels and exposed to Kodak XAR-5 film for 16-20 hr.

Southern analysis of BALB/c mouse DNA

BALB/c mouse DNA was extracted from spleen as described (15). 10 µg were restricted with 50 units of HindIII (Bethesda Research Laboratories) for 6 hrs, electrophoresed through 0.5% agarose gels and transferred to nitrocellulose membranes as described by Southern (12). Restriction fragments which were used as probes were isolated from 1.0% low melting point agarose gels (Bethesda Research Laboratories), and purified by melting gel slices at 65°C, extraction with phenol/chloroform and ethanol precipitation. The fragments were nick-translated according to Rigby, et al. (9) and hybridized to nitrocellulose membranes as described for the genomic library screening. Membranes were exposed to Kodak XAR-5 film for 3-7 days.

RESULTS

Identification and characterization of mouse α -IFN cluster

A library of BALB/c mouse embryo DNA in phage λ Charon 28 was screened with a 292 bp BglII/HincII fragment from the coding region of a MuIFN- α_2 cDNA (4). Nine positive plaques identified by this screening were plaque purified and characterized by restriction endonuclease analysis and Southern hybridization with the MuIFN- α_2 probe. HpaII restriction digest and hybridization patterns have shown that eight of the genomic clones contain different genomic inserts while one clone, λ MIF-9, is the same as λ MIF-1 (Fig. 1). Further restriction analysis of these clones with BamHI, HincII, BglII, PstI, PvuII, EcoRI, HindIII, and EcoRI/HindIII confirm-

Figure 1 A) *Hpa*II restriction digest analysis of positive phage DNAs from a genomic library screening. Lane 1, *Hind*III restricted λ DNA and *Hae*III digested ϕ X174 DNA markers; lane 2, λ MIF-1; lane 3, λ MIF-2; lane 4, λ MIF-3; lane 5, λ MIF-4; lane 6, λ MIF-5; lane 7, λ MIF-6; lane 8, λ MIF-7; lane 9, λ MIF-8 and lane 10, λ MIF-9. B) Southern blot analysis of the gel in A with the MuIFN- α_2 *Bgl*I/*Hinc*II fragment (4). The lanes are the same as lanes 2-10 in A. The location of *Hind*III restricted λ DNA and *Hae*III digested ϕ X174 DNA markers are indicated on the left, and are 23.7, 9.5, 6.7, 4.3, 2.3, 2.1, 1.353, 1.078, 0.878, 0.6 and 0.31 kb in length.

ed that λ MIF-1 and λ MIF-9 contain the same genomic DNA insert (data not shown). Since none of the identified human α -IFN genes contain internal *Hpa*II sites, we assumed that the same would be true of the murine α -IFN genes. Seven of the genomic clones identified contained more than one *Hpa*II fragment which is homologous to the MuIFN- α_2 probe (Fig. 1), suggesting that there are several α genes in each of these genomic segments. λ MIF-3 contains three *Hpa*II restriction fragments which hybridize weakly to the MuIFN- α_2 probe, and these probably represent one or more pseudogenes (16).

Restriction endonuclease analysis of λ MIF-1 with *Eco*RI, *Hind*III, and *Eco*RI/*Hind*III revealed that this genomic clone contains the MuIFN- α_1 gene which restriction map and sequence have been previously published (3). MuIFN- α_1 is present within the 9.5 kb *Hpa*II fragment of λ MIF-1; however, this genomic clone also has two additional *Hpa*II fragments

Figure 2 A) Restriction digest maps of a 28 kb genomic region of BALB/c mouse DNA and the two overlapping clones, λ MIF-1 and λ MIF-7, which were combined to give this map. The restriction maps of λ MIF-1 and λ MIF-7 were generated by multiple restriction digests of subclones of *Hpa*II fragments, which are indicated above each clone. MuIFN- α genes are indicated as filled arrows (▬); sequences homologous to a 972 bp intergenic *AccI*/*EcoRI* fragment (from pKK16) are shown as cross-hatched boxes. Wavy lines (~~~~) are λ Charon 28 vector sequences; the *Hpa*II sites at the 3' ends of λ MIF-1 and λ MIF-7 are also from the vector. The *Bam*HI site in MuIFN- α_5 (pKK8) and at the 3' end of λ MIF-1 are *Mbo*I cloning sites which regenerated *Bam*HI sites during construction of the genomic library. B) Restriction digest maps and sequencing strategies of *Hpa*II fragment subclones. Filled boxes are MuIFN- α genes and the small empty boxes on their 5' ends indicate the signal sequences of these genes. Cross-hatched boxes, are the 972 bp *AccI*/*EcoRI* fragment of pKK16 and a homologous region 5' to MuIFN- α_6 in pKK22. Wavy lines (~~~~) indicate λ Charon 28 vector sequences. Sequencing was performed on 5' end labeled fragments, indicated by filled circles (●), or in some cases by 3' labeling with terminal deoxynucleotidyl transferase, indicated by open circles (○). Arrows indicate direction and extent of sequencing.

which are homologous to the MuIFN- α_2 probe (Fig. 1). The *Hpa*II restriction fragments of λ MIF-1 were subcloned into the *Cla*I site of pBR322 and analyzed by restriction digest mapping and sequencing. In addition to the MuIFN- α_1 gene, λ MIF-1 contains a second gene in the 1.6 kb *Hpa*II fragment,

-720 -700 -680
 α5 GATGGTCCTAAAATCATCAGATAAATAGCTAACATACATTGAGGATTAAAA

-660 -640 -620
 α5 TAAATTGTATATAAATTTATATTAGACTTAATGAAGGCAACTCGAGAAACCAATTTTAGGACTTTTTATTTTTT

-600 -580 -560 -540
 α5 TACTTTTATCGAGACAGGGTTTCTCTGTATAGCTCTGGCTGCTCGAACACAGTTTGTACACCTTTACTTCAA-
 α6 GTATACCTTTACTTCAAT

-520 -500 -480 -460
 α4 CCCCCACACTTTACTTTTTTGACAGAAATATTTATGTAAATAGTATA
 α5 GTATATTAATTGAGAGACACTTGCAACAATACCTTGCATTA-TTATTT-AC-----ATACTT-TGTAAACAATACA
 α6 GTTTATTAATTGAGAGACACTTGAAACAATACCTTGCATTA-TTATTT-AC-----ATACTT-TGTAAACAATACA

-440 -420 -400 -380
 α1 AGACTATT-TTACAGA--AACTTTG-ATTTA
 α4 AATAAAACAAAAAGGTTGTAATAAAT-TACAAAAC--TT-TAAAAACAGTCATAGGGTAAATTTTGTGTTTG
 α5 AATAAAACAAAA--TTGTCAA-TCAATATAGTTAACGTTTACAGAAATATA-TTACAGG--AAATTTTG-ATTTA
 α6 AATAAAACAAAA--TTGTCAATTCAATATAGTTTAACTTACAAAATACA-TTACAGG--AAATTTG-ATTTA

-360 -340 -320
 α1 TACTAAATATGTAAATTTCAAGTAAGGACAGTTTATCAGTACTGTAACAGTTAATGAAAG-TT-----CAAATTA
 α4 CAGTAGATGTGCAGTTTTTCACACAGAGCTGTATATCTGTGGAGTAGTGGTTAATGACAAATTAAGGTAAATTA
 α5 CAATAAATATGTAAATTCACAAACAAAGACAGTTTATGGTGTCTGGAACAGTTAATGAAAG-TT-----CAAATTA
 α6 CAATAAATATGTAAATTTATAACAAGACAGTTTATGGAGCCGAAATAGTTATTGAAAG-TT-----CAAATTA

-300 -280 -260 -240
 α1 GGAAAAAGAAA--TGTAAGAAAGCTTT-TGATGAGGACCAAGTGAAGAGGAAGCAATAATGAAACCCACAATGGT
 α4 --AAGAGAAACCTCGGAGAGTAGCTTTCTGA-GGGCAGCAGTGAACCTGAAAGCAATGATTGAA--CCACATCC
 α5 GGAAGAAAAACCTTCAAAACCAAGCTTT-GGGTATGGAGCAGTGAAGAGGAAGCAATAATGGAATCTCAATGGG
 α6 GGAAGAAAAACCTTCAAAACCAAGCTTT-GGGTGTGAGCAGT-AAAGAGGAAGCAATAATGAAATCTCAATGGG

-220 -200 -180 -160
 α1 TTA-----GA--AAACACCCAGACGCAAGCAGAGAAATGAGTTAAAGAAAGTGAAGAACACAA
 α4 CCAGGGGGGAGGGGAGGGGAGGACAAAAT-CCACAGACACAAGCAGAGAGTGAAGTAAAGAAAGTGAAGAACACAA
 α5 TTA-----GA-AAAAATGCCCAAACAAGTGGAGAAATGAGTTAAAGAAAGTGAAGAACACAA
 α6 TTA-----GA-AAAAATGCCCAAACAAGTGGAGAAATGAGTTAAAGAAAGTGAAGAACACAA

-140 -120 -100 -80
 α1 GTGGAAAGTGATGGAAGGGCATTGAGAAAGTAAAAACCAAGTGTGTTG-CCCTATTTTAAGACACATTCACCCAGGAT
 α4 TTGGAAAGCAAGGGGAGGGTATTCGGAAGGAGAAACTTCTATTTGTCCTATTTTAAGAGAGATGTACACAGCA-
 α5 GTGGAAAGTGATGGAAGGGCATTGAGAAAGTAAAAACCAAGTGTGTTG-CCCTATTTTAAGACACATTCACCCAGGAT
 α6 GTGGAAAGTGATGGAAGGGTATTCGGAAGGAGAAACTAGTGTGTTG-CCCTATTTTAAGACACATTCACCCAGGAT

▼ -60 -40 -20
 α1 GGTCTTCAGAGAACCCTAGAGGGGAAGGATCAGGACCAAAACAGTCCAGAAAGACCAAGGCTTTGGCAACACTCACC
 α4 GGCTCTCAGAGAACCCTGAGGAGAAGACTCA-AACC-CACAGCCCAAGAGAGGCCAGCAGCATCTACAAGACCCACA-
 α5 GGTCTTCAGAGAACCCTAGAGGGGAAGGATCAGAAACCAACAGCCCAAGAAAGACCAAGCAGCATTTGGCAACACTCACC
 α6 GGCTCTCAGAGAACCCTAGAGGGGAAGGATCAAGATCAAAACAGCCCAAGAAAGACCAAGCAGCATTTGGCAACACTCACC

+1 +20 +40
 α1 ATG GCT AGG CTC TGT GCT TTC CTG ATG GTC CTG GCG GTG ATG AGC TAC --- TGG CCA
 α4 ATG GCT AGG CTC TGT GCT TTC CTC ATG ATC CTA GTA ATG ATC TAC TAC TGG TCA
 α5 ATG GCT AGG CTC TGT GCT TTC CTG ATG GTC CTG CCG GTG CTG AGC TAC --- TGG CCA
 α6 ATG GCT AGG CTC TGT GCT TTC CTG ATG GTC CTG GCG GTG CTG AGC TAC --- TGG CCA

S20 S10

+60 +80 +100
 α1 ACC TGC TCT CTA GGA TGT GAC CTG CCT CAG ACT CAT AAC CTC AGG AAC AAG AGA GCC
 α4 GCC TGT TCT CTA GGA TGT GAC CTG CCT CAC ACT TAT AAC CTC GGG AAC AAG AGG GCC
 α5 ACC TGC TCT CTA GGA TGT GAC CTT CCT CAG ACT CAT AAC CTC AGG AAC AAG AGA GCC
 α6 ACC TGC TCT CTA GGA TGT GAC CTG CCT CAG ACT CAT AAC CTC AGG AAC AAG AGA GCC

1 10

+120 +140 +160
 α1 TTG ACA CTC CTG GTA CAA ATG AGG AGA CTC TCC CCT CTC TCC TGC CTG AAG GAC AGG
 α4 TTG ACA CTC CTG GAA GAA ATG AGA AGA CTC CCC CCT CTT TCC TGC CTG AAG GAC AGG
 α5 TTA ACC CTC GTA AAA ATG AGG AGA CTC TCC CCT CTC TCC TGC CTG AAG GAC AGA
 α6 TTA ACC CTC CTG GTA AAA ATG AGG AGA CTC TCC CCT CTC TCC TGC CTG AAG GAC AGA

20 30

+180 +200 +220
 α1 AAG GAC TTT GGA TTC CCG CAG GAG AAG GTG GAT GCC CAG CAG ATC AAG AAG GCT CAA
 α4 AAG GAT TTT GGA TTC CCC TTG GAG AAG GTG GAT AAC CAA CAG ATC CAG AAG GCT CAA
 α5 AAG GAC TTT GGA TTC CCA CAG GAG AAG GTG GGT GCC CAG CAG ATC CAG AAG GCT CAA
 α6 AAG GAC TTT GGA TTC CCA CAG GAG AAG GTG GGT GCC CAG CAG ATC CAG GAG GCT CAA

40 50

[illegible]

Figure 3 Comparison of MuIFN- α nucleotide sequences, which were determined by the strategy shown in Fig. 2B. A variant of the Goldberg-Hogness sequence (TATTTAA), the start and stop codons, and the polyadenylation signal (AATAAA) are boxed. An arrowhead at position -67 indicates the 5' end of the MuIFN- α messages. Insertions and deletions have been added to allow maximum homologies.

HpaII restriction fragments from the remaining seven genomic clones were also subcloned into the ClaI site of pBR322 and analyzed by restriction mapping. The 3.0 kb HpaII fragment (pKK15) from λ MIF-7 had a similar restriction map as the partial gene, MuIFN- α_5 , in λ MIF-1. In addition, λ MIF-7 contains a second complete gene, MuIFN- α_6 (Fig. 2A). Sequence analysis revealed that pKK15 contains the entire MuIFN- α_5 gene and that the two genomic clones, λ MIF-1 and λ MIF-7, have a small region of common sequences within this gene. The restriction maps of λ MIF-1 and λ MIF-7 were generated by multiple restriction enzyme digests of both genomic clone and subclone DNAs, and these maps were combined at the overlapping region within MuIFN- α_5 to yield a 28 kb segment of mouse genomic DNA containing four complete MuIFN- α genes (Fig. 2A).

813

were sequenced according to the strategies outlined in Fig. 2B, and the nucleotide sequences are compared in Fig. 3. A total of 1503, 1591, 2016, and 1748 basepairs (bp) were sequenced for the α_1 , $-\alpha_4$, $-\alpha_5$, and $-\alpha_6$ genes, respectively. The sequence of the MuIFN- α_1 gene is exactly the same as published previously except for a single T at position -269 which is a C in the published MuIFN- α_1 sequence (3). We have also sequenced an additional 118 bp from the 5' and 192 bp from the 3' untranslated regions of the MuIFN- α_1 . There are no introns in any of the four MuIFN- α described here, as is the case with other cloned murine and human type I IFN genes.

All four genes contain a variant of the Goldberg-Hogness sequence, TATTTAA, at position -94 which is the same in most of the human IFN- α genes (17,18,19). All of the MuIFN- α genes show a high degree of homology within the first 200 bp upstream of the start codon; however, the α_4 gene shows some nucleotide variations which are not present in the other promoter regions. In addition, the α_4 gene contains a 19 bp insertion consisting of purines (16 of the 19 bases are G's), at position -203 to -222. The MuIFN- α_4 , $-\alpha_5$ and $-\alpha_6$ genes show a high degree of homology in their 3' untranslated regions; however, the MuIFN- α_4 gene has an insertion of 21 T's at positions +1260 to +1281. These three α -IFN genes have a polyadenylation signal (AATAAA) located 333, 316, and 308 bp downstream of their stop codons, respectively. The 3' untranslated region of MuIFN- α_1 has little homology to the other identified α -genes, and no polyadenylation signal for MuIFN- α_1 was found in the previously published sequence (3) within 380 bp downstream of its stop codon; we have sequenced an additional 192 bp from the 3' untranslated region of this gene, and no polyadenylation signals of the AATAAA (20) or ATTAAA type (21,22) were found within the 572 bp of 3' non-coding sequences.

The MuIFN- α_4 gene has a 15 bp deletion in the coding region at positions +379 to +393, thus the mature protein contains a 5 amino acid internal deletion; the α_4 gene also contains an additional amino acid both in the signal sequence and at the carboxyl terminus of the mature protein

	S20										S10										1									
Mua1	M	A	R	L	C	A	F	L	M	V	L	A	V	M	S	Y	-	W	P	T	C	S	L	G	C	D	L	P		
Mua4	M	A	R	L	C	A	F	L	M	I	L	V	M	M	S	Y	Y	W	S	A	C	S	L	G	C	D	L	P		
Mua5	M	A	R	L	C	A	F	L	M	V	L	P	V	L	S	Y	-	W	P	T	C	S	L	G	C	D	L	P		
Mua6	<u>M</u>	<u>A</u>	<u>R</u>	<u>L</u>	<u>C</u>	<u>A</u>	<u>F</u>	<u>L</u>	<u>M</u>	<u>V</u>	<u>L</u>	<u>A</u>	<u>V</u>	<u>L</u>	<u>S</u>	<u>Y</u>	-	<u>W</u>	<u>P</u>	<u>T</u>	<u>C</u>	<u>S</u>	<u>L</u>	<u>G</u>	<u>C</u>	<u>D</u>	<u>L</u>	<u>P</u>		
	10										20										30									
Mua1	Q	T	H	N	L	R	N	K	R	A	L	T	L	L	V	Q	M	R	R	L	S	P	L	S	C	L	K	D		
Mua4	H	T	Y	N	L	G	N	K	R	A	L	T	V	L	E	E	M	R	R	L	P	P	L	S	C	L	K	D		
Mua5	Q	T	H	N	L	R	N	K	R	A	L	T	L	L	V	K	M	R	R	L	S	P	L	S	C	L	K	D		
Mua6	<u>Q</u>	<u>T</u>	<u>H</u>	<u>N</u>	<u>L</u>	<u>R</u>	<u>N</u>	<u>K</u>	<u>R</u>	<u>A</u>	<u>L</u>	<u>T</u>	<u>L</u>	<u>L</u>	<u>V</u>	<u>K</u>	<u>M</u>	<u>R</u>	<u>R</u>	<u>L</u>	<u>S</u>	<u>P</u>	<u>L</u>	<u>S</u>	<u>C</u>	<u>L</u>	<u>K</u>	<u>D</u>		
	40										50										60									
Mua1	R	K	D	F	G	F	P	Q	E	K	V	D	A	Q	Q	I	K	K	A	Q	A	I	P	V	L	S	E	L		
Mua4	R	K	D	F	G	F	P	L	E	K	V	D	N	Q	Q	I	Q	K	A	Q	A	I	L	V	L	R	D	L		
Mua5	R	K	D	F	G	F	P	Q	E	K	V	G	A	Q	Q	I	Q	E	A	Q	A	I	P	V	L	S	E	L		
Mua6	<u>R</u>	<u>K</u>	<u>D</u>	<u>F</u>	<u>G</u>	<u>F</u>	<u>P</u>	<u>Q</u>	<u>E</u>	<u>K</u>	<u>V</u>	<u>G</u>	<u>A</u>	<u>Q</u>	<u>Q</u>	<u>I</u>	<u>Q</u>	<u>E</u>	<u>A</u>	<u>Q</u>	<u>A</u>	<u>I</u>	<u>P</u>	<u>V</u>	<u>L</u>	<u>T</u>	<u>E</u>	<u>L</u>		
	70										80																			
Mua1	T	Q	Q	I	L	N	I	F	T	S	K	D	S	S	A	A	W	N	A	T	L	L	D	S	F	C	N	D		
Mua4	T	Q	Q	I	L	N	L	F	T	S	K	D	L	S	A	T	W	N	A	T	L	L	D	S	F	C	N	D		
Mua5	T	Q	Q	V	L	N	I	F	T	S	K	D	S	S	A	A	W	N	A	T	L	L	D	S	F	C	N	D		
Mua6	<u>T</u>	<u>Q</u>	<u>Q</u>	<u>I</u>	<u>L</u>	<u>T</u>	<u>L</u>	<u>F</u>	<u>T</u>	<u>S</u>	<u>K</u>	<u>D</u>	<u>S</u>	<u>S</u>	<u>A</u>	<u>A</u>	<u>W</u>	<u>N</u>	<u>A</u>	<u>T</u>	<u>L</u>	<u>L</u>	<u>D</u>	<u>S</u>	<u>F</u>	<u>C</u>	<u>N</u>	<u>D</u>		
	90										100										110									
Mua1	L	H	Q	Q	L	N	D	L	Q	G	C	L	M	Q	Q	V	G	V	Q	E	F	P	L	T	Q	E	D	A		
Mua4	L	H	Q	Q	L	N	D	L	K	A	C	V	M	Q	-	-	-	-	-	E	P	P	L	T	Q	E	D	S		
Mua5	V	H	Q	Q	L	N	D	L	K	A	C	V	M	Q	Q	V	G	V	Q	E	S	P	L	T	Q	E	D	S		
Mua6	<u>L</u>	<u>H</u>	<u>Q</u>	<u>L</u>	<u>L</u>	<u>N</u>	<u>D</u>	<u>L</u>	<u>Q</u>	<u>G</u>	<u>C</u>	<u>L</u>	<u>M</u>	<u>Q</u>	<u>Q</u>	<u>V</u>	<u>E</u>	<u>I</u>	<u>Q</u>	<u>A</u>	<u>L</u>	<u>P</u>	<u>L</u>	<u>T</u>	<u>Q</u>	<u>E</u>	<u>D</u>	<u>S</u>		
	120										130										140									
Mua1	L	L	A	V	R	K	Y	F	H	R	I	T	V	Y	L	R	E	K	K	H	S	P	C	A	W	E	V	V		
Mua4	L	L	A	V	R	T	Y	F	H	R	I	T	V	Y	L	R	E	K	K	K	H	S	L	C	A	W	E	V	I	
Mua5	L	L	A	V	R	K	Y	F	H	R	I	T	V	Y	L	R	E	K	K	H	S	P	C	A	W	E	V	V		
Mua6	<u>L</u>	<u>L</u>	<u>A</u>	<u>V</u>	<u>R</u>	<u>T</u>	<u>Y</u>	<u>F</u>	<u>H</u>	<u>R</u>	<u>I</u>	<u>T</u>	<u>V</u>	<u>Y</u>	<u>L</u>	<u>R</u>	<u>E</u>	<u>K</u>	<u>K</u>	<u>H</u>	<u>S</u>	<u>P</u>	<u>C</u>	<u>A</u>	<u>W</u>	<u>E</u>	<u>V</u>	<u>V</u>		
	150										160																			
Mua1	R	A	E	V	W	R	A	L	S	S	S	A	N	V	L	G	R	L	R	E	E	K								
Mua4	R	A	E	V	W	R	A	L	S	S	S	T	N	L	L	A	R	L	S	K	E	K	E							
Mua5	R	A	E	V	W	R	A	L	S	S	S	V	N	L	L	A	R	L	S	K	E	E								
Mua6	<u>R</u>	<u>A</u>	<u>E</u>	<u>V</u>	<u>W</u>	<u>R</u>	<u>A</u>	<u>L</u>	<u>S</u>	<u>S</u>	<u>S</u>	<u>A</u>	<u>K</u>	<u>L</u>	<u>L</u>	<u>A</u>	<u>R</u>	<u>L</u>	<u>N</u>	<u>E</u>	<u>D</u>	<u>E</u>								

Figure 4 Comparison of the MuIFN- α amino acid sequences. Amino acids are written in one letter notation (36) and common residues are underlined. The glycosylation site Asn-Ala-Thr is boxed.

(Fig. 4). The additional amino acid at the carboxyl end has also been found in cDNA clones of the MuIFN- α_2 gene, which is expressed in mouse cells (3,4). MuIFN- α_1 , - α_5 , and - α_6 all encode mature proteins that are 166 amino acids long (Fig. 4) which is the same length as the HuIFN- α 's.

One major difference between the human and mouse α -IFNs is that the mouse proteins are glycosylated whereas the human are not (23,24,25,26). All four of the MuIFN- α genes in the 28 kb genomic segment contain a single Asn-Ala-Thr glycosylation site at positions 78-80 (Fig. 4).

Identification of repetitive intergenic region

In an attempt to identify unique fragments that could be used as probes for extending the 28 kb region, nick-translated BALB/c mouse DNA was hybridized to HpaII digests of the eight genomic clones and to an MspI/AccI/EcoRI digest of subclone pKK16, which contained the MuIFN- α_1 gene at the 3' end of the cluster (Fig. 2A). A 972 bp AccI/EcoRI and an 850 bp AccI fragment from the α_1 subclone (Fig. 2B) were identified which gave very weak signals with the genomic DNA probe (data not shown). When BALB/c genomic DNA restricted with HindIII was analyzed by Southern hybridization, the 850 bp AccI fragment hybridized to a 2.4 kb fragment, while the 972 bp AccI/EcoRI fragment hybridized to five fragments which were 17, 9.5, 5.8, 2.7 and 2.4 kb in length (Fig. 5). The 17, 9.5, 5.8 and 2.7 kb HindIII fragments of BALB/c genomic DNA also hybridized to a 1040 bp EcoRI/AccI fragment containing the coding region of MuIFN- α_1 and to the BglII/HincII fragment of the coding region of MuIFN- α_2 (Fig. 5), indicating that this repeated fragment is linked to some, but not all of the MuIFN- α genes. Sequences homologous to the 972 bp AccI/EcoRI fragment were also found between α_5 and α_6 , and 5' to the α_6 gene, but not between α_4 and α_5 . The intergenic regions between α_1 and α_4 , and α_5 and α_6 (7.5 and 6.8 kb, respectively) are larger than the distance between α_4 and α_5 (6kb) suggesting that this repeated sequence may have been deleted from this area. The 17 kb HindIII fragment from BALB/c DNA (Fig. 5) contains the α_5 and α_6 genes as well as the repeated sequence between these genes and part of the repeated sequence upstream of α_6 (Fig. 2A). The 5.8 kb HindIII fragment contains the α_4 gene and part of the repeated sequence between α_4 and α_1 , and the 2.4 kb HindIII fragment is a part of the 972 bp AccI/EcoRI sequence upstream of the α_1 gene (Fig. 2A). One of the remaining two HindIII fragments

Figure 5 Southern blot analysis of 10 μ g of HindIII digested BALB/c mouse DNA with (lane 1) the MuIFN- α_2 BglIII/HincII coding region probe, (lane 2) a 1040 bp EcoRI/AccI probe containing the MuIFN- α_1 gene, (lane 3) a 972bp AccI/EcoRI probe from pKK16 (see Figs. 2A and B) and (lane 4) an 850 bp AccI probe from the 3' untranslated region of MuIFN- α_1 (pKK-16). Sizes of HindIII restricted λ DNA markers are indicated in kb on the left. The sizes of the HindIII fragments which hybridize to the 972 bp intragenic AccI/EcoRI fragment (lane 3) are shown on the right.

(2.7 and 9.5 kb) contains the repeated sequence upstream of MuIFN- α_6 . The 972 bp AccI/EcoRI fragment from the 5' flanking region of MuIFN- α_1 and a homologous 822 bp region from the 5'upstream region of MuIFN- α_6 were sequenced according to the strategy in Fig. 2B, and the sequences are compared in Fig. 6. There is a high degree of DNA homology

```

      10      20      30      40      50      60      70
A  GTAGACTTTGTGAGGATTTTAATTATTGGATAAATAGCTATTTGTGTGACTACTAAACATGTTTAAGATATGAC

      80      90      100     110     120     130     140     150
A  TTCTGTACTCTTTCTTTGGTGGTTATTTTAAATGTAACCTTTCATGGTTAAATAAAAAAAATGAAATAGC

      160     170     180     190     200     210     220
A  TCAAGATATCAACATGCTCAGCAAAACATGAAGTTTCCACACCTTTGATTCAATAACCTTTTCTTATTAACAAA
B  CTTTCACCTCGTCTTTATTAATTAATAGCTCACTCTTAAATAATAA

      230     240     250     260     270     280     290     300
A  TTTGGGAATCAAGAGAAGA--AATGGAAGGACTGAGATAGGAGAATGCAGTAGGTGAGC-----TAT
B  CTGTAGAAATAAAGACAGAGAAATTAAGGACTGAGACAGGAGAATGCAGTAGGTGAGAACTACTGGACTATAT

      310     320     330     340     350     360     370     380
A  CATTCTACACTCATAAAGCAACCTTGATTTAATGCACACCCAGGTTACGTAGAAAGGATATTAGAGTCAGGGAGC
B  CATTCCTGCTCCATTAAGCAGAGCTTGATTCATCGACACTCAGTTTATATAAAAGTACCTTAGAATCGATAAGC

      390     400     410     420     430     440     450     460
A  ATCACACAGCTGTCCATGAAGACACATCTGTGTACAGTAACACCATGTAAGCTGGAGACTGCCAAGTTTGTCAAG
B  ATCACACAGC-----ATGAAGACACATCTGTGTACAGTAACACTGTGTGTGCTGGAGACTGACAGGII-GTCAAG

      470     480     490     500     510     520     530
A  CTTCAGGATTGAGAGTGTGAACATGGCCTTATCCAGGTCTCTGTGTCTGCTATTGCTGAGCTGCAGTGACCCAGTA
B  CTTCAGGACTGAGAGTG--AACATGGCCATATCCAGGTTCCTGTATCTGCTATCGCTGAGGCTGGAGTGAATCAGTA

      540     550     560     570     580     590     600     610
A  CCCACTGACATAGGGCAGAACACAGTAACATCTTAACATAAGGAGGAAGTTACCCCAATGATCTGGCCAACATC
B  CCCACTGATGTAAGCAGGACAGAGAAAC-----TAAGGAGGGAAGTCCACCG--AATGATCTGGCCCAACATC

      620     630     640     650     660     670     680
A  CTAGCTTAATCAAGAAAGCTTCAAGGCCAGTGTGAGGCACCTGTGGGGAAGGTATCCAGAAAGAGAATGGTAGAGTG
B  CTAACATAAICA-GTAGCTTCAAGGCCAGTGTGAGGCACCTGTGGGGAAGGTATCCAGAAAGAGAATGGTAGAGTG

      690     700     710     720     730     740     750     760
A  TACGTGACACCCATCTCTGTTTTCT-----ACATGCGCAGTTTCAGCACATCTATAGCCCTTT
B  TACCTGACACCCATCTCTGTTTTCATACACACATGCACATGCATGACAGGCAGGACAGGTGAAGTCTII

      770     780     790     800     810     820     830
A  GCACCTTGCCAATGAGTTGAAGCCTATTCACCACATCTCTTTATCAGATTAGTGATCAAGTCATATATCAAG
B  GCACCTTGCCAGGTGGTGAAGCCTATTCACCACATCTCTTTATCAGATTAGTGATCAAGTCATATATCAAG

      840     850     860     870     880     890     900     910
A  GTTCTTGATGCTAGGTATTTTCAAGTCTTTTTCAGGATGC---TGGATCTATTTTCATTCTTTAGAAAGCAACCATC
B  GACTTGATGCTAGGTCAGTTTGAATGTTAAACATGATTATGAATCCACCTTTTATTTTII-GAGGAAACCCAC

      920     930     940     950     960     970     980
A  TAATTAGATGAGCACCATTTCATT---TGGTCTCTTATTTCCATTGTGTGCTTTTATATTCCTTCTTAAAAATA
B  TAATCATTGAGCACCATTTCATTGAAGACGATCACCCTATTTCTGTGTGAACCTTTTGTATCTTCTTAAAAATG

      990     1000    1010
A  AGATTCTCAGTAAGTGAAGGAATTTC
B  AAA-TCTCAGTAAGTGAAGAAATTTATGGCTGGG

```

Figure 6 Comparison of the nucleotide sequences of the 972 bp AccI/EcoRI intergenic fragment from PKK16(A) and its homologous region from PKK22(B). Underlined nucleotides are common to both sequences. Insertions and deletions were introduced into the sequence alignments to achieve maximum homology.

(80%) between these two regions, although several insertions and deletions are seen. The location of this repeated fragment between the MuIFN- α genes on this 28 kb cluster and the homology between these areas indicates that this cluster probably evolved through tandem duplication of one of the MuIFN- α genes.

The unique 850 bp AccI fragment which was identified in

the 3' untranslated region of MuIFN- α_1 (Fig. 2B) was used to screen the eight purified genomic clones as well as an additional 16 positive plaques from a second screening of the genomic library. Only λ MIF-1, which contains the MuIFN- α_1 gene, was positive in this screening, indicating that there are no additional MuIFN- α genes closely linked to MuIFN- α_1 on its 3' end (at least within the 16-20 kb fragments of genomic DNA which were used to construct the library). The 972 bp AccI/EcoRI repeated sequence is associated with at least 6 of the MuIFN- α genes and of these, two have not yet been identified. Although no α genes seem to be localized immediately downstream from α_1 , there is some evidence that two more α genes are linked to the 5' end of the 28 kb segment examined in this study.

DISCUSSION

Four MuIFN- α genes have been identified within a 28 kb region of mouse chromosomal DNA (Fig. 2A). The cloning of a genomic segment containing one of these genes, MuIFN- α_1 , has been previously reported (2). Three of the genes; MuIFN- α_1 , - α_5 and - α_6 are 93-95% homologous at the nucleotide level, whereas MuIFN- α_4 has approximately 88% homology with these three genes. The coding region of MuIFN- α_4 shows a number of single nucleotide alterations from the other three α genes, and these differences are reflected in the amino acid sequence (Fig. 4). MuIFN- α_1 , - α_5 and - α_6 are 90-95% homologous at the amino acid level, while the MuIFN- α_4 protein shows lower homology to the other three α peptides.

An intergenic fragment between MuIFN- α_1 and MuIFN- α_4 was identified, which is also present between MuIFN- α_5 and MuIFN- α_6 , and upstream of MuIFN- α_6 (Fig. 2A). The DNA sequence of this fragment contains an open reading frame, although no homology to any other known DNA or protein sequence has been found. The question whether this sequence is involved in the regulation of the α interferon gene expression remains to be answered. However, the presence of the homologous sequence between three α genes showing a high degree of homology indicates that this gene cluster may have

evolved through tandem duplications of a MuIFN- α gene and the intergenic regions (Fig. 5). MuIFN- α_4 has lower homology to the other three genes and does not have the repeated intergenic region associated with its 5' end region. In addition, the MuIFN- α_4 gene is flanked by sequences not present in the other α -IFN genes (a G rich cluster in the 5' upstream region and a T rich cluster in the 3' untranslated region); the sequence differences both in the coding and the flanking region indicate that the α_4 gene may have been inserted into a previously existing MuIFN- α cluster. The area of the polyadenylation signals for MuIFN- α_4 , - α_5 and - α_6 and the 3' end flanking regions of these genes have little obvious homology with the 3' end untranslated sequences of the α_1 gene (Fig. 3). A unique 850 bp AccI fragment from the 3' end region of MuIFN- α_1 did not hybridize to any of 23 positive phage λ plaques from a screening of a genomic library with the MuIFN- α_2 probe. This evidence indicates that there are probably no MuIFN- α genes closely linked to the 3' end of the cluster which has been identified and characterized. MuIFN- α_1 may be the last gene in this tandemly duplicated region and the lack of apparent homology in the 3' untranslated region (including the lack of a polyadenylation signal) could be explained by recombinational events during unequal crossing over in which these sequences were lost. From the results of genomic blots using a 972 bp AccI/EcoRI probe from the intergenic region between α_1 and α_4 , however, it is obvious that there is at least one more gene (probably two) which is linked to the 5' end of this cluster, immediately upstream of MuIFN- α_6 . If there are 10-15 MuIFN- α genes, then nearly half of them are present in this tandemly duplicated region.

MuIFN- α 's and HuIFN- α 's exist as multigene families (25,27,28) and similar clusters of HuIFN- α genes have been identified (16,29). The presence of similar tandem duplications in both species implies that the duplication events which gave rise to the multigene IFN- α family may have occurred before the divergence of the human and mouse lines or as suggested by Shaw, et al. (3), that a large IFN- α gene

family may have existed with some members giving rise to the MuIFN- α 's and others to the HuIFN- α 's when these lines diverged. It is already known that the HuIFN- α gene family and HuIFN- β are on chromosome 9 (30,31,32,33), and that the MuIFN- α gene family and MuIFN- β are both localized on chromosome 4 (4, and Kelley and Pitha, in preparation). Restriction fragment length polymorphisms between BALB/c and C57BL/6 mice were recently identified and used to link the MuIFN- α locus to histocompatibility locus, H-15, on chromosome 4 (5). However, none of the MuIFN- α genes identified in the present study represent the polymorphic genes used to link the IFN- α locus to H-15. Regions of homology between human chromosome 9p and mouse chromosome 4 have been determined. This region contains the Aco and Galt loci both in mouse and human (34), and the murine IFN- α locus has been localized to this region also (35), implicating that the human IFN- α locus may be in this region as well. As more becomes known about the MuIFNs, it will be interesting to see if there are any obvious connections at the nucleotide level between the genomic organizations of the mouse and human type I interferons.

ACKNOWLEDGEMENTS

This study is dedicated to the memory of Dr. Wallace P. Rowe who brought our attention to the mouse system. This work represents partial fulfillment of doctoral requirements for K.A.K. and was supported by NIH grant 50I-AI-19737-02.

*To whom correspondence should be addressed

REFERENCES

1. Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. and Weissmann, C. (1980) *Nature* 285, 547-549.
2. Higashi, Y., Sokawa, Y., Watanabe, Y., Kawade, Y., Ohno, S., Takaoda, C. and Taniguchi, T. (1983) *J. Biol. Chem.* 258, 9522-9529.
3. Shaw, G. D., Boll, W., Taira, H., Mantei, N., Lengyel, P. and Weissmann, C. (1983) *Nuc. Acids Res.* 11, 555-573.
4. Kelley, K. A., Kozak, C. A., Dandoy, F., Sor, F., Skup, D., Windass, J. D., DeMaeyer-Guignard, J., Pitha, P. M. and DeMaeyer, E. (1983) *Gene* 26, 181-188.

5. Dandoy, F., Kelley, K. A., DeMaeyer-Guignard, J., DeMaeyer, E. and Pitha, P. M. (1984) *J. Exp. Med.* 160, 294-302.
6. Skup, D., Windass, J. O., Sor, F., George, H., Williams, B. R. G., Fukuhara, H., DeMaeyer-Guignard, J. and DeMaeyer, E. (1982) *Nuc. Acids Res.* 10, 3069-3084.
7. Rimm, D. L., Horness, D., Kucera, J. and Blattner, F. R. (1980) *Gene* 12, 301-309.
8. Benton, W. D. and Davis, R. W. (1977) *Science* 196, 180-182.
9. Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* 113, 237-251.
10. Williams, B. G., Blattner, F. R. (1979) *J. Virol.* 29, 555-575.
11. Holmes, D. S. and Quigley, M. (1981) *Anal. Biochem.* 114, 193-197.
12. Southern, E. M. (1975) *J. Mol. Biol.* 98, 503-517.
13. Smith, H. O. and Birnstiel, M. L. (1976) *Nuc. Acids Res.* 3, 2387-2398.
14. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* 65, 499-560.
15. Heisterkamp, N., Groffen, J., Stephenson, J. R., Spurr, N. K., Goodfellow, P. N., Solomon, E., Casritt, B. and Bodmer, W. F. (1982) *Nature* 299, 747-749.
16. Weissmann, C., Nagata, S., Boll, W., Fountoulakis, M., Fujisawa, A., Fujisawa, J.-I., Haynes, J., Henco, K., Mantei, N., Ragg, H., Schein, C., Schmid, J., Shaw, G., Streuli, M., Taira, H., Todokoro, K. and Weidle, U. (1982) In: *Interferons* (T.C. Merigan and R. M. Friedman, eds.) Academic Press, p.295-326.
17. Nagata, S., Mantei, N. and Weissmann, C. (1980) *Nature* 287, 401-408.
18. Lawn, R. M., Adelman, J., Dull, T. J., Gross, M., Goeddel, D. V. and Ullrich, A. (1981) *Science* 212, 1159-1162.
19. Lawn, R. M., Gross, M., Houck, C. H., Franke, A. E., Gray, P. W. and Goeddel, D. V. (1981) *Proc. Natl. Acad. Sci. USA* 78, 5435-5439.
20. Proudfoot, N. J. and Brownlee, G. G. (1976) *Nature* 263, 211-214.
21. Hagenbuchle, O., Bovey, R. and Young, R. A. (1980) *Cell* 21, 179-187.
22. Nunberg, J. H., Kaufman, R. J., Chang, A. C. Y., Cohen, S. N. and Schimke, R. T. (1980) *Cell* 19, 355-364.
23. Knight, E., Jr. (1976) *Proc. Natl. Acad. Sci. USA* 73, 520-523.
24. Havel, E. A., Yamazaki, S. and Vilcek, J. (1977) *J. Biol. Chem.* 252, 4425-4427.
25. Tan, Y. H., Baraket, F., Berthold, W., Smith-Johannsen, H. and Tan, C. (1979) *J. Biol. Chem.* 254, 8067-8073.
26. Taira, H., Broeze, R., Jayaram, B. M., Lengyel, P., Hunkapiller, M. W. and Hood, L. E. (1980) *Science* 207, 528-530.
27. Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A., Yelverton, E. and Gray, P. W. (1981) *Nature* 290, 20-26.

-
28. Brack, C., Nagata, S., Mantei, N. and Weissmann, C. (1981) *Gene* 15, 379-394.
 29. Ullrich, A., Gray, A., Goeddel, D. V. and Dull T. J. (1982) *J. Mol. Biol.* 156, 467-486.
 30. Owerbach, D., Rutter, W. J., Shows, T. B., Gray, P., Goeddel, D. V. and Lawn, R. M. (1981) *Proc. Natl. Acad. Sci. USA* 78, 3123-3127.
 31. Meager, A., Graves, H. E., Burke, D. C. and Swallow, D. (1979) *Nature* 280, 493-494.
 32. Pitha, P. M., Slate, D. L., Raj, N. B. K. and Ruddle, F. H. (1982) *Mol. Cell. Biol.* 2, 564-570.
 33. Trent, J. M., Olson, S. and Lawn, R. M. (1982) *Proc. Natl. Acad. Sci. USA* 79, 7809-7813.
 34. Nadeau, J. H. and Eichner, E. M. (1982) *Cytogenet. Cell Genet.* 34, 271-281.
 35. D'Eustachio, P. and Ruddle, F. (1983) *Science* 220, 919-924.
 36. Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, National Biomedical Res. Found., Washington.