Volume 13 Number 3 1985 Nucleic Acids Research

Organization, structure and expression of murine interferon alpha genes

Ellen C.Zwarthoff, Arno T.A.Mooren and Jan Trapman

Department of Pathology, Erasmus University, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Received 12 November 1984; Accepted 9 January 1985

ABSTRACT

Using a human interferon-α probe we have isolated recombinant phages containing murine interferon-alpha (Mu IFN-α) genes from a genomic library. One of these phages contained two complete Mu IFN-α genes and part of a third gene. The insert of a second phage held two IFN genes. This indicates that the Mu IFN-α genes are clustered in the genome as is the case for the analogous human genes. The nucleotide sequences of these 5 genes were determined. They show that the genes are all different, albeit highly homologous. The deduced amino acid sequences show that four of the five genes contain a putative glycosylation site. Three genes were transiently expressed in COS cells and they gave rise to protein products showing antiviral properties.

The expression of the five Mu IFN-α genes and the Mu IFN-β gene was studied in virus-induced mouse L cells. The individual mRNAs were visualized in a nuclease S1 experiment, using a specific probe for each gene. In RNA preparations from induced cells mRNAs for each of the five α genes and the β gene were present. However, substantial differences in the amounts of the individual mRNAs were observed.

INTRODUCTION

Interferons (IFNs) are proteins with numerous biological properties (for reviews, see refs. 1 and 2). IFNs are secreted by the producing cells and can, after binding to surface receptors on recipient cells, induce in the latter a variety of responses. First they were discovered for their ability to induce an antiviral state in the recipient cells, later they were found also to inhibit cell proliferation and to affect immunological processes. IFNs have also been shown to have an antitumor effect in certain human and animal neoplasias (3,4). Interferons generally are highly species specific. Thus, Mu IFNs hardly show any antiviral or growth inhibitory activity on human cells and vice versa. When regarded in the molecular perspective, IFNs are able to, selectively, affect the expression of several genes. For instance they can induce or increase the expression of the 2-5 A synthetase gene (5) and of genes of the major histocompatibility complex (6).

IFNs can be divided into 3 antigenically distinct classes known as IFN-α,

-β and -γ. IFN-α and β (also known as Type I IFNs) are produced after induction of cells with a virus or a double-stranded polyribonucleotide (1). IFN-α comprises a group of closely related proteins, which is reflected genetically in the IFN-α gene family, consisting of at least 14 members (7). IFN-β presumably represents one protein which is the product of a single gene (8,9,10). Between species the IFN-α genes seem to be highly conserved during evolution , this is much less the case with the IFN-β genes (11). Both the IFN-α and -β genes do not contain introns. They are located on chromosome 9 in the human genome (11) and on chromosome 4 in the mouse (12,13,14). IFN-γ (also known as Type II IFN) is secreted by lymphocytes after antigenic or mitogenic stimulation (1). The IFN-γ protein is encoded by a single gene, which is interrupted by 3 introns (15,16,17). The IFN-γ gene has been located on human chromosome 12 (18,19) and its murine equivalent on chromosome 10 (20).

Most human (Hu) IFN genes have been molecularly cloned and large quantities of the proteins have been obtained after expression of the genes in bacteria (21). Because many aspects of the interferon system cannot be evaluated in man studies on non-human IFNs are urgently needed. The mouse offers an excellent model in view of the broad knowledge that already excists on its genetics. For this reason we found it of importance to investigate the properties of the Mu IFN gene family. So far Mu IFN-β (10) and -γ cDNA clones (17), 2 Mu IFN-α genes (22,23) and a Mu IFN-α cDNA clone (22) were isolated and characterized. In this study we describe the molecular cloning of 2 clusters of Mu IFN-α genes, the properties of the genes located in these clusters and their expression.

MATERIALS AND METHODS

Isolation of Mu IFN-α genes. For the isolation of Mu IFN-α genes we screened a Balb/c genomic library with a Hu IFN-α probe. The library was provided by L. Hood (California Institute of Technology), and was made by partial digestion of mouse DNA with Hae III and Alu I, addition of Eco RI linkers to size-selected (15-20 kb) fragments and ligation of Charon 4A arms. The Hu IFN-α probe was a 55 nucleotides long synthetic DNA fragment homologous to a highly conserved part in the Hu IFN-α genes, and was provided by P. Boseley and A. Easton (Searle, High Wycombe, UK). The library was screened essentially as described by Benton and Davies (24). Hybridization was performed at 55°C in 5 x SSC, because of the shortness and the heterologous nature of the probe. Phage DNA was isolated according to

standard procedures (25). Restriction enzyme sites were mapped by analysis of fragments generated by single, double and triple digestions on 1% agarose gels, Southern blotting and hybridization. All enzymes were from Boehringer Mannheim.

Nucleotide sequence analysis. Eco R1 fragments hybridizing to the Hu IFN-α probe were cloned into the Eco R1 site of pBR328 and a detailed map of these inserts was made to determine the exact localization of the genes. Suitable restriction enzyme fragments were chosen from the physical maps of the pBR328 subclones. These fragments were isolated and cloned into the M13 vectors mp8 and mp9 (26). The nucleotide sequences of the M13 inserts were determined using the chain termination method described by Sanger et al. (27).

RNA isolation and S1 analysis. Total RNA was isolated from monolayers by the guanidinium isothiocyanate/cesium chloride method (28). When cells were induced for interferon expression, 100 HA (haemaglutination) units Sendai virus (Flow laboratories) were added per ml culture medium during 1.5 h and RNA was isolated 7 h later. Expression of the individual IFN mRNAs was analyzed in a nuclease S1 experiment (29). To this end a specific probe for each of the 5 Mu IFN-α genes and the Mu IFN-β gene was prepared (see also Fig. 4). The specific activity in the relevant 5'-ends of the probes varied from 0.3 to 0.9 x 10^6 dpm/pmol. The Mu IFN-β cDNA was provided by Y. Higashi and Y. Kawade (Kyoto University). 15 μg total RNA was annealed to 0.4 pmol of each fragment. The mixtures were digested with 10,000 units nuclease S1 (Boehringer , Mannheim). Separation of the fragments was performed on a 6% polyacrylamide gel.

RESULTS

Isolation and organization of Mu IFN-α genes

A genomic mouse library was screened with a Hu IFN-α probe as described in Materials and Methods. This approach was possible because of the extensive homology between human and mouse IFN-α genes (11). Out of 150,000 plaques, six hybridizing recombinant phages were isolated after repeated cycles of plaque purification. The DNA of these phages was isolated and analyzed by digestion with several restriction enzymes, Southern blotting and hybridization (results not shown). Two of these phages proved to be particularly interesting. One (called no.10) contained 3 hybridizing fragments in every digest, suggesting 3 separate genes. The second (called no.19) contained 2 fragments that hybridized to the Hu IFN-α probe. The other 4 phages contained one gene each and have not been studied so far.

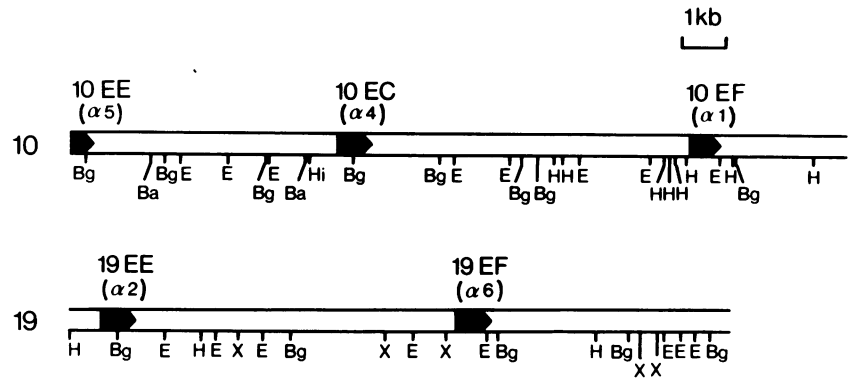

Figure 1. Physical maps of the mouse DNA inserts of phage 10 and 19. Restriction endonuclease cleavage sites are designated: E, Eco RI; Ba, Bam HI; Bg, Bgl II; H, Hind III; X, Xba I. Gene regions are represented by solid boxes. Arrowheads indicate the direction of transcription.

A physical map of the mouse DNA inserts of phages 10 and 19 was established and is shown in Fig. 1. The cloned portion of mouse DNA in phage 10 is 17.8 kb in length. Hybridizing Eco RI fragments of phage 10 were 4.3, 2.6 and 1.6 kb in length and were called 10EC, 10EE and 10EF, respectively. Complete IFN genes are present on fragments 10EC and 10EF. Fragment 10EE, located next to the right arm of the phage, contains the 3' part of a third gene. These genes were called after the Eco R1 fragments on which they were located. The coding regions of genes 10EF and 10EC are separated by 7.6 kb, gene 10EC and 10EE are separated by 5.6 kb. The mouse DNA insert of phage 19 is 15.1 kb in length. Hybridizing Eco R1 fragments of phage 19 were 2.4 and 1.6 kb long and called 19EE and 19EF, both contained a complete IFN gene. The genes 19EE and 19EF are separated from each other by 7.6 kb.

Structure of the Mu IFN-α genes

The primary structure of the Mu IFN-α genes and their flanking sequences were determined by the enzymatic method of Sanger et al. (27) using the single-stranded phages M13 mp8 and mp9 (26). The nucleotide sequences of the genes are shown in Fig. 2. In all four complete genes an uninterrupted open reading frame was found encoding 186 to 190 amino acids. As expected fragment 10EE only contained the 3' part of an IFN gene, but here also an open reading frame was found. Within the coding regions the sequences are aligned to maximize homology. ATG and TGA triplets are boxed. At approximately 90 nucleotides upstream from the initiating ATG codon the sequence TATTTAA is found. This sequence is found in that position in other

IFN-α genes and is thought to represent a modified Goldberg-Hogness box (30). Gene 10EC has a deletion of 15 nucleotides between nucleotides 377 to 393 when compared to the other genes. This results in a deletion of 5 amino acids in the putative protein. 10EC has an additional triplet at position 49-50 which would be in the signal sequence of the protein.

In Fig. 3 the amino acid sequences of the Mu IFN-α proteins as deduced from the nucleotide sequences are shown. Residues identical in all sequences are underlined. By comparison with the published N-terminal sequence of a Mu IFN-α protein (31) the first 23 (24 in the case of 10EC) amino acids presumably represent the signal peptide. The mature proteins are 166 or 167 amino acids in length, except the protein encoded by gene 10EC which would be 162 amino acids long. There is a putative glycosylation site (N A T) at position 78 - 80 in the proteins encoded by genes 10EC, 10EE, 10EF and 19EE, but not in the protein encoded by gene 19EF. To establish that the cloned genes code for genuine IFNs, they have to be expressed in prokaryotic or eukaryotic cells. Such experiments were carried out with genes 10EF, 19EE and 19EF, using transient expression in the monkey COS cell system (32). Preliminary data show that all three genes code for a protein with antiviral properties (data not shown). The protein product of gene 10EF has already been studied in detail (33).

IFN mRNA transcription in mouse L cells

To determine whether the 5 genes are actually expressed in mouse cells, we did a nuclease S1 experiment. To this end a specific probe for each gene was prepared (see Fig. 4). Since mouse L cells produce both IFN-α and -β upon induction (34,35), we also used a Mu IFN-β probe (10). Figure 5 shows the autoradiograph of the gel. For each gene a specific protected fragment can be seen in lanes with RNA from induced cells (lanes +), but not in those with mock-induced RNA (lanes -). As a control for the specificity of the experiment we used RNA from a hamster cell clone transcribing large quantities of the mRNA from gene 10EF (lane CHO-10EF, see also ref. 33). This RNA was hybridized to the probe from gene 19EE and no protection of the DNA fragment against degradation by nuclease S1 was observed. Thus we conclude from these data that all five Mu IFN-α genes and also the Mu IFN-β gene are transcriptionally active in induced mouse L cells. The lengths of the protected fragments suggest a transcription start site between nucleotides -67 to -74, which would agree with transcription starting at the sequence AGA also found at the transcription start site of the Hu IFN-α1 mRNA (36). With the probe for gene 19EF two protected fragments are seen; at

```
          -180                                                -140
10EC (α4)    GCAGAGAGTGAAGTAAAGAAAGTGAAAAGAGAATTGGAAAGCAAGGGGAGGGTATTCCGAAA
10EE (α5)
10EF (α1)    AGAGAATGAGTTAAAGAAAGTGAAAAGACAAGTGGAAAGTGATGGAAGGGCATTCAGAAAGT
19EE (α2)    CAGAGAGTGAAGTAAAGAAAGTGAAAAGAGAATTAGGAAGCAAGGGGAGGGTATTTGGAAAG
19EF (α6)    CTAGACACAAGTAGAGAATAAGTTAAAGTGAAAGGACAGTTGGAAAGTTAGAGGAGGGCATT

               -100                                              -60
GGAGAAACTTCTATTTGTCCCTATTTAAGAGAGATGTACACAGCAGGCTCTCAGAGAACCTGTAGGAGAAGACTC

AAAAACCAGTGTTTGCCCTATTTAAGACACATTCACCCAGGATGGTCTTCAGAGAACCTAGAGGGAAGGATCAG
GAGAAACTTGTGTTTGTCCCTATTTAAGATAGATGCACACAGCAGGCTTCTCAGAGAACCTAGAGGAGAAGACTC
CAGAAAGTAAAAACTAGTGTTTGCCCTATTTAAGACACAGGATGGTCTTCAGAAGACCTAGAGGGAAGGAGCAG

                        -20                  1                        20
AAACCCACAGCCCAGAGAGCGACCAGCATCTACAAGACCCACA ATG GCT AGG CTC TGT GCT TTC CTC

GACCAAACAGTCCAGAAGACCAGAAGCTTTGGCAACACTCACC ATG GCT AGG CTC TGT GCT TTC CTG
AAACACACAGTCCAGAGAGCCATCAACCTCTGCAAGACCCACA ATG GCT AGA CTC TGT GCT TTC CTC
GACTAAACAGCCCAGAAGACCAGAAGCATTGGCAACATTCACC ATG GCT AGG CTC TGT GCT TTC CTG

                40                                  60                          80
ATG ATC CTG GTA ATG ATG AGC TAC TAC TGG TCA GCC TGT TCT CTA GGA TGT GAC CTG

ATG GTC CTG GCG GTG ATG AGC TAC     TGG CCA ACC TGC TCT CTA GGA TGT GAC CTG
GTG ATG CTG ATA GTG ATG AGC TAC     TGG TCA ATC TGT TCT CTA GGA TGC GAT CTG
ATG GTT CTG GCT GTG ATG AGC TAC     TGG TCA ACC TGC TCT CTA GGA TGT GAC CTG

                       100                        120
CCT CAC ACT TAT AAC CTC GGG AAC AAG AGG GCC TTG ACA GTC CTG GAA GAA ATG AGA

CCT CAG ACT CAT AAC CTC AGG AAC AAG AGA GCC TTG ACA CTC CTG GTA CAA ATG AGG
CCT CAC ACT TAT AAC CTC AGG AAC AAG AGG GCC TTG AAG GTC CTG GCA CAG ATG AGG
CCT CAG ACT CAT AAA CTC AGG AAC AAG AGA GCC TTG ACA CTC TTG ATA CAA ATG AGG

                              160                        180
AGA CTC CCC CCT CTT TCC TGC CTG AAG GAC AGG AAG GAT TTT GGA TTC CCC TTG GAG

AGA CTC TCC CCT CTC TCC TGC CTG AAG GAC AGG AAG GAC TTT GGA TTC CCG CAG GAG
AGG CTC CCC TTT CTC TCC TGC CTG AAG GAC AGG CAG GAC TTT GGA TTC CCC CTG GAG
AGA CTC TCC CCT CTC TCC TGC CTG AAG GAC AGG AAG GAC TTT GGA TTC CCC CAG GAG

   200                           220                           240
AAG GTG GAT AAC CAA CAG ATC CAG AAG GCT CAA GCC ATC CTT GTG CTA AGA GAT CTT
                                                                      CTG
AAG GTG GAT GCC CAG CAG ATC AAG AAG GCT CAA GCC ATC CCT GTC CTG AGT GAG CTG
AAG GTG GAT AAC CAG CAG ATC CAG AAG GCT CAA GCC ATC CCT GTG CTG CGA GAT CTT
AAG GTG GAT ACC CTG AAG ATC CAG AAA GAA AAA GCC ATC CCT GTC CTA AGT GAG GTG

       260                           280                           300
ACC CAG CAG ATT TTG AAC CTC TTC ACA TCA AAA GAC TTG TCT GCT ACT TGG AAT GCA
ACC CAG CAG GTC CTG AAC ATC TTC ACA TCA AAG GAC TCA TCT GCT GCA TGG AAT GCA
ACC CAG CAG ATC CTG AAC ATC TTC ACA TCA AAG GAC TCA TCT GCT GCT TGG AAT GCA
ACT CAG CAG ACC TTG AAC CTC TTC ACA TCA AAG GCT TCA TCT GCT GCT TGG AAT GCA
ACC CAG CAG ATC CTG AAC ATC TTC ACA TCA AAG GAT TCA TCT GCT GCT TGG GAT GCA
```

```
          320                                 340                                 360
ACT CTC CTA GAC TCA TTC TGC AAT GAC CTC CAT CAG CAG CTC AAT GAT CTC AAA GCC
ACC CTC CTA GAC TCA TTC TGC AAT GAA GTC CAT CAG CAG CTC AAT GAC CTC AAA GCC
ACC CTC CTA GAC TCA TTC TGC AAT GAC CTC CAC CAG CAG CTC AAT GAC CTG CAA GGT
ACC CTC CTA GAC TCA TTC TGC AAT GAC CTC CAC CAG CAG CTC AAT GAC CTG CAA ACC
ACC CTC CTA GAT ACA TTT TGT AAT GAC CTC TAC CAG CAG CTC AAT GAC CTG CAA GCC

               380                                 400                                 420
TGT GTG ATG CAG                                GAA CCT CCT CTG ACC CAG GAA GAC TCC CTG
TGT GTG ATG CAA CAG GTC GGG GTG CAG GAA TCT CCC CTG ACC CAG GAA GAC TCC CTG
TGT CTG ATG CAG CAG GTG GGG GTG CAG GAA TTT CCC CTG ACC CAG GAA GAT GCC CTG
TGT CTG ATG CAG CAG GTG GGG GTG CAG GAA CCT CCT CTG ACC CAG GAA GAC GCC CTG
TGT CTA GTG CAG CAG GTG AGG TTG CAG GAA CCT CCT CTG ACC CAG GAA GTC TCC CTG

                    440                                 460                                 480
CTG GCT GTG AGG ACA TAC TTC CAC AGG ATC ACT GTG TAC CTG AGA AAG AAG AAA CAC
CTG GCT GTG AGG AAA TAC TTC CAC AGG ATC ACT GTG TAC CTG AGA GAG AAG AAA CAC
CTG GCT GTG AGG AAA TAC TTC CAC AGG ATC ACT GTG TAC CTG AGA GAG AAG AAA CAC
CTG GCT GTG AGG AAA TAT TTC CAC AGG ATC ACT GTG TAC CTG AGA GAG AAG AAA CAC
CTG GCT GTG AGG AAA TAC TTC CAC AGG ATC ACT GTG TAC CTG AGA GAG AAG AAA CAC

                         500                                 520
AGC CTC TGT GCC TGG GAG GTG ATC AGA GCA GAA GTC TGG AGA GCC CTC TCT TCC TCA
AGC CTC TGT GCC TGG GAG GTG GTC AGA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCA
AGC CCC TGT GCC TGG GAG GTG GTC AGA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCT
AGC CCC TGT GCC TGG GAG GTG GTC AGA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCA
AGC CCC TGT GCC TGG GAG GTG GTC AGA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCT

540                                 560                                 580
ACC AAC TTG CTG GCA AGA CTG AGT GAG GAG AAG GAG [TGA] GTCCTGAGACAAAGTGTGGAGAG
GTT AAC TTG CTG GCA AGA TTG AGC AAG GAG GAG [TGA] GTC CTGAGACAAAGTGTGGAGAGGAC
GCC AAT GTG CTG GGA AGA CTG AGA GAA GAG AAA [TGA] GTC CTGAGCCAAAGTGTAGAGGACTC
GTC AAC TTG CTG CCA AGA CTG AGT GAA GAG AAG GAG [TGA] GTGTGAGACAAAGTGTGGAGAGA
GCC AAT GTG CTG GGA AGA CTG AGA GAA GAG AAA [TGA] GTC CTGAGCCAAAGTGTAGAGGACTC

                    620
GACTTTTCTGGACCAGAACACTGCATCTCACTTTATAAGC
TTTTCTGAACCAGAACACTGCATCTCACTTTATAAGATCT
TTCCAGACTAGAACACTGTACCTCTCTGCTCATATCTCTG
CCTCCCCTGGACTAGAAACTGCATCTCATTTTATAAGCTC
TTCCAGACTAGAACACTGTACCTCTCTGCTCATATCTCTG
```

Figure 2. Nucleotide sequences of the Mu IFN-α genes present on phages 10 and 19. ATG and TGA triplets are boxed. Within the coding region the sequences are aligned to maximize homology.

present we cannot discriminiate whether this represents an additional transcription start site at approximately 107 nucleotides upstream from the ATG codon, a splice acceptor site or a virus-induced transcript from the other DNA strand. The intensity of the signals on the autoradiograph shown in Fig. 5 varies considerably between the genes. When the specific activity in the relevant 5' ends of the probes is taken into account, it appears that the mRNA for the β gene is the most abundant, closely followed by the 10EC

```
                S1                  S10                 S20        1                  10
10EC (α4)       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 H T Y N L G N K R A L T V L
10EE (α5)
10EF (α1)       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 Q T H N L R N K R A L T L L
19EE (α2)       M A R L C A F L V M L I V M S Y   W S I C S L G C D L P H T Y N L R N K R A L K V L
19EF (α6)       M A R L C A F L M V L A V M S Y   W S T C S L G C D L P Q T H K L R N K R A L T L L

 20                 30                  40                  50                  60
E E M R R L P P L S C L K D 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 T Q Q I L N L F
                                                                    L T Q Q V L N I F
V Q M R R L S P L S C L K D 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 T Q Q I L N I F
A Q M R R L P P L S C L K D R Q D F G F P L E K V D N Q Q I Q K A Q A I P V L R D L T Q Q T L N L F
I Q M R R L S P L S C L K D R K D F G F P Q E K V D T L K I Q K E K A I P V L S E V T Q Q I L N I F

 70                 80                  90                  100                 110
T S K D L S A T W N A T L L D S F C N D L H Q Q L N D L K A C V M Q         E P P L T Q E D S L L
T S K D S S A A W N A T L L D S F C N E 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 L L
T S K D S S A A W N A T L L D S F C N D 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 L L
T S K A S S A A W N A T L L D S F C N D L H Q Q L N D L Q T C L M Q Q V G V Q E P P L T Q E D A L L
T S K D S S A A W D A T L L D T F C N D L Y Q Q L N D L Q A C L V Q Q V R L Q E P P L T Q E V S L L

 120                130                 140                 150                 160
A V R T Y F H R I T V Y L R K K K H S L C A W E V I R A E V W R A L S S S T N L L A R L S E E K E
A V R K Y F H R I T V Y L R E K K H S L C A W E V V R A E V W R A L S S S V N L L A R L S K E E
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 R A E V W R A L S S S A N V L G R L R E E K
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 R A E V W R A L S S S V N L L P R L S E E K E
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 R A E V W R A L S S S A N V L G R L R E E K
```

Figure 3. Amino acid sequences of the Mu IFN-α proteins predicted from the DNA sequence shown in Fig. 2. The sequence of the signal peptides is designated S1-S24, and the mature proteins as 1-167. Putative N-glycosylation sites are enclosed. Residues identical in all sequences are underlined.

mRNA and that the level of the latter is at least 10 times higher than that of the other α mRNAs.

DISCUSSION

We have isolated recombinant phages from a Balb/c genomic library containing Mu IFN-α genes. Two phages (called 10 and 19), which contained four complete and one partial IFN-α gene were studied in detail. The

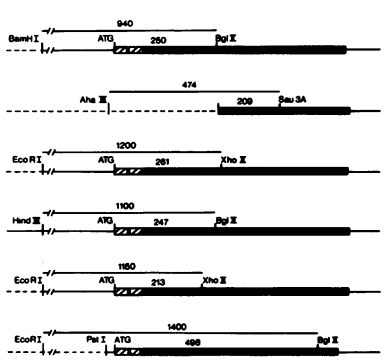

Figure 4. Specific probes utilized for S1 mapping. The probes and their lengths are indicated above the diagram of each gene. The regions encoding the signal peptide and mature protein are indicated by hatched and solid boxes, respectively. Mouse DNA sequences are represented by solid lines and bacterial DNA by stippled lines. The distance from ATG codon to the relevant 5'-end of the probe is indicated between these landmarks, except for gene 10EE where this distance represents the number of nucleotides that are colinear with the probe.

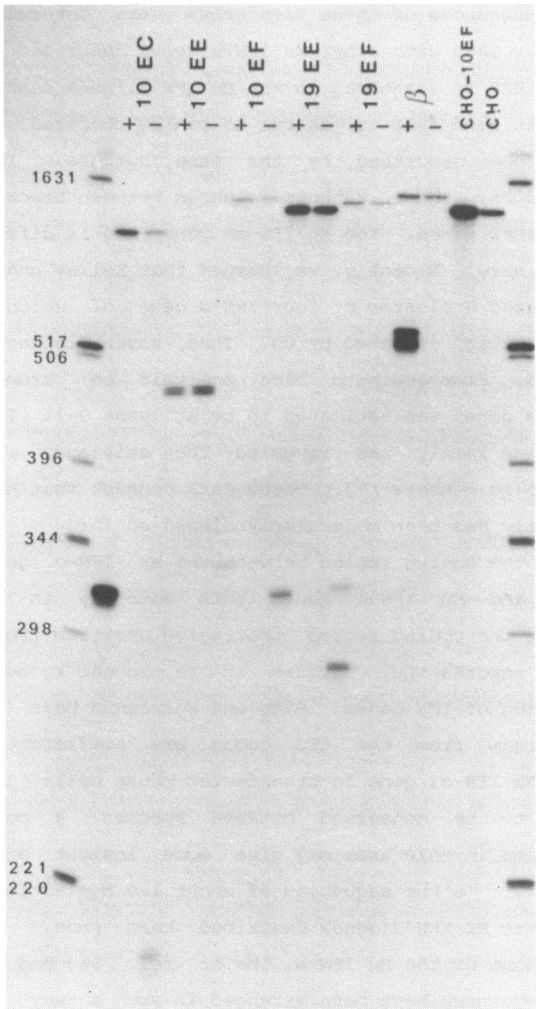

Figure 5. Specific S1 mapping of the individual Mu IFN mRNAs. Total RNA from Sendai virus-induced (lanes +) or mock-induced (lanes -) mouse L cells was hybridized to the probes shown in Fig. 4 and digested with nuclease S1. The probe used is indicated at the top of the lanes. Lanes CHO-10EF and CHO are controls (see text). The marker lanes consist of a 5'-labeled Hinf I digest of pBR322. The length of these fragments is indicated.

presence of more than one gene within a small piece of mouse DNA suggests that the Mu IFN-α genes are tightly linked in the genome, as was shown to be the case for the Hu IFN-α gene family (37). This also agrees with the clustering of the IFN-α genes on mouse chromosome 4 (12,13,14).

The nucleotide sequences of these five genes were determined (Fig. 2). Comparison of our data with other Mu IFN-α genes described recently (22,23) showed that gene 10EF is identical to the Mu IFN-α1 gene described by Shaw et al. (22) and that gene 19EE is, as far as can be compared, identical to the Mu IFN-α2 cDNA clone described by the same authors. In Fig. 2 the nomenclature used by these authors is shown between brackets and has been extended to the other genes. The Mu IFN-αA gene (23) is different from the genes described here. Recently, we learned that Kelley and Pitha (personal communication) cloned a cluster of four IFN-α genes of which three overlap the phage 10 cluster isolated by us. Thus, seven distinct Mu IFN-α genes have been isolated. From Southern blot analysis on chromosomal DNA the number of Mu IFN-α genes was estimated to be at least 9-11 (14,22). Analysis of the Hu IFN-α gene family has revealed the existence of at least 14 distinct non-allelic members (7). These data suggest that about half of the Mu IFN-α gene family has been molecularly cloned so far.

The homology in the coding region between the Mu IFN-α genes is 88-94%. Between the Hu and Mu IFN-α genes this homology is 70-75%. From the sequences flanking the coding region, the region upstream from the TATA box is particularly interesting, because it is thought to be involved in the inducible expression of IFN genes. Ragg and Weissmann have shown that 177 base pairs upstream from the ATG codon are sufficient to allow viral induction of the Hu IFN-α1 gene in transfected mouse cells (38). Since the induction seems to be conserved between species, a comparison of the nucleotide sequences in this area may give some insight on the mechanism involved. In Fig. 6 the sequences of about 180 nucleotides upstream from the ATG codon of the Mu IFN-α genes described here (nos. 1-4) are shown together with those of the Hu IFN-α1 (no.5; ref. 36) and -α2 genes (no.6; ref. 30). The sequences have been arranged in such a way as to maximize alignment. Nucleotides identical in all sequences are underlined; the TATA box is enclosed. The homology between the murine sequences upstream from the TATA box is about 75%, this homology is 65% between the TATA box and the ATG. Comparison of human and murine sequences shows that upstream from the TATA box about 45% of the sequence is conserved. An interresting feature of the sequences shown in Fig. 6 is that notably the first 50 nucleotides (-180 - -130) are extremely purine rich: 80-88% of the nucleotides are A or G. In the Hu IFN-β gene (39,40,41) this region contains 75% purines. Although the overall homology between these regions is not very striking, their purine

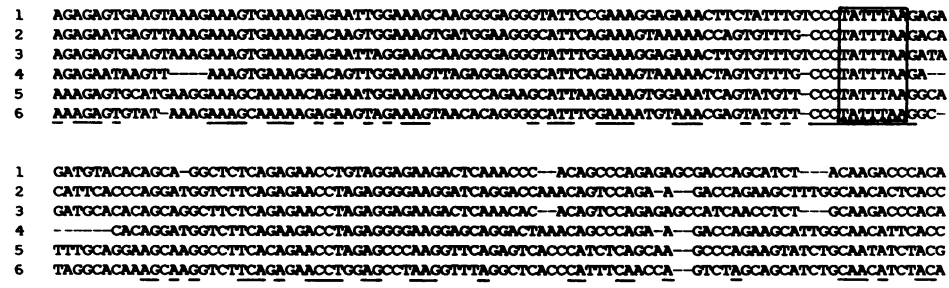

Figure 6. Comparison of 5'-non-coding sequences. Approximately 180 nucleotides preceeding the ATG codon of 6 IFN-α genes are compared. Gaps (-) were introduced to maximize alignment. Nucleotides identical in all sequences are underlined. The presumed TATA box is enclosed. Sequences 1 to 4 are from this paper and represent 10EC (Mu IFN-α4), 10EF (Mu IFN-α1), 19EE (Mu IFN-α2) and 19EF (Mu IFN-α6), respectively. Sequence 5 is the Hu IFN-α1 gene from Nagata et al. (36) and no. 6 the Hu IFN-α2 gene from Lawn et al. (30).

richness is a common feature and may be involved in the regulation of the genes.

Expression of the cloned Mu IFN-α genes and the Mu IFN-β gene was studied in mouse L cells. The results indicate that all genes are transcribed upon induction of the cells with Sendai virus (see Fig. 5). Interestingly, there are differences in the levels of the individual mRNAs. IFN-β mRNA and the 10EC mRNA are the predominant species in these cells. The amount of 10EC mRNA is at least 10 times higher than that of the other α mRNAs. Similar data were found by Hiscott et al. (42) for the Hu IFN-α genes. However, the mechanism by which these differences in expression are brougth about remains to be elucidated. So far we expressed three genes (10EF, 19EE and 19EF) transiently in monkey COS cells and found that they gave rise to protein products with antiviral activity. This confirms the results of Shaw et al. (22) and shows that 19EF codes for a biologically active protein.

Comparison of the amino acid sequences of the Mu IFN-α proteins shows that over 70% of the amino acids are conserved (Fig. 3). Between the Hu IFN-α and the Mu IFN-α proteins there is a homology of 50-60%. The conservation between Mu IFN-α proteins and Mu IFN-β (10) is low, about 27% of the residues are conserved throughout the sequence. A similar degree of homology was found between the Hu α and β IFNs (43). However, the region AWEVVRAE present in Hu and Mu IFN-α proteins between positions 140-147 is almost identical to the sequence AWMVVRAE in Mu IFN-β (amino acids 137-144, ref. 10) and a similar region is found in Hu IFN-β (8). This region might be important in

one of the biological activities of the IFNs.

There is a potential N-glycosylation site in four Mu IFN-α proteins described here. In the protein encoded by gene 19EF this site is absent. The protein product of the Mu IFN-αA gene also contains no glycosylation site (23). The presence of putative glycosylation sites in part of the Mu IFN-α subspecies is in good agreement with earlier studies on the Mu IFN proteins (44,45). That the presence of a glycosylation site is not necessary for the biological activity of Mu IFN-α is demonstrated by the antiviral properties of the proteins encoded by genes 19EF and A. A similar glycosylation site is found in the Hu and Mu IFN-β proteins, but not in Hu IFN-α (10,21). The Mu IFNs described so far have 5 cystein residues at positions 1, 29, 86, 99 and 139. Most Hu IFN-α proteins have four cysteins at positions 1, 29, 99 and 139 (21). It is thought that cysteins 1 + 99 and 29 + 139 form S-S bridges. The fifth cystein in two Hu IFN-α proteins (α1 and α13) is also in position 86.

The molecular cloning of Mu IFN-α genes will enable us to obtain large quantities of the different Mu IFN-α subspecies. This has already been achieved for the protein product of gene 10EF (33). The pure subspecies can be studied in detail regarding their various biological activities amongst which are their *in vivo* antiviral and antitumor properties.

ACKNOWLEDGEMENTS

This work was sponsored in part by The Netherlands Cancer Foundation (KWF). The authors are indebted to L. Hood for the Balb/c genomic library, to P. Boseley and A. Easton for the Hu IFN-α probe and to Y. Higashi and Y. Kawade for the Mu IFN-β cDNA clone.

REFERENCES

1. Stewart II, W.E. (1979) The Interferon System, Springer Verlag, New York.
2. Lengyel, P. (1982) Ann. Rev. Biochem. 51, 251-282.
3. Kirkwood, J.M. (1984) J. Clin. Oncology 2, 336-352.
4. Gresser, I. and Tovey, M.G. (1978) Biochem. Biophys. Acta 516, 231-247.
5. Shulman, L. and Revel, M. (1980) Nature, 287, 98-100.
6. Fellous, M., Nir, U., Wallach, D., Merlin, G., Rubinstein, M. and Revel, M. (1982) Proc. Natl. Acad. Sci. USA 79, 3082-3086.
7. Todokoro, K., Kioussis, D. and Weissmann, C. (1984) EMBO J. 3, 1809-1812.
8. Houghton, M. Jackson, I.J., Porter, A.G., Doel, S.M., Catlin, G.M., Barber, C. and Carey, N.H. (1981) Nucleic Acids Res. 9, 247-266.
9. Tavernier, J., Derynck, R. and Fiers, W. (1981) Nucleic Acids Res. 9, 461-471.
10. Higashi, Y., Sokawa, Y., Watanabe, Y., Kawade, Y., Ohno, S., Takaoka, C.

and Taniguchi, T. (1983) J. Biol. Chem. 258, 9522-9529.
11. Owerbach, D., Rutter, W.J., Shows, T.B., Gray, P.W., Goeddel, D.V. and Lawn, R.M. (1981) Proc. Natl. Acad. Sci. USA 78, 3123-3127.
12. Kelley, K.A., Kozak, C.A., Dandoy, F., Sor, F., Windass, J.D., DeMaeyer-Guignard, J., Pitha, P.M. and DeMaeyer, E. (1983) Gene 26, 181-188.
13. Lovett, M., Cox, D.R., Yee, D., Boll, W., Weissmann, C., Epstein, C.J. and Epstein, L.B. (1984) EMBO J. 3, 1643-1646.
14. Van der Korput, J.A.G.M., Hilkens, J., Kroezen, V., Zwarthoff, E.C. and Trapman, J. (1984) J. Gen. Virol. in press.
15. Gray, P.W., Leung, D.W., Pennica, D., Yelverton, E., Najarian, R., Simonsen, C.C., Derynck, R., Sherwood, P.J., Wallace, D.M., Berger, S.L., Levinson, A.D. and Goeddel, D.V. (1982) Nature 295, 503-508.
16. Devos, R., Cheroutre, H., Taya, Y., Degrave, W., Van Heuverswyn, H. and Fiers, W. (1982) Nucleic Acids Res. 10, 2487-2501.
17. Gray, P.W. and Goeddel, D.V. (1983) Proc. Natl. Acad. Sci. USA 80, 5842-5846.
18. Trent, J.M., Olson, M. and Lawn, R.M. (1982) Proc. Natl. Acad. Sci. USA 79, 7809-7813.
19. Naylor, S.L., Sakaguchi, A.Y., Shows, T.B., Law, M.L., Goeddel, D.V. and Gray, P.W. (1983) J. Exp. Med. 57, 1020-1027.
20. Naylor, S.L., Lalley, P.A., Sakaguchi, A.Y. and Gray, P.W. (1984) Cytogen. Cell Gen. 37, 550.
21. Pestka, S. (1983) Archiv. Biochem. Biophys. 221, 1-37.
22. Shaw, G.D., Boll, W., Taira, H., Mantei, N., Lengyel, P. and Weissmann, C. (1983) Nucleic Acids Res. 11, 555-573.
23. Daugherty, B., Martin-Zanca, D., Kelder, B., Collier, K., Craig Seamans, T., Hotta, K. and Pestka, S. (1984) J. Interferon Res. 4, 635-643.
24. Benton, W.D. and Davies, R.W. (1977) Science 196, 180-182.
25. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, Cold Spring Harbor, New York.
26. Messing, J. (1983) Methods in Enzymology, Wu, R., Grossman, C. and Moldave, K., Eds., Vol. 101, pp. 20-77, Academic Press, New York.
27. Sanger, F. Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
28. Chirgwin, J.M., Przybyla, A.E., Mac Donald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
29. Berk, A.J. and Sharp, P.A. (1977) Cell, 12, 721-732.
30. Lawn, R.M., Gross, M., Houck, C.M., Franke, A.E., Gray, P.W. and Goeddel, D.V. (1981) Proc. Natl. Acad. Sci. USA 78, 5435-5439.
31. Taira, H., Broeze, R., Jayaram, B.M., Lengyel, P., Hunkapillar, M.W. and Hood, L.E. (1980) Science 207, 528-530.
32. Gluzman, Y. (1981) Cell 23, 175-182.
33. Zwarthoff, E.C., Bosveld, I.J., Vonk, W.P. and Trapman, J. (1984) J. Gen. Virol., in press.
34. Yamamoto, Y. and Kawade, Y. (1980) Virology 103, 80-88.
35. Trapman, J. (1980) Febs Lett. 109, 137-140.
36. Nagata, S., Mantei, N. and Weissmann, C. (1980) Nature 287, 401-408.
37. Brack, C., Nagata, S., Mantei, N. and Weissmann, C. (1981) Gene 15, 379-394.
38. Ragg, H, and Weissmann, C. (1983) Nature 303, 439-442.
39. Gross, G. Mayr, U., Bruns, W., Grosveld, F., Dahl, H.H.M. and Collins, J. (1981) Nucleic Acids Res. 9, 2495-2507.
40. Degrave, W., Derynck, R., Tavernier, J., Haegeman, G. and Fiers, W. (1981) Gene 14, 137-143.
41. Ohno, S. and Taniguchi, T. (1981) Proc. Natl. Acad. Sci. USA 78, 5305-5309.

42. Hiscott, J., Cantell, K. and Weissmann, C. (1984) Nucleic Acids Res. 12, 3727-3746.
43. Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. and Weissmann, C. (1980) Nature 285, 547-549.
44. Fujisawa, J.I., Iwakura, Y. and Kawade, Y. (1978) J. Biol. Chem. 253, 8677-8679.
45. Fujisawa, J.I. and Kawade, Y. (1981) Virology 112, 480-487.