

---

**Structure and expression in *Escherichia coli* of a cloned rat interferon- $\alpha$  gene**

---

R.Dijkema, P.Pouwels, A.de Reus\* and H.Schellekens\*

---

Medical Biological Laboratory TNO, P.O. Box 45, 2280 AA and \*Primate Centre TNO, P.O. Box 5815, 2280 HV Rijswijk, The Netherlands

---

Received 21 October 1983; Accepted 28 November 1983

---

**ABSTRACT**

DNA synthesized by *in vitro* transcription on rat interferon (IFN) mRNA has been cloned and amplified as recombinant DNA. The nucleotide sequence of these rat IFN cDNA clones revealed i. the partial presence of the coding region of the gene and ii all cDNA clones were derived from the same subtype of rat IFN- $\alpha$  mRNA. Purified inserted fragments were used as a hybridisation probe against chromosomal "Southern blots" to show that at least twelve rat IFN- $\alpha$ -related sequences are present in the genome. A  $\lambda$ -linked rat gene library was screened with the cDNA probes, resulting in an equivalent number of rat IFN- $\alpha$ -related hybrid phages. By use of a 3'-noncoding region as a probe, the chromosomal counterpart of the cDNA clones could be detected and the nucleotide sequence of its coding region has been determined. Expression of the coding region in *E. coli* yielded biologically active IFN, when tested for *in vitro* or *in vivo* antiviral activity.

**INTRODUCTION**

On exposure to viruses or other specific inducers, cells of almost all vertebrates secrete one or more proteins, known as interferons (IFNs). IFNs are being intensively studied because of their potent ability to confer a virus-resistant state in their target cells<sup>1,2</sup>. In addition, they exhibit other effects such as inhibition of cell proliferation, and modulation of the immune response<sup>2,3</sup>.

Sofar three interferons with different antigenic specificity have been recognized,  $\alpha$ ,  $\beta$  and  $\gamma$ . Recently, several groups have reported the construction and characterization of cDNA clones of human IFNs<sup>4-8</sup>, which revealed the existence of a multigene family in the case of human IFN- $\alpha$ <sup>9,10</sup>. At least 8 distinct human IFN- $\alpha$  subtype genes have been identified in a cDNA library prepared from a myeloblastoid cell line<sup>10</sup> as well as a gene bank constructed in phage  $\lambda$ <sup>9</sup>, whereas only one<sup>11,12</sup> or more<sup>13</sup> human IFN- $\beta$ , and one human IFN- $\gamma$ <sup>14</sup> gene exist.

The biological activities of human IFN- $\alpha$  subtype proteins produced in *Escherichia coli* bacteria have been compared with human IFN- $\alpha$  proteins deri-

ved from induced buffy coats and were found to be identical by in vitro and in vivo experiments<sup>5,15</sup>; the efficacy of human IFN- $\alpha_2$  and naturally derived human IFN- $\alpha$  has been demonstrated in a limited animal testing that makes use of squirrel and rhesus monkey. In addition, phase I studies in man showed the same side effects for bacterial human IFN- $\alpha_2$  and natural human IFN- $\alpha$ <sup>16</sup>. Although these side effects could be reproduced in chimpanzees, they cannot when lower primates were used as a model<sup>17</sup>, indicating a species specific response for IFN treatment. This problem can be partly solved by the use of species specific IFNs in other animal test models.

We report in this article the identification of the rat IFN- $\alpha$  gene family comprising presumably twelve members and describe the isolation and characterization of the rat IFN- $\alpha$  subtype gene predominantly expressed after viral infection. Expression of this gene in the procaryotic host E. coli yields biologically active IFN when tested for in vitro and in vivo antiviral potency.

#### MATERIALS AND METHODS

Cells and viruses. A rat embryo cell line (Ratc), a high IFN producer has been described before<sup>18</sup>. Induction of the cells by Sendai or New Castle Disease (NDV) virus and poly I:C were carried out essentially as described<sup>18</sup>.

Preparation and partial purification of IFN mRNA. Total cytoplasmic RNA was extracted from induced Ratc cells by an isotonic buffer A (10 mM Tris-HCl (pH 7.9), 1 mM EDTA, 150 mM NaCl) containing 0.65% Nonidet P<sub>40</sub>. After centrifugation, the supernatant was mixed with an equal volume of buffer B (20 mM Tris-HCl (pH 7.9), 20 mM EDTA, 0.3 M NaCl, 7 M ureum, 1% SDS) and extracted twice with phenol. The RNA was precipitated with ethanol and polyadenylated RNA was prepared by repeated batch-wise adsorption to oligo(dT)-cellulose according to Nagata et al.<sup>4</sup>. Size purification of the IFN mRNA was achieved by use of a continuous PAAGE-SDS system as described<sup>19</sup>. Different size fractions were collected and RNA in each fraction was precipitated with ethanol.

Translation of mRNA in Xenopus laevis oocytes. To detect IFN mRNA activity, part of each fraction (about 0.5  $\mu$ g) was dissolved in 1-5  $\mu$ l 15 mM Tris-HCl (pH 7.5), 88 mM NaCl, and injected in 10-20 oocytes (50 nl per oocyte)<sup>20</sup>. Oocytes were incubated for 16-18 h in Barth's medium<sup>21</sup>, homogenised in 0.2 ml 50 mM Tris-glycine buffer (pH 8.9), and the supernatant solution was assayed for IFN activity.

Interferon assay. IFN was determined by the vesicular stomatitis virus (VSV) plaque reduction assay<sup>22</sup> on Ratec cells. Since there is no international standard for rat IFN available, every rat IFN assay was titrated against a laboratory rat IFN reference<sup>18</sup> and expressed as (arbitrary) units.

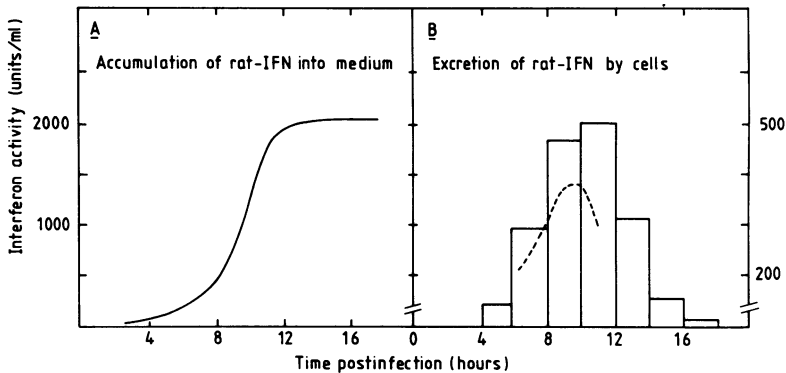
Isolation of IFN- $\alpha$  cDNA clones. Partially purified IFN mRNA yielding approximately 10,000 units/ $\mu$ g of IFN activity in the oocyte assay was used as a template for cDNA synthesis<sup>23</sup> and cloned in the PstI restriction site of pBR322 by dC/dG tailing. Tetracycline resistant transformants were screened by colony hybridization<sup>24</sup> for the presence of rat IFN- $\alpha$  sequences with a 690 bp HindIII-EcoRI fragment of the chromosomal mouse IFN- $\alpha_1$  DNA<sup>25</sup> as a probe.

DNA blot analysis. High molecular weight DNA was prepared from rat liver (strain WagRij) and digested with an excess of restriction endonuclease. Complete digestion of DNA was tested by simultaneous incubation of rat liver DNA and phage  $\lambda$  DNA and the latter pattern was compared to a standard gel pattern. The resulting DNA fragments were separated by electrophoresis through a 0.6% agarose gel and then transferred onto a Schleicher and Schull (BA 85) nitrocellulose filter<sup>26</sup>. Conditions for hybridization with the chromosomal mouse IFN- $\alpha_1$  DNA and rat IFN- $\alpha$  cDNA as a probe and washing of the filters were as described<sup>25</sup>.

Phage library screening. A rat gene library of partially EcoRI-cleaved rat liver DNA (strain Sprague-Dawley) in phage  $\lambda$  Charon 4A<sup>27</sup>, was screened by in situ plaque hybridization<sup>28</sup> with nick-translated probes described above. Approximately 450,000 recombinant phages were screened with a density of 20,000 plaques per 14 cm petri dish. Adsorption of phages was performed by sequentially placing two nitrocellulose filters on each plate for 3 min and 5 min at room temperature, respectively. Further treatment of the filters was as described by Benton and Davis<sup>28</sup>. Positive clones were plaque-purified twice.

DNA sequence analysis. Sequence analysis was performed according to the procedure described by Maxam and Gilbert<sup>29</sup>.

Expression of a rat IFN- $\alpha$  gene in E. coli. The plasmid pMBL604 (M. Broekhuijsen, personal communication), containing a hybrid tetracycline-tryptophan promoter and a synthetic Shine-Dalgarno (SD) sequence, was linearized with EcoRI (see Results section). Plasmid rat IFN- $\alpha_1$  contains a 2.5 kb EcoRI chromosomal rat DNA fragment including the rat IFN- $\alpha_1$  gene. The 2.5 kb EcoRI fragment was digested with HphI which removes the first 17 (out of 22) codons of the sequence coding for the signal peptide. The resulting fragment



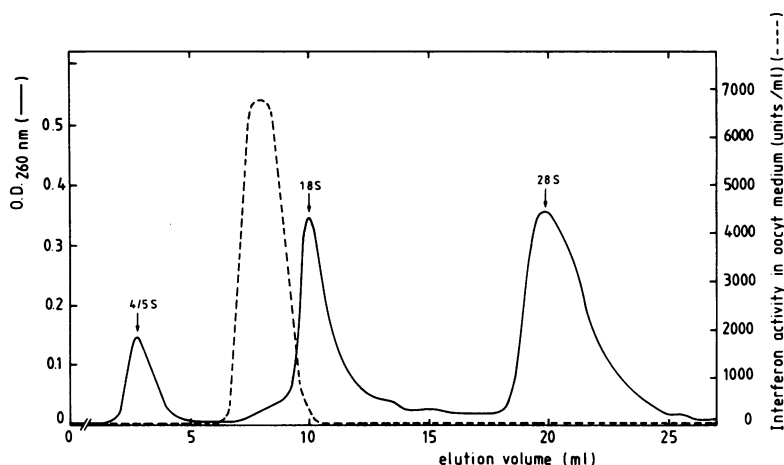
**Figure 1.** Time course of interferon production. A roller bottle of post-confluent Ratec cells was infected with Sendai virus. After a 1 h adsorption period, the cells were washed and incubated in 50 ml serum-free MEM at 37 °C.

- At different times post-infection, samples (2 ml) were removed from the culture medium, adjusted to pH 2.0, kept for 5 days at 4 °C, neutralized, and then tested for IFN activity.
- At different time intervals, the culture medium (50 ml) was removed and replaced by fresh medium. Results of duplicate determinations of withdrawn culture medium are shown (—). In addition the relative amount of IFN mRNA present in the cells was determined by the oocyte assay (---).

was treated with Bal31 in order to remove the additional 6 amino acids signal codons leading to the mature rat IFN- $\alpha_1$  coding sequence. An artificial ATG codon was reintroduced in the form of a synthetic DNA adapter: 5' AATTCTAGACATG GAICTGIAC. Ligation of the adapter to the population of Bal31-treated molecules and the subsequent ligation into EcoRI-linearized pMBL604 was followed by transformation into *E. coli* JA221<sup>30</sup>. Transformants were screened for ampicilline resistance, rat IFN- $\alpha_1$  DNA sequences, and rat IFN- $\alpha_1$  protein synthesis, respectively.

## RESULTS AND DISCUSSIONS

**Kinetics of rat IFN production.** Different viral and non-viral inducers were tested for the efficient stimulation of Ratec cells for IFN. Viral inducers, in particular Sendai virus, gave rise to a marked enhancement of IFN production leading to IFN titers of 2500–5000 units/ml culture medium. Figure I shows the production of rat IFN by Ratec cells after Sendai virus infection. Accumulation of medium IFN and time-dependent IFN secretion by these cells reaches an optimum about 11 h post infection. Therefore, in experiments aimed at the extraction of IFN mRNA, cells were harvested about 10 h after viral induction.



**Figure 2.** Analytical gel electrophoresis of poly(A)-containing RNA from Sendai-induced R6tec cells. Samples containing poly(A)RNA and rRNA markers were applied after heating at 56 °C and eluted (—). RNA from each fraction was precipitated with ethanol, and part of it was used in the oocyte assay to determine the location of IFN mRNA (- - -).

Analysis of IFN mRNA by PAAGE-SDS fractionation. RNA was extracted from  $10^{10}$  Sendai virus-induced R6tec cells. A total of 52 mg of RNA was obtained which, upon processing by oligo(dT) cellulose chromatography, yielded 250  $\mu$ g of mRNA that was subsequently fractionated by a continuous PAAGE-SDS system<sup>19</sup>. The absorbance profile of the RNA after fractionation is shown in Figure II. A single peak of IFN mRNA activity, which contained 8  $\mu$ g of RNA, was found at a position corresponding to a sedimentation coefficient of approximately 14S; the fractionation by this type of gel system resulted in a 50-fold enrichment of IFN mRNA as was determined by injection of crude RNA and 14S mRNA into *X. laevis* oocytes (data not shown).

Construction and identification of clones containing rat IFN- $\alpha$  sequences. Four  $\mu$ g of 14S mRNA (with an IFN titer of 10,000 units/ $\mu$ g in the *X. laevis* oocyte assay) was used to prepare ds cDNA. The cDNA was electrophoresed on a 5% polyacrylamide gel and 125 ng of material ranging from 800-1200 bp in size was recovered by electroelution. Hybrid plasmids were constructed by annealing 125 ng of dC-tailed ds cDNA to 400 ng of PstI-cleaved, dG-tailed pBR322 DNA. After transformation of *E. coli* HB101, approximately 15,000 tetracycline resistant transformants were obtained. These were screened for the presence of rat IFN- $\alpha$  sequences with a nick-translated probe that was derived from the chromosomal mouse IFN- $\alpha_1$

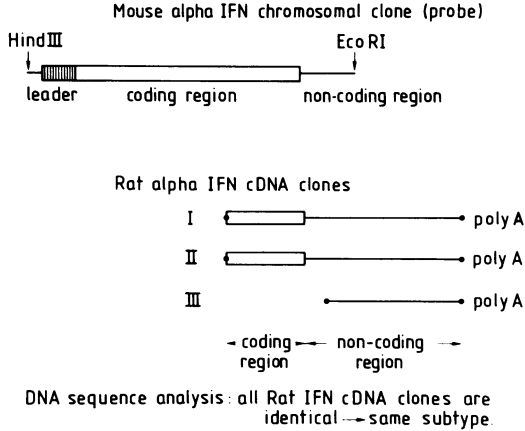
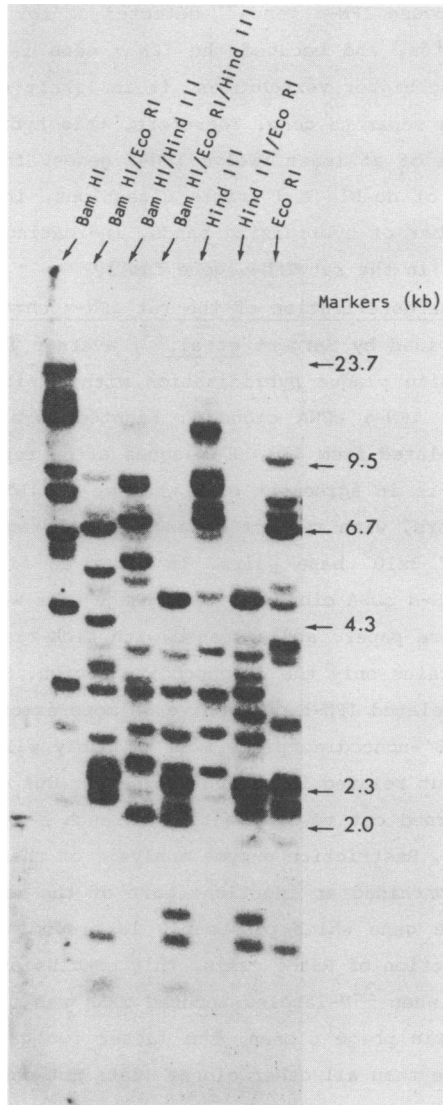


Figure 3. Size determination of the different rat IFN- $\alpha$  cDNA clones. On top is shown the structural organization of the chromosomal mouse IFN- $\alpha_1$  gene that was used as a probe. Below we show the relative positions of the three rat IFN- $\alpha$  cDNA clones.

gene. Three colonies (I-III) were identified which hybridized to the same extent with this probe. When the 750 bp PstI insert from clone I was used again for screening the 14S mRNA-derived cDNA library, no additional IFN-containing transformants could be detected. In order to determine whether the cDNA clones contained a complete coding region, we partly sequenced the inserts by the Maxam-Gilbert procedure. Figure III shows the location of the different rat IFN cDNA inserts with respect to the chromosomal mouse IFN- $\alpha_1$  gene. All cDNA clones were found to contain rat IFN- $\alpha$  sequences; however, they were incomplete with respect to the coding region. Since in general, the 3'-noncoding regions of homologous genes diverge more strongly compared to their coding regions, we performed a DNA sequence analysis of the 3'-noncoding regions of the different cDNA clones. This analysis revealed identical sequences for all three rat IFN- $\alpha$  cDNA clones, indicating their common origin with respect to rat IFN- $\alpha$  mRNA subtype.

Evidence for multiple rat IFN- $\alpha$  genes. Only one subtype rat IFN- $\alpha$  gene could be detected from cloned mRNA of viral-infected Rateg cells, however, a multigene family of IFN- $\alpha$  genes is present in human<sup>9,10</sup>. We therefore determined the total number of rat IFN- $\alpha$  genes present in the rat genome. To this end, rat liver DNA was digested with various restriction endonucleases, electrophoresed through agarose gels and transferred to nitrocellulose paper. Hybridizations were performed by using nick-translated rat IFN- $\alpha$  cDNA



**Figure 4.** Hybridization of rat DNA with IFN- $\alpha$  related sequences as a probe. Rat liver DNA was digested with the indicated restriction endonucleases and processed for "Southern blotting" as described in the Materials and Methods section. Sizes in kilobase pairs (kb) of bacteriophage  $\lambda$  DNA digested with HindIII were included in the gel as size standards.

clone I and the chromosomal mouse IFN- $\alpha_1$  gene as probes (Figure 4). The number of IFN- $\alpha$  genes can be estimated from the number of hybridizing bands in the lanes containing BamHI and HindIII digests. Since none of the human

IFN- $\alpha$  genes<sup>9,10</sup> and mouse IFN- $\alpha$  genes<sup>25</sup> detected so far contains either of these restriction sites, and because the IFN- $\alpha$  gene family appears to be highly conserved among higher vertebrates, it is likely to assume that each band corresponds to a separate gene. Therefore, this hybridization analysis suggests the presence of at least twelve IFN- $\alpha$  genes. In addition, the hybridization patterns of double and triple digestions, including EcoRI, result in the same number of hybridizing bands, indicating again the lack of this restriction site in the rat IFN- $\alpha$  gene family.

Isolation and characterization of the rat IFN- $\alpha$  chromosomal genes. The rat gene library provided by Sargent *et al.*<sup>27</sup>; average fragment size 15 kb was screened by in situ plaque hybridization with a nick-translated probe derived from the rat IFN- $\alpha$  cDNA clone I. Eighteen hybridization-positive phage clones were isolated from 450,000 plaques after repeated plaque purification. This value is in agreement with what we should expect for a gene family of 12-13 members, with respect to an average fragment size of 15 kb and a genome size of  $3 \times 10^9$  base pairs. In order to find the chromosomal counterpart of the IFN- $\alpha$  cDNA clones, hybrid phage DNAs were isolated, spotted onto nitrocellulose paper, and screened with nick-translated IFN- $\alpha$  cDNA clone III, which contains only the 3'-noncoding region. Since the 3'-noncoding regions of the related IFN- $\alpha$  genes diverge more strongly than their coding regions<sup>10</sup> this 3'-noncoding probe most probably will discriminate between the different but related IFN- $\alpha$  subtype genes. Out of 18 IFN- $\alpha$  related hybrid phages two turned out to be positive in such a hybridization screening (data not shown). Restriction enzyme analysis of these two phage clones revealed that they contained an identical part of the rat genome including the rat IFN- $\alpha$  subtype gene which presumably is predominantly expressed as mRNA after viral infection of Ratec cells. This conclusion was reinforced by the observation that when <sup>32</sup>P-labeled induced mRNA was used as a probe for screening the 18 hybrid phage clones, the latter two gave a much stronger hybridization response than all other clones (data not shown).

Nucleotide sequence of a rat IFN- $\alpha$  gene. The two  $\lambda$  hybrid phage clones designated  $\lambda$  rat IFN- $\alpha_1$  contain a 15.5 kilobase (kb) sequence of the rat genome including the structural gene for IFN- $\alpha$ . Digestion of  $\lambda$  rat IFN- $\alpha_1$  with EcoRI produces fragments of 20 and 10.5 kb deriving from  $\lambda$  DNA, and fragments of 8.0, 2.9, 2.5 and 2.1 kb from the rat DNA insert. Only the 2.5 kb fragment hybridizes to the IFN- $\alpha$  cDNA probe. This 2.5 kb EcoRI fragment was subcloned into the EcoRI site of pBR322 (plasmid rat IFN- $\alpha_1$ ). A map of restriction endonuclease sites in the coding region of the gene and the sequen-



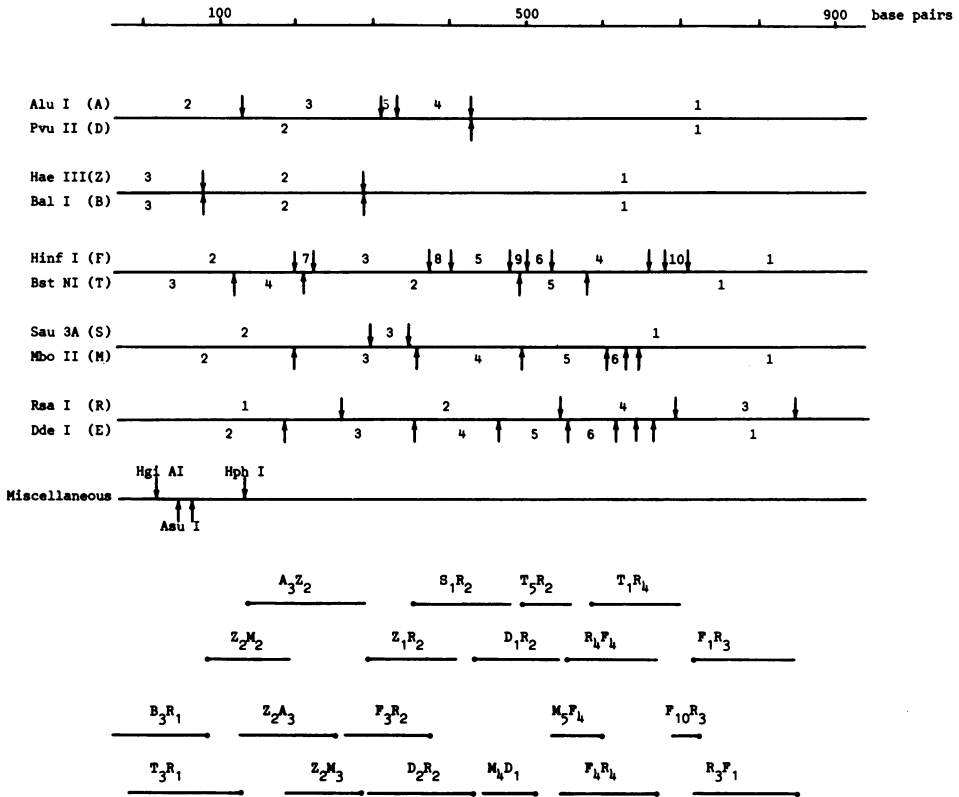


Figure 5. Restriction enzyme map and sequence strategy for the coding region of the rat IFN- $\alpha_1$  chromosomal gene. The relative positions of some restriction endonuclease sites were inferred from Southern blotting analysis of plasmid rat IFN- $\alpha_1$  with the mouse IFN- $\alpha_1$  gene as a probe (data not shown). Nucleotide sequences obtained from these fragments revealed the presence of many other restriction enzyme sites that were used in subsequent sequence analysis studies. Below is shown a schematic representation of the sequenced tracts. In the fragment names, the first capital denotes the endonuclease used for primary cleavage, the second capital the endonuclease used for secondary scission, which separates the end labels on the primary restriction sites, e.g., fragment A<sub>3</sub>Z<sub>2</sub> is the result of HaeIII cleavage of AluI fragment A<sub>3</sub>; it is that portion of A<sub>3</sub> that overlaps with HaeIII fragment Z<sub>2</sub>.

cing strategy used are shown in Figure 5; the 2.5 kb EcoRI fragment was cleaved by an appropriate restriction enzyme, labeled with <sup>32</sup>P at the 5' termini, and digested with a second restriction enzyme to yield single-labeled fragments that serve as template for the chemical degradation procedure of Maxam and Gilbert<sup>29</sup>. The nucleotide sequence obtained in this way is shown in Figure VI. An open translational reading frame is present between

Rat IFN chromosomal clone 6 and 11



**Figure 6.** The partial nucleotide sequence of plasmid rat IFN- $\alpha_1$ . The nucleotide sequence was determined as indicated in the legend to Figure V. The amino acid sequence was deduced from the nucleotide sequence. Indicated are the putative starts of the IFN mRNA, pre IFN- $\alpha_1$  protein, mature IFN- $\alpha_1$  protein, and stopcodon. The arrow indicates the position of the largest rat IFN- $\alpha$  cDNA (clone I).

```

1
Met Ala Arg Leu Cys Ala Phe Leu Met Val Leu Ala Val Met Ser Tyr Trp Pro
ATG GCT AGG CTC TGT GCT TTC CTG ATG GTC GCG GTG ATG AGC TAC TGG CCA
* * * * *
ATG GCT CCG CTC TGT GCT TTC TTG ATG TCC CTG GTG GTG GTG ATG TGG TCA
Met Ala Arg Leu Cys Ala Phe Leu Met Ser Leu Val Val Val Ser Tyr Trp Ser
1

19
Thr Cys Ser Leu Gly Cys Asp Leu Pro Gln Thr His Asn Leu Arg Asn Lys Arg
ACC TGC TCT CTA GGA TGT GAC CTG CCT CAG ACT CAT AAC CTC AGG AAC AAG AGA
* * * * *
GCC TGC TGT CTA GGA TGT GAC CTG CCT CAT ACT CAT AAC CTC AGG AAC AAG AGA
Ala Cys Cys Leu Gly Cys Asp Leu Pro His Thr His Asn Leu Arg Asn Lys Arg
19

37
Ala Leu Thr Leu Leu Val Gln Met Arg Arg Leu Ser Pro Leu Ser Cys Leu Lys
GCC TTG ACA CTC CTG GTA CAA ATG AGG AGA CTC TCC CCT CTC TCC TGC CTG AAG
* * * * *
GTC TTC ACA CTC CTG GCA CAA ATG AGG AGA CTC TCC CCT GCT TCA TGC CTG AAG
Val Phe Thr Leu Leu Ala Gln Met Arg Arg Leu Ser Pro Val Ser Cys Leu Lys
37

55
Asp Arg Lys Asp Phe Gly Phe Pro Gln Val Lys Val Asp Ala Gln Gln Ile Lys
GAC AGG AAG GAC TTT GGA TTC CCG CAG GTG GAT GCC CAG CAG ATC AAG
* * * * *
GAC AGA AAG TAC TTT GGG TTC CCT TTG GAG AAG GTG GAT GGC CAG CAG ATC CAG
Asp Arg Lys Tyr Phe Gly Phe Pro Leu Glu Lys Val Asp Gly Gln Gln Ile Gln
55

73
Lys Pro Gln Ala Val Pro Val Leu Ser Glu Leu Thr Gln Gln Ile Leu Asn Ile
AAG CCT CAA GCC GCT CCT GTC CTG AGT GAG CTG ACC CAG GAC ATC CTG AAC ATC
* * * * *
AAG GCT CAA GCT ATC CCT GTC CTG CAT GAG CTG ACC CAG CAG ATC CTC AGC CTC
Lys Ala Gln Ala Ile Pro Val Leu His Glu Leu Thr Gln Gln Ile Leu Ser Leu
73

91
Phe Thr Ser Lys Asp Ser Ser Ala Ala Trp Asn Thr Thr Leu Leu Asp Ser Phe
TTC ACA TCA AAG GAC TCA TCT GCT TGG AAT ACA ACC CTC CTA GAC TCA TTC
* * * * *
TTC ACA TCA AAG GAG TCA TCT ACT GCT TGG GAT GCA ACC CTC CTA GAC TCA TTC
Phe Thr Ser Lys Glu Ser Ser Thr Ala Trp Asp Ala Thr Leu Leu Asp Ser Phe
91

109
Cys Asn Asp Leu His Gln Gln Leu Asn Asp Leu Gln Gly Cys Leu Met Gln Gln
TGC AAT GAC CTC CAC CAG CAG CTT AAT GAC CTG CAA GGT TGT CTG ATG CAG CAG
* * * * *
TGT AAT GAC CTC CAG CAG CAG CTG AGT GGT CTG CAA GCC TGT CTG ATG CAG CAG
Cys Asn Asp Leu Gln Gln Gln Leu Ser Gly Leu Gln Ala Cys Leu Met Gln Gln
109

127
Val Gly Val Gln Glu Phe Pro Leu Thr Gln Glu Asp Ala Leu Leu Ala Val Arg
GTG GGG GTG GAC GAA TTT CCC CTG ACC CAG GAA GAT GCC CTC GTG GCT GTG AGG
* * * * *
GTA GGG GTG CAG GAA TCT CCC CTG ACC CAG GAA GAC TCC CTA CTG GCT GTG AGG
Val Gly Val Gln Glu Ser Pro Leu Thr Gln Glu Asp Ser Leu Leu Ala Val Arg
127

145
Lys Tyr Phe His Arg Ile Thr Val Tyr Leu Arg Glu Lys Lys His Ser Pro Cys
AAA TAC TTC CAC AGG ATC ACT GTG TAC CTG AGA GAG AAG AAA CAC AGC CCC TGT
* * * * *
GAA TAC TTC CAC AGA ATC ACT GTG TAC CTG AGA GAG AAT AAA CAC AGC CCC TGT
Glu Tyr Phe His Arg Ile Thr Val Tyr Leu Arg Glu Asn Lys His Ser Pro Cys
145

163
Ala Trp Glu Val Val Arg Ala Glu Val Trp Arg Ala Leu Ser Ser Ser Ala Asn
GCC TGG GAG GTG GTC AGA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCC GCC AAT
* * * * *
GCC TGG GAG GTG GTC AAA GCA GAA GTC TGG AGA GCC CTG TCT TCC TCA GCC AAC
Ala Trp Glu Val Val Lys Ala Glu Val Trp Arg Ala Leu Ser Ser Ser Ala Asn
163

181
Val Leu Gly Arg Leu Arg Glu Glu Lys ***
GTG CTG GGA AGA CTG AGA GAA GAG AAA TGA
* * * * *
TTG ATG GGA AGA CTG AGA GAA GAA AGA AAT GAG TCC TGA
Leu Met Gly Arg Leu Arg Glu Glu Arg Asn Glu Ser ***
181

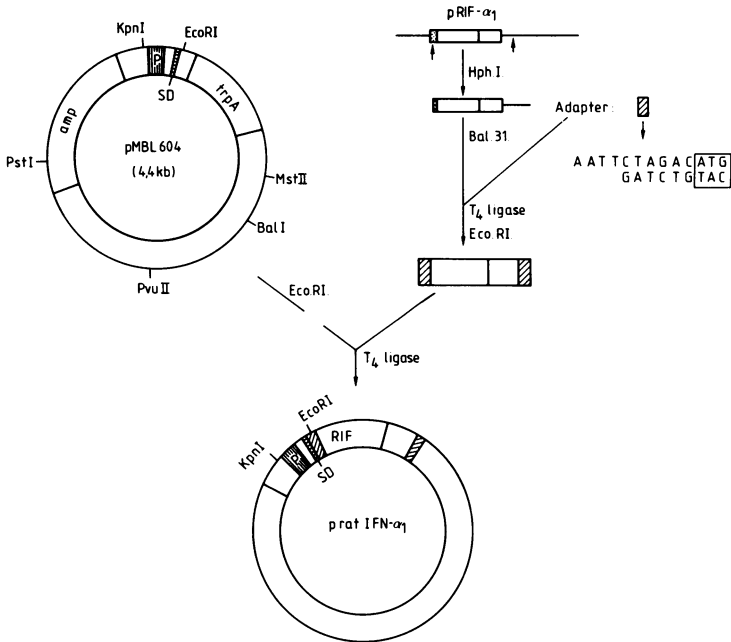
```

Figure 7. Comparison of the nucleotide and amino acid sequences of the chromosomal genes of mouse IFN- $\alpha$  (top) and rat IFN- $\alpha_1$  (bottom). The sequence of the mouse IFN- $\alpha_1$  genes was determined by Shaw *et al.* (); the sequence of the rat IFN- $\alpha_1$  gene was determined in the legends to figures V and VI. The asterisks between the two sequences indicate identical amino acid positions.

an ATG startcodon at position 130 and a TGA stopcodon at position 706. The sequence codes for a protein of 192 amino acids, most probably comprising a leader or signal peptide and the mature IFN protein (see below). Note that a 5'TATTTAA-3' sequence is present at position 30-36, a sequence which may correspond to the Goldberg-Hogness box that was reported to occur some 30 nucleotides upstream from the 5'-end of eucaryotic mRNA<sup>31</sup>; this suggests that initiation of transcription does occur around position 60, which would reinforce the conclusion that the ATG at position 130 is the initiation codon for the IFN protein, as it represents the first available ATG present in the IFN mRNA. A more direct evidence that the 192 amino acid protein indeed represents the rat IFN protein comes from a comparison of the nucleotide and amino acid sequence with the chromosomal mouse IFN- $\alpha_1$  gene<sup>25</sup> (Figure 7).

There is limited knowledge about the primary structure of mature mouse IFN- $\alpha$  proteins, however, amino terminal sequences of mouse IFN-C<sup>32</sup> allowed the identification of the mature mouse chromosomal IFN- $\alpha_1$  protein<sup>25</sup>. This gene encodes a 23 amino acid long leader peptide followed by a mature mouse IFN- $\alpha_1$  probe of 166 amino acids starting with a Cys residue at position 24. Since all cloned mature human IFN- $\alpha$  proteins<sup>10</sup> and mouse IFN- $\alpha$  proteins<sup>25</sup> start with a Cys residue, and since this gene family is highly conserved among higher vertebrates, the mature rat IFN- $\alpha_1$  protein most probably starts with the Cys residue at position 24 and counts for 169 amino acids. With this assumption, the signal peptides of mouse and rat IFN- $\alpha_1$  differ in 6 out of 23 residues (74% amino acid homology; 88% DNA sequence homology), whereas their mature proteins occupy 135 out of 166 amino acids (81%) at identical positions (DNA sequences homology is 90%). In Figure 6 we also show the position of the longest cDNA (clone I; Figure 3) with regard to the coding region of the rat IFN- $\alpha_1$  gene. Comparison of the coding region in the cDNA sequence with its chromosomal counterpart revealed 6 base pair differences, 4 of which are third letter positions not changing the amino acid composition; the two base changes leading to different amino acids are located at positions 632 (A  $\rightarrow$  G; lys  $\rightarrow$  arg) and 672 (A  $\rightarrow$  C; met  $\rightarrow$  leu), both base changes, however, can now perfectly be matched with the chromosomal mouse IFN- $\alpha_1$  sequence. These differences in nucleotide sequence of the cDNA and the chromosomal rat IFN- $\alpha_1$  gene presumably are caused by polymorphism between the two rat strains that were used for either mRNA induction (WagRij) and gene library construction (Sprague Dawley).

Expression of cloned rat IFN- $\alpha_1$  coding sequences in E. coli. For the expression in E. coli, the mature coding sequence of the chromosomal rat



**Figure 8.** Construction of rat IFN- $\alpha_1$  expression plasmid. The construction is described in the Materials and Methods section. Only the relevant HphI sites present on the 2.5 kb chromosomal EcoRI fragment are indicated.

IFN- $\alpha_1$  gene was brought under control of the portable promoter present in plasmid pMBL604 (see Materials and Methods section; Figure 8). Transformants were screened for their ability to produce biologically active rat IFN- $\alpha_1$  protein by the cytopathic effect reduction (CPE) assay<sup>22</sup>. Two transformants were found to produce rat IFN- $\alpha_1$  protein with antiviral activities of 5000 and 250 units/ml culture, respectively. Subsequent plasmid DNA analysis of both transformants revealed the exact nature of the constructions made (Table 1). In plasmid DNA from transformant #13, producing 5000 units/ml culture the initiation codon AUG is followed by the TGT (Cys) codon of mature IFN. Plasmid DNA from transformant #10 contains, one additional codon in between the ATG and TGT codons, namely the last codon of the leader peptide GGA (Gly).

**Antiviral activity in vivo.** Rats younger than 6 weeks of age can be lethally infected by numerous viruses, e.g. VSV, herpes virus and pseudorabies virus. The antiviral potency of rat IFN against a lethal dose of virus has been tested; administration of 1000 units natural rat IFN (derived

Table 1. Structure of N-terminal amino acids as deduced from the determined DNA sequence and rat IFN- $\alpha_1$  protein activity as extracted by the procedure of Derynck *et al.*<sup>3,4</sup>

clone number	DNA construction	rat IFN- $\alpha_1$ protein (units/ml)
	Met Cys	
13	AATTCTAGACATG-TGT-	5000
10	AATTCTAGACATG-GGA-TGT-	250
	Met Gly Cys	

from Sendai or NDV-induced Ratec cells) at days -1, 0, and +1 resulted in a complete protection<sup>33</sup>. When we used recombinant DNA-derived rat IFN- $\alpha_1$  in one of such experimental viral infections (e.g. pseudorabies virus) we again found complete protection of the animals against a lethal dose of virus, at IFN- $\alpha_1$  doses comparable to the natural IFN preparations. From these experiments we conclude that rat IFN- $\alpha_1$  encodes a functional rat IFN- $\alpha$ .

DISCUSSION

With 14S poly(A)mRNA from Sendai-induced Ratec cells, three partial rat IFN- $\alpha$  cDNAs have been synthesized and cloned in the *E. coli* vector pBR322. DNA sequence analysis of these recombinant plasmids revealed their common origin with respect to rat IFN- $\alpha$  mRNA subtype. To estimate the number of rat IFN- $\alpha$  genes in the rat genome Southern blots of restricted rat DNA were hybridized using labeled mouse and rat IFN- $\alpha$  DNA probes. Since the human and mouse IFN- $\alpha$  genes studied sofar do not contain restriction enzyme sites for BamHI and HindIII, and the chromosomal rat IFN- $\alpha_1$  gene (based on DNA homology with other IFN- $\alpha$  genes) contains no intron, we tentatively conclude that the rat IFN- $\alpha$  gene family consist of at least twelve members. This conclusion was reinforced, when we found an equal number of rat IFN- $\alpha$  related hybrid phages per genome length upon screening of a rat gene library constructed in phage  $\lambda$ . Since all three cDNAs were derived from the same subtype of rat IFN- $\alpha$  mRNA, a single chromosomal rat IFN- $\alpha$  gene in particular is induced when Ratec cells become infected with Sendai virus. This gene was identified by the use of labeled 3' non-coding cDNA and induced mRNA as a probe. At present we do not know whether this narrow spectrum of induced gene(s) is characteristic for this particular virus or that other viral and non-viral inducers will act in the same specific way on other genes. Characterization of the chromosomal rat IFN- $\alpha_1$  gene by DNA sequence analysis allowed a comparison with the human and mouse chromosomal counterpart. There

is more homology between the different coding regions compared to signal sequences indicating a higher degree of evolutionary constraint for the coding region<sup>25,10</sup>. The mature rat IFN- $\alpha_1$  protein is encoded by 169 amino acids, 3 amino acids longer than most of the mature human and mouse IFN- $\alpha$  genes. This length difference for the rat IFN- $\alpha_1$  protein can be easily explained by the insertion of 2A residues at position 692-693, resulting in a different translational reading frame that terminates 3 amino acids later. Surprisingly these three additional amino acids (Asn-Glu-Ser) may represent a glycosylation site for the rat IFN- $\alpha_1$  polypeptide. Expression of the coding sequence of the chromosomal rat IFN- $\alpha_1$  gene in E. coli resulted in an antiviral activity in vitro as was demonstrated by the CPE-bioassay. Therefore, if glycosylation takes place, as one would expect in the case of natural rat IFN- $\alpha$  preparations but not for E. coli rat IFN- $\alpha$ , this does not influence antiviral activity in vitro. Preliminary experiments on the antiviral activity in vivo also show that the degree of glycosylation is independent for antiviral activity. In future, the rat IFN- $\alpha_1$  expression levels obtained in E. coli enable the further characterization of possible antiviral and antitumor potency.

#### ACKNOWLEDGEMENTS

We thank C. Weissmann for the chromosomal mouse IFN- $\alpha_1$  probe, T. Sargent, B. Wallace and J. Bonner for the rat genomic library, M. Salden for help with the Xenopus oocyte assay and H. Dodemont for a gift of rat liver DNA.

#### REFERENCES

1. Isaacs, A., and Lindenmann, J. (1957) Proc. R. Soc. B 1479, 258-267.
2. Stewart, W. E. II (1980) in The Interferon System, Springer, New York.
3. Baron, S., and Dianzani, F. (1977) in Tex. Rep. Biol. Med. 35, 1-573.
4. Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K., and Weissmann, C. (1980) Nature Lond. 284, 316-320.
5. Goedel, O. V., Yelverton, E., Ullrich, A., Heyneker, H. L., Miozarri, G., Holmes, W., Seeburg, P. H., Dull, T., May, L., Stebbing, N., Crea, R., Maeda, S., McCandliss, R., Sloma, A., Tabor, J. M., Gross, M., Familletti, A. C., and Pestka, S. (1980) Nature Lond. 287, 411-417.
6. Ohno, S., and Taniguchi, T. (1981) Proc. Natl. Acad. Sci. USA 78, 5305-5309.
7. Derynck, R., Content, J., De Clercq, E., Volckaert, G., Tavernier, J., Devos, R., and Fiers, W. (1980) Nature Lond. 285, 542-547.
8. 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 Goedel, D. V. (1982) Nature Lond. 295, 503-506.

9. Nagata, S., Mantei, N., and Weissmann, C. (1980) *Nature Lond.* 287, 401-408.
10. Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. M., Ullrich, A., Yelverton, E., and Gray, P. W. (1981) *Nature Lond.* 290, 20-26.
11. Houghton, M., Jackson, I. J., Porter, A. G., Doel, S. M., Cattlin, G. H., Barber, C., and Carey, N. H. (1981) *Nucl. Acids Res.* 9, 247-266.
12. Lawn, R. M., Adelman, J., Franke, A. E., Houck, C. M., Gross, M., Najarian, R., and Goeddel, D. V. (1981) *Nucl. Acids Res.* 9, 1045-1052.
13. Sagar, A. D., Sehgal, P. B., Slate, D. L., and Ruddle, F. M. (1982) *J. Exp. Med.* 156, 744-755.
14. Gray, P. W., and Goeddel, D. V. (1982) *Nature Lond.* 298, 859-863.
15. Schellekens, H., de Reus, A., Bolhuis, R., Fountoulakis, M., Schein, C., Ecsödi, J., Nagata, S., and Weissmann, C. (1981) *Nature Lond.* 292, 775-776.
16. Edelstein, M., and Schellekens, H. (1983) *Int. J. of Cancer*, in press.
17. Schellekens, H., Van Eerd, P., de Reus, A., Weck, P. and Stebbing, N. (1982) *Antiviral Res.* 2, 313-318.
18. Schellekens, H., De Wilde, G. A., and Weimar, W. (1980) *J. Gen. Virol.* 46, 243-247.
19. Wieringa, B., Mulder, J., Van der Ende, A., Bruggeman, A., AB, G., and Gruber, M. (1978) *Eur. J. Biochem.* 89, 67-79.
20. Gurdon, J. B., Lane, C. D., Woodlans, M., and Marbaix, G. (1971) *Nature Lond.* 233, 177-179.
21. Barth, L. G., and Barth, L. J. (1959) *J. Embryol. Exp. Morphol.* 16, 183-195.
22. Cantell, K., Hirvonen, S., Mogensen, K. E., and Pyhälä, L. (1974) *In Vitro Monogr.* 3, 35-38.
23. Dijkema, R., Dekker, B. M. M., Van Ormondt, H., De Waard, A., Maat, J., and Boyer, H. W. (1980) *Gene* 10, 63-67.
24. Hanahan, D., and Meselson, M. (1980) *Gene* 10, 63-67.
25. Shaw, G. D., Boll, W., Taira, H., Mantei, N., Lengyel, P., and Weissmann, C. (1983) *Nucl. Acids Res.* 11, 555-573.
26. Southern, E. M. (1975) *J. Mol. Biol.* 98, 503-517.
27. Sargent, T. D., Wu, J. R., Sala-Trepat, J. M., Wallace, R. B., Reges, A. A., and Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* 76, 3256-3260.
28. Benton, W. D., and Davis, R. W. (1977) *Science* 196, 180-182.
29. Maxam, A. M., and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564.
30. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L., and Boyer, H. W. (1977) *Gene* 2, 95-113.
31. Gannon, F., O'Hare, K., Perrin, F., Le Pennec, J. P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Canni, B., and Chambon, P. (1979) *Nature Lond.* 278, 428-434.
32. Taira, H., Broeze, R., Jayaram, B. M., Lengyel, P., Hunkapiller, M. W., and Hood, L. E. (1980) *Science* 207, 528-530.
33. Schellekens, H., Smiers-de Vreede, E., De Reus, A., and Dijkema, R. (1982) *J. Gen. Virol.*, submitted.
34. Derynck, R., Remaut, E., Saman, E., Stanssens, P., De Clercq, E., Content, J., and Fiers, W. (1980) *Nature Lond.* 287, 193-197.