Reprinted from Nature, Vol. 285, No. 5766, pp. 542-547, June 19 1980
© *Macmillan Journals Ltd., 1980*

Isolation and structure of a human fibroblast interferon gene

Rik Derynck*, Jean Content†, Erik DeClercq‡, Guido Volckaert*, Jan Tavernier*, René Devos* & Walter Fiers*

*Laboratory of Molecular Biology, State University of Ghent, Ledeganckstraat 35, B-9000 Ghent, Belgium
†Institut Pasteur du Brabant, rue du Remorqueur 28, B-1040 Brussels, Belgium
‡Rega Institute for Medical Research, Katholieke Universiteit Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

Chimaeric plasmids containing double-stranded cDNA copies of mRNA induced in human fibroblasts by poly I·C were screened by an RNA selection method. A series of clones to which human fibroblast interferon mRNA selectively hybridized was identified. From the nucleotide sequence of the gene, the complete amino acid sequence of human fibroblast interferon was deduced. The protein is 166 amino acids long and is preceded by a 21-amino acid signal sequence.

At least three interferons with different antigenic specificity have so far been recognized. Leukocyte interferon, lymphoblastoid interferon and fibroblast interferon represent the type I interferons, whereas mitogen-induced immune interferon is a type II interferon (see ref. 1 for review). Although the structure of leukocyte and fibroblast interferons may be sufficiently similar for them to be recognized by the same receptor[2], there is evidence that they are encoded by different genes[3]. Biologically, they have a different host-cell range[1], different dose–response curves when assayed for antiviral activity on cell culture[4,5] and a varying degree of cytostatic action in several types of cells[6,7]. Both interferons have immunomodulating activities, both are able to activate 'natural killer' cells[8–10] and both have promising clinical antiviral and antitumour activities[11–15].

The differences and similarities in activity of the two interferons will be better understood when their structures are known. All that is known so far of the structure of HF-IF (human fibroblast interferon) is the sequence of the 13 NH_2-terminal amino acids[16] and the fact that it is a glycoprotein of about 20,000 molecular weight (MW)[17–20].

Two laboratories have recently used recombinant DNA technology to clone interferon cDNA[21,22]. We have been working along similar lines, and here describe the construction and isolation of hybrid plasmids containing cDNA of HF-IF. The nucleotide sequence of the mature mRNA is derived from the analysis of the cloned DNA. On the basis of this information, the total amino acid sequence of HF-IF is deduced.

Isolation of RNA and construction of chimaera plasmids

Total RNA was isolated from VGS cells, a human diploid fibroblast cell line[23]. The cells were primed with homologous human fibroblast interferon, induced for interferon production with poly I·C and superinduced with the antibiotic cycloheximide. Induction by poly I·C has been shown to induce only HF-IF in human fibroblasts, in contrast to viral induction, which results in a certain level of leukocyte interferon production[24]. Confluent VGS fibroblasts passage 23–26 (usually 20 roller bottles, 670 cm^2 each) were primed with 100 units ml^{-1} HF-IF for 16 h (ref. 23), and induced with 100 μg ml^{-1} poly I·C in the presence of 50 μg ml^{-1} cycloheximide for 4 h (ref. 25). The cells were scraped off, lysed with NP40, and the total cytoplasmic RNA was extracted with phenol. Usually 60–80 μg of total RNA was obtained from a confluent roller bottle. Purification over oligo(dT)–cellulose yielded polyadenylated mRNA representing 3–5% of the total RNA. This mRNA fraction was further enriched for interferon mRNA by sedimentation in a 5–20% sucrose gradient in 50% formamide[26]. Alternatively, the mRNA fraction was separated on a polyacrylamide gel and the fractionated RNA eluted from 2-mm gel slices[27] (the average decrease in mRNA length per successive gel slice was about 45 nucleotides). In all cases, the presence and amount of interferon mRNA were monitored by translation in *Xenopus laevis* oocytes[28], followed by assaying of interferon activity, based on reduction of the cytopathogenic effect caused by vesicular stomatitis virus (VSV). The use of human fibroblasts trisomic for chromosome 21 gave a 5–10-fold enhanced sensitivity to HF-IF in comparison with the diploid human fibroblasts[29]. The peak fractions of these gradients were further used for the synthesis of double-stranded cDNA and cloning. The *in vitro* synthesis of cDNA was carried out by sequential treatment with avian myeloblastosis virus (AMV) reverse transcriptase, RNases T1 and A, *Escherichia coli* DNA polymerase I and S_1 nuclease, essentially as described elsewhere[30]. This double-stranded cDNA was fractionated by polyacrylamide gel electrophoresis. The full-size interferon double-stranded cDNA was considered to be about 850 base pairs long on the basis of a prominent band on polyacrylamide gel with cDNA synthesized from the active mRNA, which was itself narrowly sized by gel fractionation (see above). This size estimate agrees with the sedimentation value of interferon mRNA in sucrose gradients[31]. The eluted double-stranded cDNA was elongated with homopolymeric (dT)-tails by terminal deoxynucleotidyl transferase. This tailed insert DNA was annealed to a double molar concentration of pBR322 plasmid DNA that had been cleaved in the β-lactamase gene by *Pst*I restriction endonuclease and treated with terminal transferase to add homopolymeric (dA)-tails at the cleavage sites[30]. The annealed hybrid plasmids were introduced into competent *E. coli* K12 HB101 cells[32]. About 97% of the tetracycline-resistant transformants were sensitive to carbenicillin, suggesting the presence of an insert cDNA in the *Pst*I site of the β-lactamase gene. In this way, a library of about 17,000 transformants was constructed using 270 fmol (128 ng) of tailed insert DNA.

Identification of clones containing human interferon cDNA

An RNA selection method was used to identify clones with an HF-IF cDNA insert. Plasmid DNA from the clones was coupled to diazobenzyloxymethyl (DBM)-cellulose powder[33]. Total RNA from induced VGS cells was hybridized to the DNA and the hybridized mRNA was eluted in 99% formamide[34]. The plasmid pSTNV-1 DNA, a pBR322

derivative containing the double-stranded cDNA from satellite tobacco necrosis virus (STNV) RNA[35], was added to the plasmid DNAs to be coupled; as the hybridization reaction was carried out in the presence of STNV RNA, we could evaluate the efficiency of hybridization, the extent of RNA degradation throughout the procedure and the possible presence of translational inhibitors. Half of the recovered RNA fractions were translated in the nuclease-treated rabbit reticulocyte lysate[36], followed by immunoprecipitation of the *in vitro* synthesized STNV coat protein and analysis on polyacrylamide gel (data not shown). The other half of the non-hybridized or hybridized RNA was assayed for the presence of interferon mRNA by translation in *Xenopus* oocytes followed by the antiviral interferon assay.

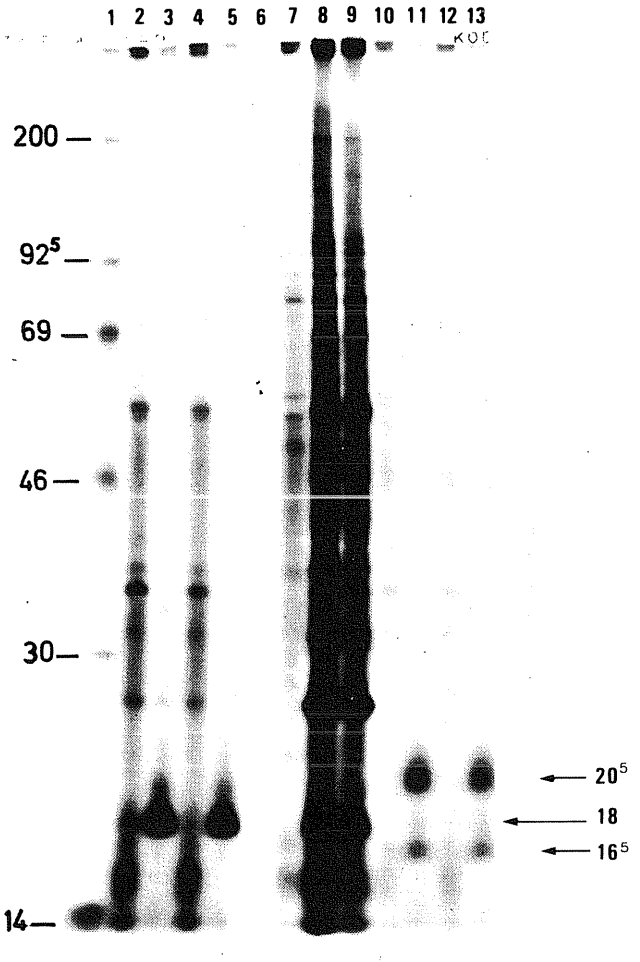


Fig. 1 Polyacrylamide gel electrophoresis of cell-free translation products from HF-IF mRNA after hybridization on different plasmid DNAs. Samples (30 µg) of total induced VGS-RNA were hybridized[34] to DBM–cellulose filters containing 3 µg DNA from pHFIF-1 or pHFIF-2. All RNA samples were divided into two equal parts and translated for 45 min at 31 C in 25-µl reaction mixtures containing 1.7×10^7 c.p.m. ^{35}S-methionine (~1,000 Ci mmol^{-1}) and 150 µg ml^{-1} calf liver tRNA in a micrococcal nuclease-treated rabbit reticulocyte lysate[36] in the absence (lanes 2–9) or presence (lanes 10–13) of 6 A_{280} ml^{-1} dog pancreas microsomes[39]. The reaction products were precipitated by addition of 2 µl of goat anti-HF-IF (200,000 HF-IF neutralizing units ml^{-1}) in the presence of 1% sodium deoxycholate and 1% NP40 for 1 h at 37 °C and 30 µl of a 10% suspension of *Staphylococcus aureus* Cowan I for 30 min at 20 °C. After extensive washing, the bacterial pellet was resuspended and boiled for 2 min in electrophoresis sample buffer, clarified for 2 min at 9,000g and applied to a 13% polyacrylamide gel. After electrophoresis, the gel was treated with En^{3}hance (NEN) and fluorographed for 5 days at −70 °C. Lane 1, ^{14}C-protein markers (Radiochemical Centre); the MW of the proteins are indicated at the left; lanes 2 and 10, pHFIF-1, non-hybridized RNA; lanes 3 and 11, pHFIF-1, hybridized RNA; lanes 4 and 12, pHFIF-2, non-hybridized RNA; lanes 5 and 13, pHFIF-2, hybridized RNA; lane 6, endogenous activity of the reticulocyte lysate; lane 7, total RNA from non-induced VGS cells (2 µg); lanes 8 and 9, total RNA from induced VGS cells (2 µg). The estimated MWs of the HF-IF precursor (lanes 3, 5) and of its putative processed, non-glycosylated and glycosylated derivatives (lanes 11, 13) are indicated at the right.

Twelve groups of 46 clones from the library were screened for the presence of an interferon cDNA-containing plasmid (Table 1). One of the two which showed a positive response was subdivided into eight groups. One of these showed a clear positive interferon response after translation of the hybridized RNA. The individual clone containing the interferon cDNA was finally detected using DBM–cellulose paper disks, allowing elution with water at 80 °C and thus reducing the RNA degradation. The bacterial clone identified in this way is designated G-HB101-pHFIF-1, the plasmid being referred to as pHFIF-1.

Other clones containing double-stranded interferon cDNA were detected using the colony hybridization technique originally described by Grünstein and Hogness[37], as modified by Hanahan and Meselson (personal communication). Restriction enzyme analysis of pHFIF-1 revealed an internal *Hin*fI fragment of about 170 base pairs. This fragment was ^{32}P-labelled by nick translation[38] and the probe was used for screening part of the library for clones related to pHFIF-1. In this way, several additional clones were selected, their plasmids being designated pHFIF-2 to pHFIF-13. In our collection an average of 1% of the clones hybridized to the 170-base pair fragment of pHFIF-1. As the mRNA used for the cloning was 20–40-fold enriched by formamide–sucrose gradient centrifugation, one may estimate that approximately 1 out of 2,000–4,000 mRNAs is HF-IF mRNA under the induction regimen used for the VGS fibroblasts. That the secondary clones, detected by colony hybridization, actually contained interferon-specific sequences was further supported by the observation that plasmid DNA of one of these clones, pHFIF-2, was equally effective in selecting interferon mRNA (data not shown). Furthermore, the mRNA hybridized to either pHFIF-1 or pHFIF-2 gave rise to the same polypeptide of about 18,500 MW after translation in the rabbit reticulocyte lysate and immunoprecipitation with antiserum against partially purified HF-IF (Fig. 1). Addition of a microsomal fraction of dog pancreas to the reticulocyte lysate[39] produced two modified forms of this polypeptide, one with an apparent MW of ~20,500 and the other ~16,500, the former corresponding in size to HF-IF secreted from induced culture cells[17–19]. The protein of lower MW suggests maturation by cleavage of a signal peptide, as will be discussed below.

Physical characterization and sequence analysis of the cDNA insert

The plasmid DNAs of a series of clones containing an interferon cDNA insert as detected by hybridization were further characterized using restriction enzymes. Most attention was given to pHFIF-1, 2, 3, 6 and 7. Using the detailed physical map of pBR322 (ref. 40), the insert length could be estimated, either by using enzymes which do not cleave the insert, such as *Msp*I (=*Hpa*II) and *Hha*I, or by summation of the length of different restriction fragments. The insert cDNA of pHFIF-1 was, on the basis of its length, considered to be incomplete. Three other plasmids, pHFIF-3, 6 and 7, all had an insert of more than 850 base pairs, approximately the expected size of a full-length cDNA. Although most restriction enzymes tested had the same number of cleavage sites in these DNAs, clear differences could be observed in the patterns of restriction fragments obtained. Detailed analysis revealed that some awkward inversions had occurred. The arrangement of these inversions all had a similar appearance (Fig. 2). In all three cases, only a single segment was transposed, for example, in pHFIF-3, the *Pst*I–*Hin*dII segment is inverted. The arrangement was confirmed and shown by nucleotide sequence analysis always to involve a cross-over point in the 5′-untranslated leader sequence. A second cross-over could conceivably have occurred in the tail region but it is also

Table 1 Outline of the screening strategy used for detecting plasmids containing interferon-specific DNA

Library (17,000 clones) →	Group (46 clones) →	Subgroup (7 or 8 clones) →	Individual clone
Plasmid DNA source	Plasmid DNA bound on DBM–cellulose	Interferon activity (log_{10} units ml^{-1})*	
		Non-hybridized RNA	Hybridized RNA
Group 0 (46 clones)	1 μg DNA per clone	0.2	0.5
Subgroup 0_1 (8 clones)	3 μg DNA per clone	0	1.2
		0	1.5
		0	0.5
Individual clone $0_1/8$ (subsequently referred to as pHFIF-1)	3 μg DNA per clone	0	1.0
		0	1.7
		0	1.2

Two groups of 46 clones were chosen arbitrarily from a total library of 17,000 clones (see scheme at top of table). As an example, the results of a positive group (group 0), subgroup (three experiments) and individual clone (also three experiments) are given. Total cytoplasmic RNA (15 μg) from induced VGS cells was used for hybridization. The non-hybridized and hybridized mRNA fractions[34] were precipitated twice with ethanol, dissolved in 2 μl water and assayed for the presence of interferon mRNA as follows. Five *X. laevis* oocytes were injected with 50 nl per oocyte[28]. After incubation, a homogenate was obtained by crushing the oocytes in 40 μl incubation medium and cleared by centrifugation at 9,000*g* for 2 min. Human fibroblast interferon was assayed by a CPE (cytopathic effect)-inhibition technique in human fibroblasts trisomic for chromosome 21 in microtitre trays. The cells were challenged with VSV (Indiana strain) and the CPE was recorded at 24 h. All assays included an internal standard of HF-IF which was itself calibrated against the NIH human fibroblast reference G023-902-527.

* The limit of detection is 0.1 log_{10} units ml^{-1}.

possible that only a single cross-over was necessary to generate the inversion. The actual reason for this phenomenon is unclear; presumably it is a cloning artefact, although we have never seen this in our other work.

Restriction fragments labelled at the 5′ end were prepared and sequenced, essentially according to Maxam and Gilbert[41]. The procedure used for *in vitro* DNA synthesis and cloning enabled us to determine the orientation of the coding strand. Indeed, the cDNA synthesis on the mRNA was initiated by a $p(dT)_{10}$ primer hybridized to the poly(A)-tail at the 3′ end of the mRNA. This results in a $(dA)_{10}$ stretch at the 3′ end of the (+)strand (the second DNA strand), followed by a dT-tract due to the elongation with terminal deoxynucleotidyl transferase. This feature allowed us to identify the 3′ end of the coding strand in several plasmids examined.

Figure 3 shows the strategy for the sequencing analysis. All sequences were read several times, either on the same strand but from different restriction sites, or on the opposite strand. These results obtained on the series of clones allowed us to deduce the total nucleotide sequence of HF-IF cDNA.

The deduced primary structure of human fibroblast interferon mRNA and protein

The primary structure of the mRNA was determined on the basis of the DNA sequence and is shown in Fig. 4. The sequence we have deduced is 828 nucleotides long, excluding the poly(A) tail at the 3′ end. One continuous reading frame was established, starting from the AUG codon at position 65 (immediately preceded by the *Hin*dII site) and ending with the stop codon UGA at position 626, followed almost immediately by the single *Bgl*II restriction site. This region is 561 nucleotides long and thus is coding for 187 amino acids. The untranslated 5′ end is 64 nucleotides long, but is presumably incomplete considering the method of *in vitro* synthesis of double-stranded cDNA. Although the total mRNA has an A+U content of 58% and the coding region an A+U content of 55% (normal values for a eukaryotic mRNA), the 5′-untranslated sequence is rather low in these bases—48% A+U. The 3′-untranslated region, excluding the poly(A) tail, is 203 nucleotides long and can be divided into two segments.

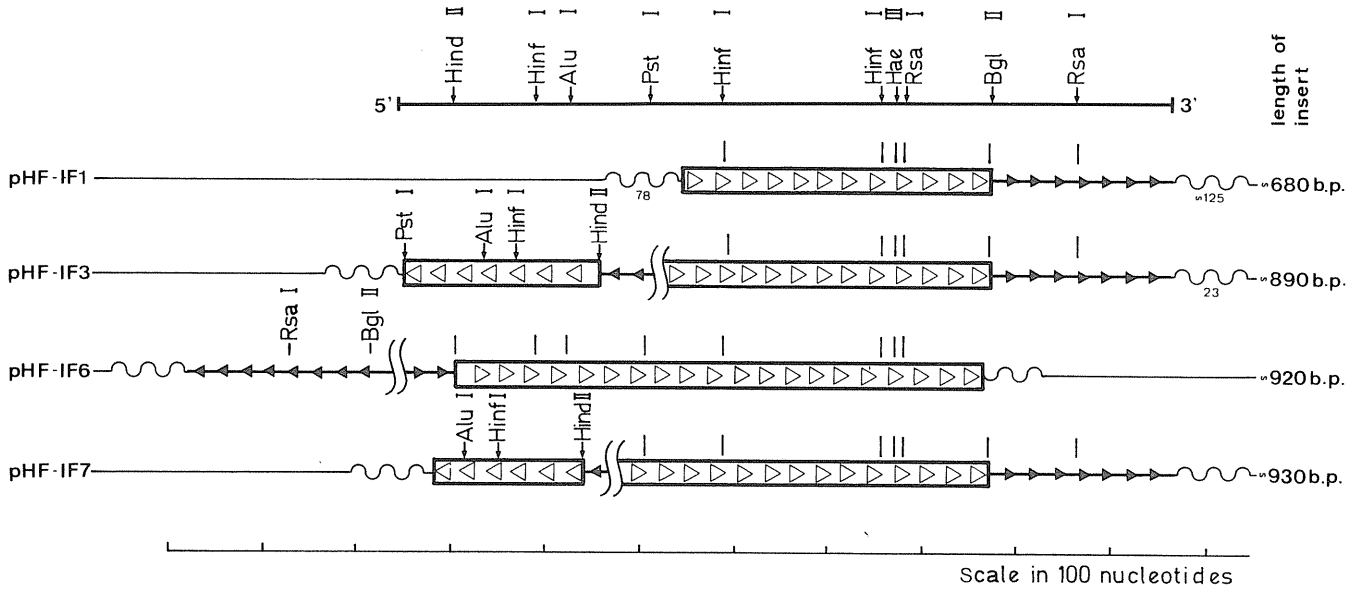
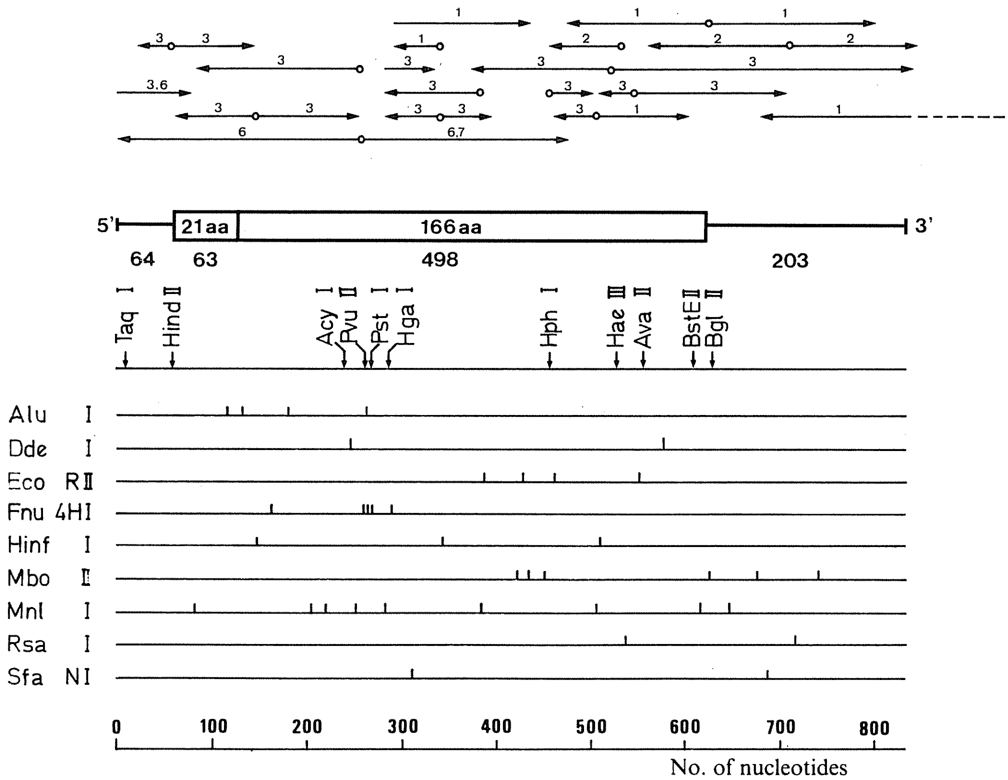


Fig. 2 Schematic representation of the organization of the cDNA inserts in pHFIF-1, 3, 6 and 7. The upper line represents a complete cDNA copy of interferon mRNA, with some reference restriction sites, and can be aligned with the more detailed map in Fig. 3. The dA·dT tails are indicated by a wavy line (the length, when known, is given underneath). The boxed segments indicate the translated sequence of the mRNA and the triangles indicate the 5′→3′ direction of the coding strand of the insert DNA, in the translated region as well as in the 5′ and 3′ untranslated regions. An interruption in the base line of the maps indicates the cross-over regions of the inversion. b.p., Base pairs.

Fig. 3 Restriction map and sequencing strategy for the HF-IF cDNA gene. The central diagram shows the organization of the mRNA for human fibroblast interferon, with the 5'-untranslated region, the boxed translated sequence with the presumed signal peptide and the 3'-untranslated segment, the number of amino acids (aa) and nucleotides being indicated for each region. At the top, the regions covered by sequencing are indicated; the open dots correspond to the ^{32}P-labelled 5' ends. The numbers above the arrows refer to the different clones analysed. Using 0.5-mm gels, it was occasionally possible to read over 300 nucleotides from the labelled site. Below is a restriction endonuclease map of the double-stranded cDNA, constructed by computer search on the basis of the complete nucleotide sequence, many of these sites being experimentally verified. Enzymes cutting the double-stranded cDNA only once are indicated on the top line and enzymes cleaving at multiple sites are indicated individually below. Recognition sequences of restriction enzymes are listed by Roberts[56].



-20 -10

MET-THR-ASN-LYS-CYS-LEU-LEU-GLN-ILE-ALA-LEU-LEU-

GCAA CCUUUCGAAG CCUUUGCUCU GGCACAACAG GUAGUAGGCG ACACUGUUCG UGUUGUCAAC **AUG** ACC.AAC.AAG.UGU.CUC.CUC.CAA.AUU.GCU.CUC.CUG. 100

-1 +1 10 20

LEU-CYS-PHE-SER-THR-THR-ALA-LEU-SER-MET-SER-TYR-ASN-LEU-LEU-GLY-PHE-LEU-GLN-ARG-SER-SER-ASN-PHE-GLN-CYS-GLN-LYS-LEU-LEU-

UUG.UGC.UUC.UCC.ACU.ACA.GCU.CUU.UCC.AUG.AGC.UAC.AAC.UUG.CUU.GGA.UUC.CUA.CAA.AGA.AGC.AGC.AAU.UUU.CAG.UGU.CAG.AAG.CUC.CUG. 190

30 40 50

TRP-GLN-LEU-ASN-GLY-ARG-LEU-GLU-TYR-CYS-LEU-LYS-ASP-ARG-MET-ASN-PHE-ASP-ILE-PRO-GLU-GLU-ILE-LYS-GLN-LEU-GLN-GLN-PHE-GLN-

UGG.CAA.UUG.AAU.GGG.AGG.CUU.GAA.UAC.UGC.CUC.AAG.GAC.AGG.AUG.AAC.UUU.GAC.AUC.CCU.GAG.GAG.AUU.AAG.CAG.CUG.CAG.CAG.UUC.CAG. 280

60 70 80

LYS-GLU-ASP-ALA-ALA-LEU-THR-ILE-TYR-GLU-MET-LEU-GLN-ASN-ILE-PHE-ALA-ILE-PHE-ARG-GLN-ASP-SER-SER-SER-THR-GLY-TRP-ASN-GLU-

AAG.GAG.GAC.GCC.GCA.UUG.ACC.AUC.UAU.GAG.AUG.CUC.CAG.AAC.AUC.UUU.GCU.AUU.UUC.AGA.CAA.GAU.UCA.UCU.AGC.ACU.GGC.UGG.AAU.GAG. 370

90 100 110

THR-ILE-VAL-GLU-ASN-LEU-LEU-ALA-ASN-VAL-TYR-HIS-GLN-ILE-ASN-HIS-LEU-LYS-THR-VAL-LEU-GLU-GLU-LYS-LEU-GLU-LYS-GLU-ASP-PHE-

ACU.AUU.GUU.GAG.AAC.CUC.CUG.GCU.AAU.GUC.UAU.CAU.CAG.AUA.AAC.CAU.CUG.AAG.ACA.GUC.CUG.GAA.GAA.AAA.CUG.GAG.AAA.GAA.GAU.UUC. 460

120 130 140

THR-ARG-GLY-LYS-LEU-MET-SER-SER-LEU-HIS-LEU-LYS-ARG-TYR-TYR-GLY-ARG-ILE-LEU-HIS-TYR-LEU-LYS-ALA-LYS-GLU-TYR-SER-HIS-CYS-

ACC.AGG.GGA.AAA.CUC.AUG.AGC.AGU.CUG.CAC.CUG.AAA.AGA.UAU.UAU.GGG.AGG.AUU.CUG.CAU.UAC.CUG.AAG.GCC.AAG.GAG.UAC.AGU.CAC.UGU. 550

150 160

ALA-TRP-THR-ILE-VAL-ARG-VAL-GLU-ILE-LEU-ARG-ASN-PHE-TYR-PHE-ILE-ASN-ARG-LEU-THR-GLY-TYR-LEU-ARG-ASN

GCC.UGG.ACC.AUA.GUC.AGA.GUG.GAA.AUC.CUA.AGG.AAC.UUU.UAC.UUC.AUU.AAC.AGA.CUU.ACA.GGU.UAC.CUC.CGA.AAC **UGA** AGAUCUCCUA GCCUG 643

UGCCU CUGGGACUGG ACAAUUGCUU CAAGCAUUCU UCAACCAGCA GAUGCUGUUU AAGUGACUGA UGGCUAAUGU ACUGCAUAUG AAAGGACACU AGAAGAUUUU GAAAU 753

UUUUA UUAAAUUAUG AGUUAUUUUU AUUUAUUUAA AUUUUAUUUU GGAAAAUAAA UUAUUUUUGG UGCAAAAGUC AAAAAAAAA$_n$ ...

Fig. 4 Nucleotide sequence of the human fibroblast interferon mRNA and corresponding amino acid sequence. The nucleotide sequence is presented as the mRNA sequence, derived from analysis of the double-stranded cDNA in several plasmids; the actual mRNA may be longer at the 5' end. The initiation and termination codons are shown in heavily outlined boxes. The AAUAAA sequence, presumably present near the 3' end in all polyadenylated eukaryotic cellular mRNAs, is shown in a dashed box. The nucleotide numbering is given at the right. The nucleotide sequence is translated into an amino acid sequence consisting of a 21-amino acid signal peptide, from the initiation codon up to the vertical arrow, and the mature polypeptide. The boxed amino acid sequence, starting from the vertical arrow, indicates the NH_2-terminal sequence derived from direct analysis of the protein[16].

	U	C	A	G	
U	Phe 4 5	Ser 1 1	Tyr 4 6	Cys 2 1	U
	Leu 3		Ochre	Opal ⊙	C
			Amber	Trp 3	A
					G
C	Leu 3 6 2 10	Pro 1	His 3 2	Arg 1	U
			Gln 3 8		C
					A
					G
A	Ile 5 4 2	Thr 2 3 2	Asn 4 8	Ser 2 5	U
	Met 4		Lys 4 7	Arg 5 5	C
					A
					G
G	Val 1 3 1	Ala 2 3 1	Asp 2 3	Gly 1 1 2 2	U
			Glu 5 8		C
					A
					G

Fig. 5 Codons used in human fibroblast interferon mRNA. The numbers refer to the frequency with which each triplet is used.

The region proximal to the UGA termination codon has a normal A+U content, whereas the distal portion is extremely A+U rich, a feature also observed, although not to this extent, in the mRNA of rat growth hormone[42]. Indeed, starting from position 744, only two C and eight G residues are present in a total of 84 nucleotides. A striking feature of this same region is the presence of many $A(U)_{2-5}$ stretches. The AAUAAA sequence presumably present in all polyadenylated eukaryotic cellular mRNAs and thought to be involved in processing or polyadenylation of the mRNA[43], is found some 20 nucleotides before the often observed C residue immediately preceding the poly(A) tail.

As in other vertebrate DNA and RNA sequences characterized so far, the dinucleotide CG is very rare in the mRNA for HF-IF. Only one of the six occurring CG dinucleotides is used within a codon (Fig. 5). This deficiency is reflected in the strong preference for AGG and AGA as codons for arginine, a preference which has also been observed in the genome of SV40 (ref. 44) and polyoma[45] and the haemagglutinin gene of a human influenza strain[46], but not, however, in the mRNA for rat growth hormone[42] or human chorionic somatomammotropin[47]. The apparently strong preference for CUG as codon for leucine, which seems to be a general phenomenon in eukaryotic cellular mRNAs but not in the mRNAs of SV40 (ref. 44) or polyoma[45] virus, is also found in HF-IF mRNA. There is also preference for AGU and AGC as codons for serine.

The amino acid sequence of HF-IF, deduced from the nucleotide sequence of the coding region, is also presented in Fig. 4. The 13 NH_2-terminal amino acids of mature HF-IF were recently directly determined by protein microsequencing techniques[16]. A corresponding nucleotide sequence begins at nucleotide 128; therefore, we conclude that this represents the NH_2-terminus of the mature HF-IF. It is preceded by a segment of 21 amino acids starting with an AUG at position 65. As interferons are secretory proteins, we postulate that this region is a signal peptide, cleaved off during or after transport of the nascent protein across the membrane. Indeed, it has been well established that the majority of secretory proteins start with a hydrophobic signal peptide which is subsequently removed[48]. The hydrophobicity of this segment, mainly of the central part, and its length of 21 amino acids are consistent with the known properties of a signal peptide[48,49]. Also, as generally observed[48,49], this signal peptide ends with an amino acid (serine in this case) of lower MW than the first NH_2-terminal amino acid of the cleaved protein. Direct evidence for this signal peptide has been obtained experimentally: addition of the microsomal fraction of dog pancreas to the *in vitro* translation mix, followed by immunoprecipitation with anti-interferon antiserum, results in a protein with an apparently lower MW than the unprocessed protein, indicating that a segment has been cleaved off (Fig. 1).

The mature HF-IF contains 166 amino acids on the basis of the deduced protein sequence. The amino acid composition is in reasonable agreement with the composition as determined by direct analysis[16,19]. It has a remarkably low content of proline and is very rich in the hydrophobic amino acids leucine and isoleucine, and also in tyrosine. The intrinsic hydrophobicity of HF-IF is well known and is also revealed by its interaction with several ligands[50]. Three cysteines are present in HF-IF, which means that there must be at least one free thiol group.

The mature fibroblast interferon has been shown to be a glycoprotein[17–19]. However, direct localization and characterization of the carbohydrate moieties on the polypeptide have not been described. Attachment of carbohydrate through *N*-glycosidic linkage is known to occur on the asparagine in the triplets Asp-X-Ser or -Thr, the presence of this sequence being a necessary but not a sufficient condition for glycosylation[51]. Only one asparagine allowing this type of glycosylation is found in fibroblast interferon, at position 80 of the amino acid sequence. Alternatively, or additionally, *O*-glycosidically linked oligosaccharides could be attached to serine and/or threonine residues, a good example of this being the human erythrocyte glycophorin[52].

We are now in a position to reconstruct a plasmid with the total HF-IF genetic information under a prokaryotic transcription signal and to test for its expression in a bacterial system.

After submission of this manuscript, we identified a clone, pHFIF-21, having a full-length insert without rearrangements. Also we learned that Taniguchi *et al.*[53] have also determined the nucleotide sequence of a cloned HF-IF gene. The deduced amino acid sequence is identical to that reported here and shows the homology (see accompanying paper[54]) with the sequence of human leukocyte interferon[55].

We thank L. De Wit, W. Zeegers and E. De Vleeschouwer for technical assistance, and Drs J. Beard, P. De Somer, W. Kuziel, E. Messens, E. Remaut, M. Verhoeyen, J. van Emmelo and M. Ysebaert for materials, help and discussions. J.T. holds a predoctoral fellowship from the IWONL of Belgium.

Received 8 April; accepted 29 April 1980.

1. Stewart, W. E. II *The Interferon System* (Springer, Berlin, 1979).
2. Wiranowska-Stewart, M. & Stewart, W. E. II *J. gen. Virol.* **37**, 629–634 (1977).
3. Cavalieri, R. L., Havell, G. A., Vilcek, J. & Pestka, S. *Proc. natn. Acad. Sci. U.S.A.* **74**, 3287–3291 (1977).
4. Vilcek, J., Havell, G. A. & Yamazaki, S. *Ann. N.Y. Acad. Sci.* **284**, 703–710 (1977).
5. Edy, V.-G., Billiau, A. & De Somer, P. *J. gen. Virol.* **31**, 251–255 (1976).
6. Einhorn, S. & Strander, H. *J. gen. Virol.* **35**, 573–577 (1977).
7. Kuwata, T., Fuse, A., Suzuki, N. & Morinaga, N. *J. gen. Virol.* **43**, 435–439 (1979).
8. Herberman, R. R., Ortaldo, J. R. & Bonnard, G. D. *Nature* **277**, 221–223 (1979).
9. Zarling, J. M., Eskra, L., Borden, E. C., Horoszewicz, J. & Carter, W. A. *J. Immun.* **123**, 63–70 (1979).
10. Huddlestone, J. R., Merigan, T. C. Jr & Oldstone, M. B. A. *Nature* **282**, 417–419 (1979).
11. De Smyter, J. *et al. Lancet* **ii**, 645–647 (1976).
12. Horoszewicz, J. S. *et al. Cancer Treat. Rep.* **62**, 1899–1906 (1978).
13. Sundmacher, R., Cantell, K., Skoda, K., Hallermann, C. & Neumann-Haefelin, D. *Albrecht v. Graefes Arch. Opthal.* **208**, 229–233 (1978).
14. Dolen, J. G. *et al. Am. J. Med.* **67**, 127–131 (1979).
15. Treuner, J. *et al. Lancet* (in the press).
16. Knight, E. Jr, Hunkapiller, M. W., Korant, B. D., Hardy, R. W. F. & Hood, L. E. *Science* **207**, 525–526 (1980).
17. Knight, E. Jr *Proc. natn. Acad. Sci. U.S.A.* **73**, 520–523 (1976).
18. Havell, E. A., Yamazaki, S. & Vilcek, J. *J. biol. Chem.* **252**, 4425–4427 (1977).
19. Tan, Y. H., Barakat, F., Berthold, W., Smith-Johannsen, H. & Tan, C. *J. biol. Chem.* **254**, 8067–8073 (1979).
20. Heine, J. W., De Ley, M., Van Damme, J., Billiau, A. & De Somer, P. *Ann. N.Y. Acad. Sci.* (in the press).
21. Nagata, S. *et al. Nature* **284**, 316–320 (1980).
22. Taniguchi, T. *et al. Proc. Jap. Acad.* **B55**, 464–469 (1979).
23. Billiau, A., Joniau, M. & De Somer, P. *J. gen. Virol.* **19**, 1–8 (1973).
24. Hayes, T. G., Yip, Y. K. & Vilcek, J. *Virology* **98**, 351–363 (1979).
25. Reynolds, F. H. Jr, Premkumar, E. & Pitha, P. M. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4881–4885 (1975).
26. Pawson, T., Harvey, R. & Smith, A. E. *Nature* **268**, 416–420 (1977).
27. Cleveland, D. W., Kirschner, M. W. & Cowan, N. J. *Cell* **15**, 1021–1031 (1978).
28. Lebleu, B. *et al. Biochem. biophys. Res. Commun.* **82**, 665–673 (1978).
29. Tan, Y. H., Schneider, E. L., Tischfield, J., Epstein, C. J. & Ruddle, F. H. *Science* **186**, 61–63 (1974).
30. Devos, R., van Emmelo, J., Contreras, R. & Fiers, W. *J. molec. Biol.* **128**, 595–619 (1979).
31. Sehgal, D., Lyles, D. & Tamm, I. *Virology* **89**, 186–198 (1978).
32. Lederberg, E. M. & Cohen, S. N. *J. Bact.* **119**, 1072–1074 (1974).
33. Goldberg, M. L., Lifton, R. P., Stark, G. R. & Williams, J. G. *Meth. Enzym.* **68**, 206–219 (1980).
34. Paucha, E. & Smith, A. E. *Cell* **15**, 1011–1020 (1978).
35. Van Emmelo, J., Devos, R., Ysebaert, M. & Fiers, W. *J. molec. Biol.* (in the press).
36. Pelham, R. B. & Jackson, R. J. *Eur. J. Biochem.* **67**, 247–256 (1976).
37. Grünstein, M. & Hogness, D. S. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965 (1975).
38. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. *J. molec. Biol.* **13**, 237–251 (1977).
39. Shields, D. & Blöbel, G. *J. biol. Chem.* **253**, 3753–3756 (1978).

40. Sutcliffe, J. G. *Nucleic Acids Res.* **5,** 2721–2728 (1978).
41. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74,** 560–564 (1977).
42. Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270,** 486–494 (1977).
43. Proudfoot, N. J. & Brownlee, G. G. *Nature* **263,** 211–214 (1976).
44. Fiers, W. *et al. Nature* **273,** 113–120 (1978).
45. Soeda, E., Arrand, J. R., Smolar, N., Walsh, J. E. & Griffin, B. E. *Nature* **283,** 445–453 (1980).
46. Min Jou, W. *et al. Cell* **19,** 683–696 (1980).
47. Shine, J., Seeburg, P. H., Martial, J. A., Baxter, J. D. & Goodman, H. M. *Nature* **270,** 494–499 (1977).
48. Blobel, G. *et al. Soc. exp. Biol. Symp.* **33,** 9–36 (1979).
49. Habener, J. F. *et al. Proc. natn. Acad. Sci. U.S.A.* **75,** 2616–2620 (1978).
50. Davey, M. W., Sulkowski, E. & Carter, W. A. *J. biol. Chem.* **251,** 7620–7625 (1976).
51. Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. in *The Glycoproteins: Their Composition, Structure and Function* (ed. Gottschalk, A.) 450–490 (Elsevier, Amsterdam, 1972).
52. Tomita, M. & Marchesi, V. *Proc. natn. Acad. Sci. U.S.A.* **72,** 2964–2968 (1975).
53. Taniguchi, T. *et al. Gene* **10,** 11–15 (1980).
54. Taniguchi, T. *et al. Nature,* **285,** 547–549 (1980).
55. Mantei, N. *et al. Gene* **10,** 1–10 (1980).
56. Roberts, R. J. *Gene* **8,** 329–343 (1980).

Printed in Great Britain by Macmillan Production Ltd., Basingstoke, Hampshire