Volume 9 Number 11 1981 Nucleic Acids Research

The structure of a thirty-six kilobase region of the human chromosome including the fibroblast interferon gene IFN-β

Gerhard Gross, Ulrich Mayr, Wolfgang Bruns, Frank Grosveld+, Hans-Henrik M.Dahl+ and John Collins*

Gesellschaft für Biotechnologische Forschung, Abt. Genetik, Mascherodor Weg 1, D-3300 Braunschweig, GFR, and +National Institute for Medical Research, The Ridgeway, Mill Hill, London, UK

Received 7 April 1981

ABSTRACT

The isolation of a human genomic cosmid hybrid containing the interferon ß gene has recently been reported (Gross et al., 1981). This hybrid was mapped using single and double digests and cross-hybridisation with the sub-cloned EcoRI and BglII fragments. Purified fragments and subclones were used as hybridisation probes against chromosomal "Southern" blots to show that at least half of the region has been cloned without alteration. This cannot at present be confirmed for the rest of the region due to the presence of highly repetitive DNA on these cloned fragments. Sequencing data on the 5'-end of the fibroblast IFN-ß gene shows a high density of direct and inverted repeats. The IFN-ß mRNA coding region contains no intron, although the possibility of other transcription starts is not ruled out. The cloned region shows no similarities to known genomic clones containing IFN-α genes.

INTRODUCTION

Human fibroblast interferon HuIFN-ß has an anti-viral activity and can also stimulate natural killer cell action against neoplastic cells 1,2,3. The Interferon ß gene(s) belong(s) to a rare class of eukaryotic genes for which the immediate induction of transcription in response to specific inducers such as poly rI: poly rC can be demonstrated 4,5,6. Recent rindings indicate that two IFN-mRNAs exist which are at the most only distantly related but are co-ordinately induced in human fibroblasts 7,8. It therefore seemed of interest to isolate the chromosomal region for the IFN-ß gene in order to study the structure ot the transcription unit and the possibility that adjacent transcription units are co-ordinately expressed.

The isolation of IFN-ß cDNA clones has recently been

demonstrated by three other groups [9,10,12]. The isolation of a clone containing a part of the IFNß cDNA and the use of this DNA as a probe in the isolation of a genomic hybrid from a cosmid gene bank has recently been described[11]. Evidence was found that the transcribed region contains no intron. Neighbouring regions were found to have no homology to the IFN-ß gene region.

A prerequisite for further studies on gene regulation in this region of the genome is the production of an accurate physical map. Such a map is presented here including the distribution of highly repetetive DNA. Such a map will also facilitate comparison with similar data obtained by other groups for other IFN genes.

MATERIALS AND METHODS

Production and screening of a genomic cosmid gene bank.

The detailed description for the production of the cosmid gene bank will be presented elsewhere (F. Grosveld, H.-H.M.Dahl, E. de Boers and R.A. Flavell, Gene, in the press). The bank consisted of 1.5×10^5 clones obtained using phosphatase treated BamHI cleaved cosmid pJB8(Julian Berg and David Ish-Horowisch, unpublished) and MboI partially cleaved human placental DNA. The BamHI cloning site in pJB8 is immediately flanked by EcoRI sites so that cloned inserts can easily be removed for subcloning and mapping.

DNA blot hybridisation.

DNA blotting and hybridisation were carried out according to Southern[18]. Nick-translation to make ^{32}P-labelled probes was carried out according to Miniatis et. al.[20]. The bank was simultaneously screened with three different labelled probes of which only one will be discussed here. One hundred suspected positive colonies were picked and rescreened by colony hybridisation with the individual probes. One colony gave a strong hybridisation with the IFN-ß cDNA probe. After two further dilutions and repicking, a single colony was isolated and designated pCos IFNß.

Isolation of purified EcoRI fragments of pCosIFNß.

Fragments form 300µg of EcoRI cleaved pCosIFNß were electro-eluted from 1 % agarose gel slices onto dialysis tubing. The eluted DNA was washed off with 0.02 M Tris-HCl pH 8.0. Each DNA sample was concentrated on about 50 µg of DEAE cellulose, washed with 4ml 0.1 M NaCl, 0.2 M Tris-HCl and eluted in 0.4 ml 1.5 M NaCl, 0.2 M Tris-HCl pH 8.0. The fragments were precipitated in 70% ethanol and resuspended in appropriate buffers for either restriction mapping or nick-translation.

Subcloning of EcoRI and BglII fragments of pCosIFNß.

To facilitate restriction mapping and DNA sequencing,EcoRI fragments of pCosIFNß were subcloned into pBR325 or pACYC184 screening for chloramphenicol sensitivity. BglII fragments were cloned into alkaline phophatase treated pJC80[20]. All BglII fragments were found amongst the hybrids by rapid screening of fragment sizes according to Birnboim and Doly[16]. BglII-B1 and BglII-B2 fragments were distinguished from one another by their PstI restriction pattern.

RESULTS

Construction of the restriction map

Fortuitously the HpaI, HindIII double and single digests allowed the constuction of an unambiguous map on which the known sites for the vector could be located. Southern hybridisation with an IFN-ß cDNA probe showed positive with HindIII-B and NpaI-B (Figure 1), thus defining the IFN-ß gene within the region 10.9 to 16 kb on the map (Figure 2). Southern hybridisation with the IFN-ß-cDNA, EcoRI-F and pJC80 (homologous to pJB8) labelled probes against single and double digests of pcosIFNß allowed the construction of nearest neighbour maps for the sites surrounding the IFN-ß gene and the vector. This analysis was facilitated by the complete absence of cross-hybridisation between the EcoRI-F fragment and other regions on the hybrid plasmid. Since this was also true when the stringency of the hybridisation was lowered (0.3 M NaCl in the final 65° C wash) it can be concluded that the cloned region contains but a single IFN-ß gene copy and that no "pseudo-genes" are present. Further

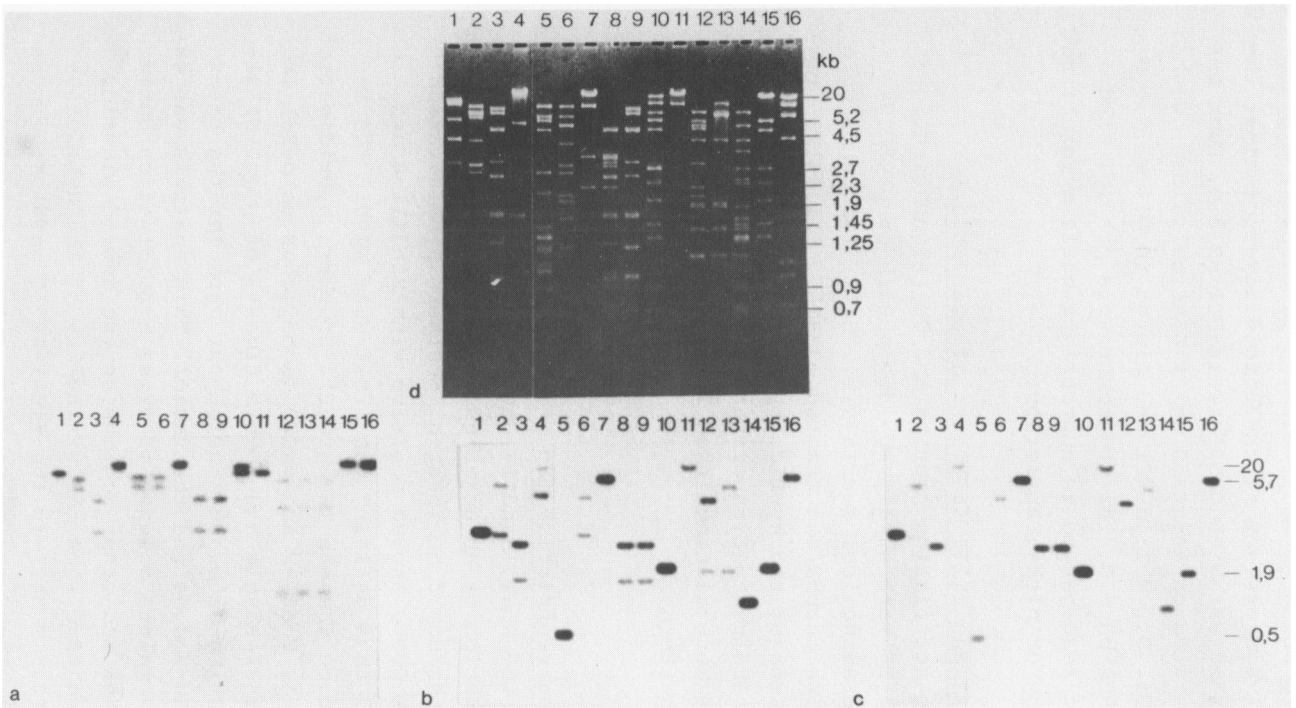


Figure 1. Southern blot hybridisation against pCosIFNß fragments using as labelled probes: a) total human genomic DNA, b) purified EcoRI-F fragment of pCosIFNß, c) a cDNA fragment containing the 3'-end of the IFN-ß gene [11]. The original gel used for the blottings is shown in the upper part of the figure (d) to the same scale. The numbered lanes correspond to cleavage by 1. SstI, 2. BglII, 3. PstI, 4. EcaI, 5. BglII + EcoRI, 6. BglII + HpaI, 7. HpaI, 8. PstI + HpaI, 9. PstI + KpnI, 10. KpnI + EcoRI, 11. KpnI, 12. PvuII + HpaI, 13. PvuII, 14. EcoRI + PvuII, 15. EcoRI, 16. HindIII.

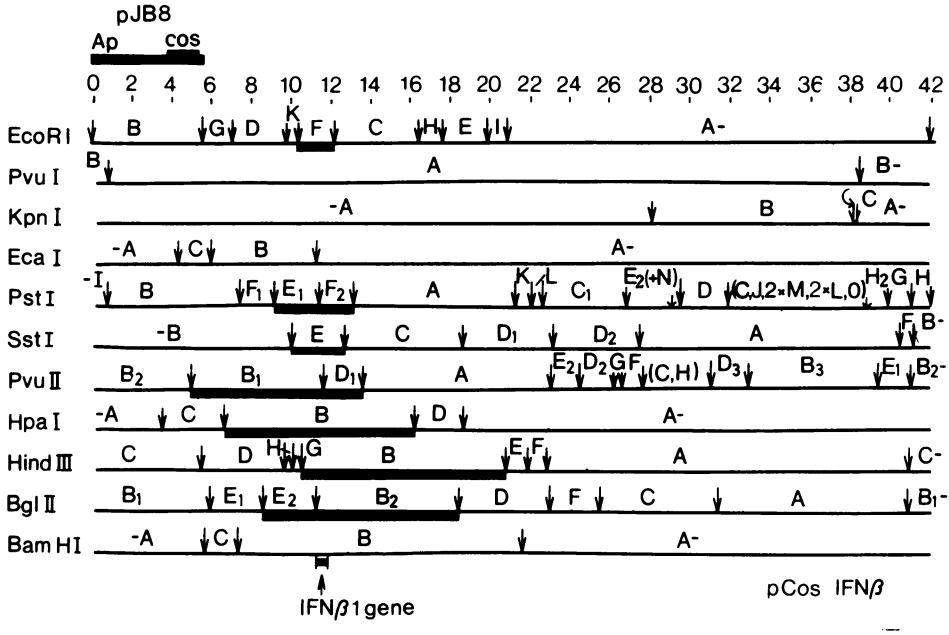


Figure 2. Restriction map of pCosIFNß. The IFN-ß gene is transcribed from right to left. The position of the cloning vector (EcoRI-B fragment) is shown in the upper left corner. Fragments hybridising with EcoRI-F are underlined. The sizes of the fragments are listed in Table 1.

Southern blot hybridisation using purified EcoRI fragments or BglII subclones as probes against multiple restriction digests allowed the completion of all the maps shown in Figure 2 with the exception of those PstI, SstI and PvuII sites located between 24 and 42 kb on the map. These maps were finally resolved by single and double cleavage of the BglII-A, -C and -F clone DNA. The ambiguities in fragment orientation for the very small PstI fragments and PvuII -C and -H have not yet been resolved. During this mapping some homology between the EcoRI-A and -G fragments was taken into consideration. The sizes of fragments measured against known and pBR322 markers are estimated as about ± 4 %. Fragments larger than 11 kb may be subject to larger errors. Some fragments smaller than 100 bp may have gone undetected.

Distribution of repetitive DNA

To map the distribution of highly repetetive DNA within the cloned region, human chromosomal DNA was used as a labelled hybridisation probe against Southern blots of pCosIFNß fragments (Figure 1a). It can be seen that strongly hybridising fragments are located in the right half of the map shown in figure 2(e.g. SstI-A, BglII-, -C and a little to -F), being mostly contained within the KpnI-B fragment. PstI-C and -D, and PvuII-B3, -C and -E1 also hybridise showing the relatively large region which contains such repetitive DNA. From the restriction map only one region contains more than copy of a particular fragment, namely the double PstI-L and -M fragments. The presence of highly repetetive DNA in these regions prevents the analysis of the fidelity of these regions when compared to the original chromosome through Southern blot analysis of total chromosome.

Fine structure of the IFNß gene region

The purified EcoRI-F fragment was analysed by restriction mapping with HincII, BglII, RsaI (Figure 3) and HinfI (not shown). This analysis yielded a restriction map as shown in Figure 4. The restriction map of the IFNß-cDNA[9,10] can be superimposed on this map (Figure 4), implying that, at least for the translated region no Intron could be present[21,22,23]. This is in agreement with other recent reports. Sequencing of various regions within the transcribed region agreed with the sequences of cDNA according to Derynck et al.[10]. We extended our sequencing into the region 5'to the coding sequence. This extends beyond the sequences contained in the hybrid isolated by Tavernier et al.[22] and Lawn et. al.[23]. The homology of this region to the comparable sequence 5'distal to the IFN-1 gene [13] is shown in Figure 5. Within the 157 bp compared (-63 to -220) 78 can be aligned by introducing a small number of gaps into the two sequences. Some of the homologies are quite extensive, such as the GTGAAAGTGG block at -140 (-73 with respect to the mRNA start). This so-called "-70 region" is found to be highly conserved for many eukaryotic genes [19]. In this example there is a 7 out of 9 base homology with the

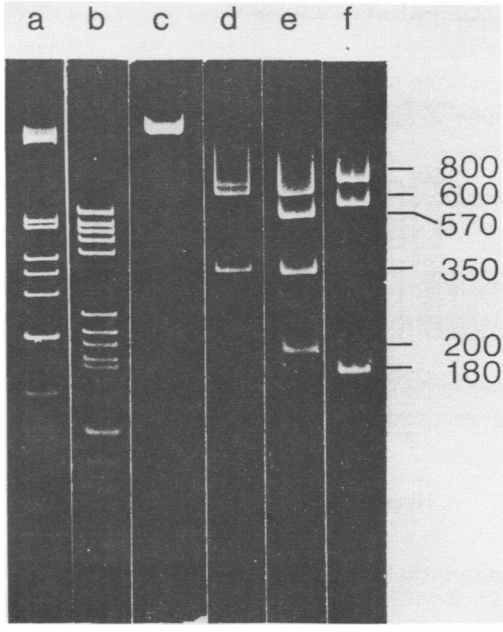
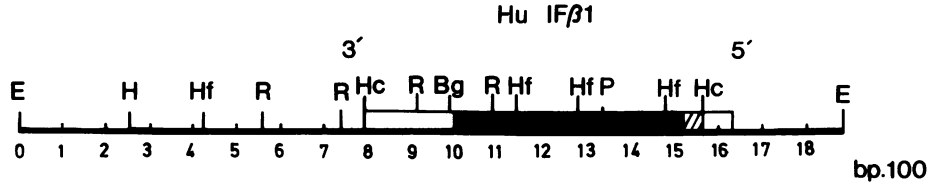


Figure 3. Restriction pattern (4-12% acrylamide gel) of the purified EcoRI-F fragment c) uncut, d) HincII, e) BglII + HincII, f) RsaI cleaved. Marker DNA: a)pBR322 + HinfI,b) pBR322 + HaeIII. Sizes in base pairs.

Adenovirus 1A early gene -73 region. The Hogness box differs markedly between the ß sequences. The ß Hogness box TATAAAT is identical with that of Sea Urchin and Drosophila H2A genes. The



R= RsaI, E= EcoRI, Bg= BglII H= HindIII, Hf= HinfI, Hc= HincII

Figure 4. Restriction map of the EcoRI-F fragment. The cDNA sequence according to Derynck et al.[10] is superimposed in this map as follows: open box, non-translated; black box, coding region for mature IFN-ß; shaded box, pre-IFN-ß coding sequence.

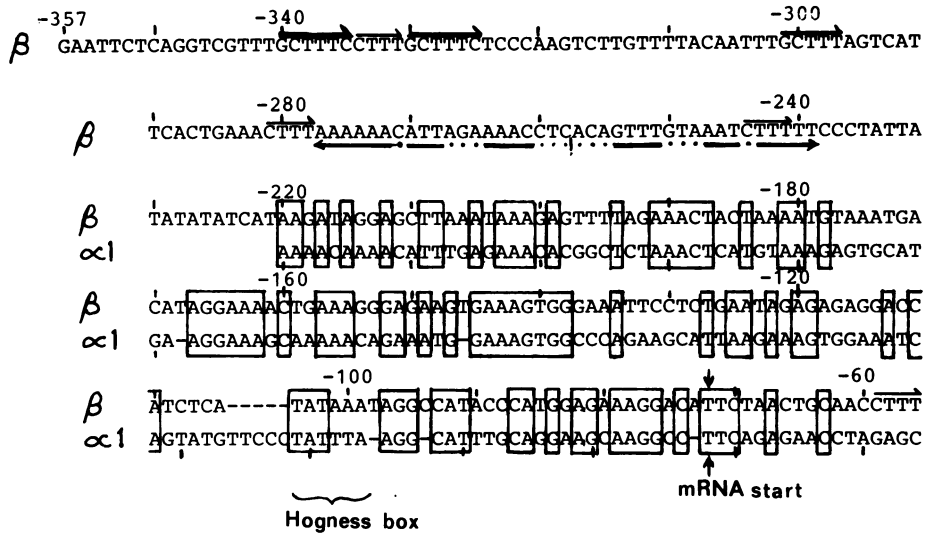


Figure 5. The sequence of the terminal 300 bp 5-prime to the IFN-ß gene within EcoRI(Figure 4). A comparison of our ß sequence with the corresponding 5-prime region of the HuIFN- 1 gene (-55 to -220 bp; Nagata et al.[13] has been made by boxing regions of homology. Gaps, shown as dashes, have been introduced to optimise alignment. Sequencing was carried out by Maxam-Gilbert methodology on fragments end-labelled at the EcoRI(-357) HincII (-6), or AvaII (-115) sites. Sequence numbering is with respect to the first base of the start codon for Pre-IFN. Direct repeats are indicated above the sequence and inverted repeats below the sequence by arrows.

presence of a large hyphenated palindrome with a 75 %-AT content is present at -239 to -277. The sequence TTTGCTTT occurs three times between -297 and -343, and subsets occur seven times in the sequence shown in Figure 5. These sequences resemble "capping" sequences [19], and in fact the 5'terminus of the capped mouse and human ß-globin gene have the sequence ACATTTGCTT which differs by only a single base insert from the sequence found at -298 to -311 (ACAATTTGCTT).

Identity of the cloned region with the human genome

In order to determine whether or not the cloned region really represents a contiguous stretch of the human genome. Southern blots of total genome were hybridised with labelled EcoRI fragments of pCosIFNß (Figure 6) or with pJC80-BglII sub-

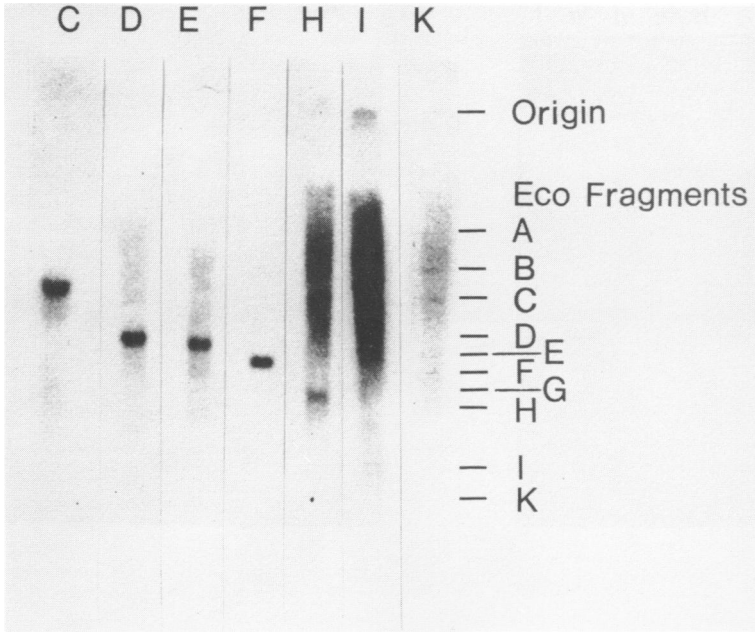


Figure 6. Southern blots of human fibroblast DNA, cut with EcoRI were hybridised with purified ^{32}P-labelled EcoRI-fragments of pCosIFNß. The lettering of the fragments is as shown in Table 1 and Figure 2, according to size.

clone DNA (Figure 7). Hybridisation with fragments EcoRI-A and -G are not shown since they contained so much homology to middle- or highly-repetitive DNA that the whole blot was radioactive and no bands were to be seen. EcoRI-C, -D, -E, -F, -H, -I and -K could all be clearly identified at their expected molecular weight. Bands I and K are not easily discernable in this figure (all 2 days exposure) but are clearly visible after a weeks exposure, at the indicated positions. The background varies markedly; the F fragment, which carries the IFNß gene, gives a very clean pattern with a single band; H, I and K give extensive heterologous hybridisation.

Southern hybridisation with subcloned BglII fragments gave only one clean result, namely with the BglII-B2 clone (Figure 7B). For comparison a Southern blot of pCosIFNß is shown to the same scale (Figure 7A). Apart from the extra band derived from

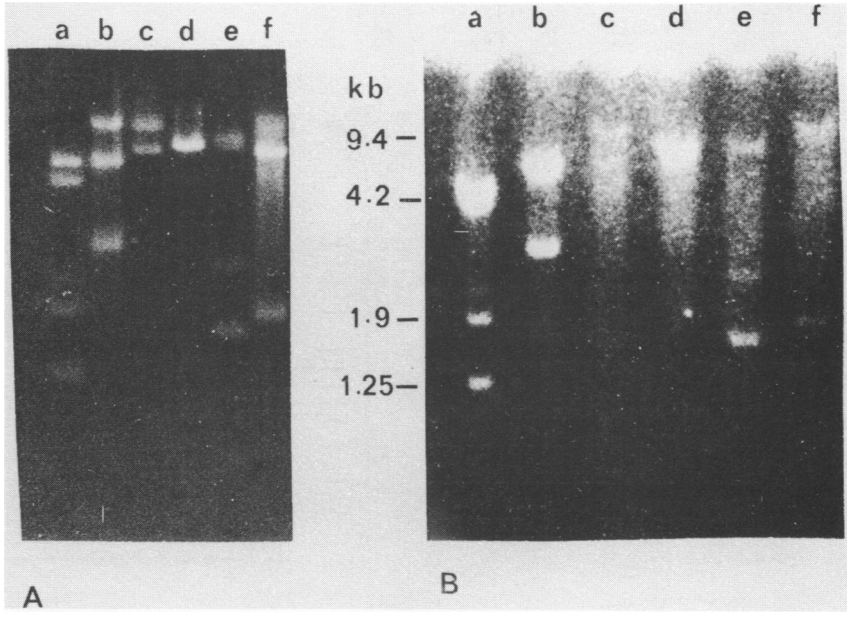


Figure 7. Southern blot hybridisation against pJC8O-BglII-B2 DNA., labelled with ^{32}P. A) Blots of pCosIFNß, B) Blots of human chromosomal DNA (fibroblast); cut with a) EcoRI, b) SStI, c) HindIII, d) BglII, e) PstI and f) PvuII. The size of fragments containing the vector pJB8, which hybridise in A) but not in B) are in lane a) 5.8, b) 11.0, c) 6.6, d) 7.2, e) 6.8 + 0,85 and f) 6.6 kb.

the vector, which only appear in Figure 7A, the patterns are identical. This blot confirms the identity of the chromosome and the cloned region between 5.8 and 23.3 kb on the map in Figure 2.

DISCUSSION

We have established that the cosmid pCosIFNß contains 36 kb of DNA from the human genome. Due to the presence of highly- or middle-repetitive DNA the possibility to confirm the identity and colinearity of cloned fragments with genomic DNA was limited to 16.5 kb. Since the gene bank was constructed using size-fractionated DNA of 35 to 50 kb it is most likely that the clone represents a continuous stretch of the human genome. A comparison of the BglII, EcoRI and HindIII pattern of

this region shows no homology with any of the eight IFN-genes so far reported [13], although the homology in sequence between the IFN-α and -ß genes (50 % for the region shown in Figure 5; 60 % for the structural coding regions [23]) and their common location on chromosome 9 (Owerbach, D. et al., Proc.Nat.Acad.Sci.U.S.A., in the press, referred to in reference 23) would lead to the prediction that close linkage is to be expected, as for example in the relatedgene clusters for the globins and immunoglobulins.

The sequence analysis to the 5'end of the IFNß gene shows a Hogness box 24 bp from the mRNA start [26,27]. Houghton et al.[21] and Lawn et al. [23] speculated that a Hogness box might exist at this position, on the basis of a comparison with the somewhat homologous IFN- 1 gene, and on the model that splicing of the 5' end of the mRNA does not take place. Our observation of the exact location of the Hogness box allows a stronger statement about the absence splicing in this system, at least when the transcription is induced in fibroblasts. On the other hand the finding that many repeated sequences based on TTTGCTTT, which are often present in the 5'leader sequences of eukaryotic mRNA is a curious phenomenon which might imply other possible transcription starts. This possibility has been raised before by the observation that Sendai virus-induced leukocytes contain mRNA species for ß Interferon which are larger than those induced in fibroblasts by polyrI:poly rC induction[29]. In our sequence an appropriately positioned Hogness box, which might help define a second transcription start point, is not observed. This site may lie just outside the sequenced region or one might invoke the existence of splicing in this region. Comparing our sequence with eight (IFN(Le)-α cDNA sequences [28], the sequence CCTTTGCTTT (-325 to -325) occurs starting 8bp after the ATG codon within the coding sequence for IFN -A, -D and -H. The significance of this, the hyphenated palindrome and the poly-adenylation recognition site[19,28]AATAAA (-201 to -208) are unclear, but resemble a short gene for an RNA species with considerable possibilities for forming secondary structures. We have not observed such an RNA species in induced fibroblasts (unpublished).

The uniqueness of the EcoRI-F fragment, containing the IFNß gene is striking in Southern blot experiments with total genomic DNA. No evidence was found for other DNA homologous to the IFNß gene with the cloned region.

One can speculate about the direction in which the IFN genes may have evolved from one another. If the gaps introduced into the IFN senquence to improve the alignment (Figure 5) can be interpreted as real deletions, then it is most likely that has evolved from ß, since deletions are more common than insertion in the evolution of eukaryotic genes[25].

The transcription patterns within the cloned region are currently under investigation, as a function of the state of differentiation of the cell and the method of induction. Preliminary observations show that Mouse LTk$^-$Aprt$^-$cells transformed with Herpes TK gene and pCosIFNß produce human Interferon-ß constitutively (H.J. Hauser et al., unpiblished). We take this as an indication that the cloned region is intact with respect to the IFN-ß gene.

ACKNOWLEDGEMENTS

We gratefully acknowledge the help and hospitality of R.A. Flavell during the screening of the cosmid-human gene bank, all members of the Genetics dept. of the G.B.F. for their support and enthusiasm and particularly for technical assistance from H. Stephan, W. Westphal and K. Littmann. We thank Guido Volckaert for introducing us to the joys of DNA sequencing. F. Grosveld and H.M. Dahl gratefully acknowledge a short term EMBO Fellowship during their stay in Braunschweig.

*Please address correspondence to J.Collins

REFERENCES

1. Stewart II,W.E.(1979)The Interferon System, Springer, New York.
2. Mayr,U.(1980)Forum Mikrobiol. 5, 269-279
3. Bloom,B.R.(1980)Nature 284, 593-595
4. Raj,N.B. and Pitha,P.M.(1977)Proc.Nat.Acad.Sci.USA, 74, 1483-1487
5. Sehgal,P.B., Doberstein,B. and Tamm,I.(1977)Proc.Natn.Acad. Sci.USA 74, 3409-3413
6. Cavalieri,R.L., Havell,E.A., Vilcek,J. and Pestka,S.(1977) Proc. Natn.Acad.Sci.USA 74, 4415-4419
7. Sehgal,P.B. and Sagar,A.D.(1980)Nature 288, 95-97

8. Weissenbach,J., Chernajovsky,Y., Zeevi,M., Shulman,L., Sereq,H., Nir,U., Wallach,D., Perricaudet,M., Tiollais,P. and Revel,M.(1980)Proc.Natn.Acad.Sci.USA 77, 7152-7156
9. Taniguchi,T., Guarente,L., Roberts,T.M., Kimelman,D., Douhan,III,J. and Ptashne,M.(1980)Proc.Natn.Acad.Sci.USA 77, 5230-5233
10. Derynck,R., Content,J., Declercq,E., Volckaert,G., Tavernier, J., Devos,R. and Fiers,W.(1980)Nature 285, 542-547
11. Gross,G., Mayr.U., Grosveld,F., Dahl,H.M., Flavell,R.A. and Collins,J.(1981)In: Molecular Biology, Pathogenicity and Ecology of Bacterial Plasmids, Clowes,R.C., Koenig,E. and Levy S. Eds., Plenum, New York. In press.
12. Goeddel,D.V., Shepard,H.M., Yelverton,E., Leung,D., Crea,R., Sloma,A. and Pestka,S.(1980)Nucleic Acids Res. 8, 4057-4074
13. Nagata,S., Mantei,N. and Weissmann,C.(1980)Nature 287, 401-408
14. Hanahan,D. and Meselson,M.(1980)Gene 10, 63-67
15. Southern,E.(1980)Meth.Enz: 68, 152-176
16. Birnboim,H.C. and Doly,J.(1979)Nucleic Acids Res. 7, 1513-1523
17. Maniatis,T., Jeffrey,A. and Kleid,D.G.(1975)Proc.Natn.Acad. Sci.USA 72, 1184-1188
18. Benoist,C., O'Hare,K., Breathnach,R. and Chambon,P.(1980) Nucleic Acids Res. 8, 127-142
19. Efstratiadis,A., Posakony,J.W., Maniatis,T., Lawn,R.M., O'Connel,C., Spritz,R.A., Deriel,J.K., Forget,B.G.,Weissman, S.M., Slighton,J.L., Blechl,A.E., Smithies,O., Baralle,F.E., Shoulders,C.C., Proudfoot,N.J.(1980)Cell 21, 653-668
20. Hohn,B. and Collins,J.(1980)Gene 11, 291-298
21. Houghton,M., Jackson,I.J., Porter,A.G., Doel,S.M., Catlin, G.H.,Barber,C. and Carey,N.H.(1981)Nucl.Acids Res.9, 247-266
22. Tavernier,J., Derynck,R. and Fiers,W.(1981)Nucl.Acids Res. 9, 461-471
23. Lawn,R.M., Adelman,J., Franke,A.E., Houck,C.M., Gross,M., Najarian,R. and Goeddel,D.W.(1981)Nucleic Acids Res. 9, 1045-1052
24. Maxam,A.M. and Gilbert,W.(1980)Meth.Enzymol. 65, 499-560
25. DeJong,W.W. and Ryden,L.(1981)Nature 290,157-159
26. Houghton,M., Stewart,A.G., Doel,S.M., Emtage,J.S.,Eaton,M. A.W., Smith,J.C., Patel,T.P., Lewis, H.M., Porter,A.G., Birch,J.R., Cartwright,T. and Carey,N.H.(1980)Nucleic Acids Res.8, 1913-1931
27. Houghton,M., Eaton,M.A.W., Stewart,A.G., Smith,J.C., Catlin, G.,Lewis,H.M., Patel,T.P., Emtage,J.S., Carey,N.H. and Porter,A.G.(1980)Nucleic Acids Res. 8, 2885-2895
28. Goeddel,D.V., Leung,D.W., Dull,T.J., Gross,M., Lawn,R.M., McCandliss,R., Seeburg,P.H., Ullrich,A., Yelverton,E. and Gray,P.W.(1981)Nature 290,20-26