At Least Three Human Type α Interferons: Structure of $\alpha 2$

Michel Streuli, Shigekazu Nagata, Charles Weissmann

Summary. The sequence of a human leukocyte-derived complementary DNA (cDNA), Hif-2h, which directs the formation in *Escherichia coli* of a polypeptide, IFN-α1, with interferon (IFN) activity has been described. A second IFN cDNA, Hif-SN206, which also elicits synthesis of a biologically active IFN, IFN-α2, is described in this article. Whereas IFN-α2 is twice as active on human as on bovine cells, IFN-α1 is 10 to 20 times more active on bovine than on human cells. As deduced from the cDNA's, the messenger RNA's for the two IFN's differ in length and in 20 percent of the nucleotides; the mature IFN polypeptides differ in 17 percent of the amino acids. Both IFN-α1 and IFN-α2 differ from the lymphoblastoid IFN described by others. Therefore, at least three different IFN-α genes are expressed in man; studies on genomic DNA reveal the presence of at least eight IFN-related genes.

Human interferons (IFN's) have been classified into three groups, α, β, and γ (*1*). Both α- and β-IFN's, previously designated type I, are acid-stable, but they differ serologically and in regard to some biological and physicochemical properties. The IFN's produced by virus-stimulated leukocytes (leukocyte IFN's) appear to be predominantly of the α-type, while those produced by lymphoblastoid cells consist of about 90 percent α- and 10 percent β-type IFN's (*2*). Induced fibroblasts produce mainly or exclusively IFN of the β-type (fibroblast IFN). Hood and his colleagues determined the NH_2-terminal sequences of fibroblast IFN (*3*) and of an IFN purified from induced lymphoblastoid cells (*4*), and they showed that α- and β-IFN's differ profoundly in the primary structure of the polypeptide chain. More recently, a comparison (*5*) between the nucleotide sequences of cloned complementary DNA's (cDNA's) coding for a leukocyte (*6, 7*) and a fibroblast IFN (*8, 9*) has confirmed the structural differences between α- and β-IFN's but also revealed distinct homologies. The γ or immune IFN or IFN's, which are produced by T lymphocytes in response to mitogens or to antigens to which they are sensitized (*10*), are acid-labile and serologically distinct from α- and β-IFN's (*11*).

The IFN's from leukocytes and lymphoblastoid cells are heterogeneous in regard to their mobilities in sodium dodecyl sulfate (SDS) polyacrylamide gels (*12*), their isoelectric points (*13–16*), their behavior on affinity chromatography (*17–19*) and high-performance liquid chromatography (*20*), and their target cell specificities (*17, 21*). Both IFN-α and IFN-β are thought to be glycoproteins (*13, 15, 18, 22, 23*); it has been suggested that the physical heterogeneity could be due, at least in part, to varying degrees of glycosylation. This variation could arise during biosynthesis or at a later stage, as a consequence of degradative processes. It is possible that one and the same type of polypeptide chain is variably glycosylated, or that there are various species of IFN polypeptides, each of which is glycosylated differently (or not at all), or that a combination of both applies.

A comparison of the amino acid sequence deduced from a leukocyte IFN cDNA (Hif-2h), which was cloned (*6*) and sequenced in our laboratory (*7*) with 34 residues of the NH_2-terminal segment of an IFN produced by lymphoblastoid cells (*4*), revealed nine differences. This discrepancy suggested that there are at least two genes for type α-IFN. Examination of our collection of IFN-α producing strains of *Escherichia coli* resulted in the identification of an IFN cDNA (Hif-SN206) with a restriction pattern different from that of Hif-2h. In this article we show that the nucleotide sequence of this cDNA differs from that of Hif-2h with respect to 20 percent of its residues. It codes for a polypeptide of 165 rather than 166 residues because it lacks the triplet coding for the Asp residue in position 44 of the Hif-2h polypeptide sequence. The amino acid sequences deduced for Hif-2h and Hif-SN206 differ in 17 percent of their positions. Since the amino acid sequence of Hif-SN206 also differs from that of the lymphoblastoid IFN in 5 out of 33 NH_2-terminal residues, we conclude that there must exist at least three genes for human α-type IFN's. We propose the designation Hu IFN-α1, Hu IFN-α2, and Hu IFN-α3 for the genes represented by cDNA Hif-2h, cDNA Hif-SN206, and the lymphoblastoid IFN partially sequenced by Zoon *et al.* (*4*), respectively.

We reported earlier on the preparation of a collection of cloned cDNA's derived from 12*S* RNA of virus-induced human leukocytes and described one clone, Hif-4c [Z-pBR322(Pst)/HcIF-4c], which gave a positive response in the hybridization-translation assay (*6*). The excised cDNA insert of this hybrid DNA, which was only about 320 base pairs (bp) long, was used as a probe to identify related hybrid DNA's in the clone collection; among 5000 clones, screened by the Grunstein-Hogness assay, 95 gave a strong hybridization response and 90 gave a weak response (*6*). A number of bacterial strains producing polypeptides with IFN activity, for example, Hif-2h, Hif-SN35, and Hif-SN42, were isolated from strongly hybridizing colonies from the set of 5000 clones as well as from other sets (*6*). We now screened the weakly hybridizing colonies and identified several strains producing IFN activity. The *E. coli* containing Hif-SN206 produced around 5000 international units of IFN activity per gram of cells, while strains containing Hif-SN208, -SN243, -SN244, -SN248, and -SN249 gave a lower yield of activity (Table 1), as determined by the cytopathic effect reduction assay (*6*) on human cells.

A comparison of the IFN activity of the weakly hybridizing strain *E. coli* (Hif-SN206) with that of two strains of *E. coli* containing hybrid cDNA plasmids of the α1-type, both on human and on bovine cells, is also shown in Table 1. The Hif-SN35 strain is an original isolate (*6*), which seems to have the same sequence as Hif-2h except for one base change (*24*); Hif-280AH3 is a derivative of Hif-2h, described in Nagata *et al.* (*6*). It is remarkable that the IFN-α1-type hybrids stimulate the production of a polypeptide with 10 to 20 times higher activity on bo-

Michel Streuli and Shigekazu Nagata are members of and Charles Weissmann is professor of molecular biology at the Institut für Molekularbiologie I, Universität Zürich, 8093 Zürich, Switzerland.

0036-8075/80/0919-1343$01.25/0

vine than on human cells, while the Hif-SN206-directed product is twice as active on human as on bovine cells. These values are measured relative to authentic, partially purified human leukocyte IFN, which is about three times more active on human than on bovine cells. The most likely explanation is that this difference in target-cell specificity of Hif-SN206 IFN and Hif-2h-type IFN is an inherent property of the polypeptides encoded by the two genes.

The Pst I digestion of Hif-SN206 released a 790-bp insert that, in comparison to Hif-2h, had two Bgl II sites rather than one and no Eco RI site rather than one. Its orientation in pBR322 is such that transcription of the IFN coding sequence is in the same direction as that of the β-lactamase gene. The restriction map of the Hif-SN206 insert is shown in Fig. 1. The restriction maps of Hif-SN208 and Hif-SN201 are similar; however, both cDNA's are oriented in the opposite sense as in Hif-SN206, and they are shorter at the 5′ end and substantially longer at the 3′ end. Nucleotide sequence analysis of Hif-SN206 was carried out by the Maxam-Gilbert procedure (*25*) from the labeled 5′ termini shown in Fig. 1. All segments were sequenced on both strands and across the restriction sites that served as origins for sequencing, except for the Bgl II site at position 185.

The sequence of the insert is shown in Fig. 2. There is an open reading frame extending from the 5′ nucleotide adjoining the deoxyguanylate tail to a termination triplet TGA (T, thymine; G, guanine; A, adenine) starting at nucleotide 547. From a comparison with the published NH_2-terminal sequence of lymphoblastoid IFN (*4*), the codon corresponding to the first amino acid of the mature IFN protein, cysteine, was located at nucleotide position 52. Thus, Hif-SN206 cDNA codes for a protein of 165 amino acids, one less than Hif-2h. A comparison of the two sequences coding for mature IFN shows that 10 percent of the nucleotide positions and 17 percent of the derived amino acid residues differ (Table 2). As can be seen in Fig. 3, the NH_2-terminal part of the SN206 protein differs in 5 of 33 residues from the lymphoblastoid IFN described by Zoon *et al.* (*4*). We conclude that at least three different IFN genes of the α-type must exist.

The IFN coding sequence of Hif-SN206 is preceded by 51 nucleotides which, by comparison to Hif-2h, most likely correspond to part of a signal sequence. The sequence does not extend far enough in the 5′ direction to reveal whether a potential initiator ATG triplet is present at position −69, as in the case of Hif-2h. The coding sequence of Hif-SN206 is followed by 196 nucleotides, which constitute part of the 3′ noncoding sequence, but most likely not all of it, since neither the AATAAA sequence, which almost always precedes the poly(A) (polyadenylate) tail by about 20 residues (*26*), nor a remnant of the poly(A) tail is represented in the insert, as is the case in Hif-2h (*7*). A comparison of the 3′ noncoding regions of Hif-2h and Hif-SN206 shows that the first segment of 90 residues is distinctly homologous (with mismatches in about 29 percent of the positions), but the remaining segments, as far as they can be compared, show very little homology (63 percent mismatches; 75 percent would be random). Both Hif-SN201 and Hif-SN208, which have restriction patterns similar to Hif-SN206, contain about 260 more base pairs at the 3′ end of the cDNA; thus, the length of the 3′ noncoding region of IFN-α2 messenger RNA (mRNA) may be around 450 nucleotides, and the total

Table 1. Interferon activity on human CCL23 and bovine embryonic kidney (BEK) cells.

Hybrid plasmid	IFN activity of *E. coli* extracts (I.U./g, wet cells)*		
	CCL23	BEK	Ratio
"Strongly hybridizing" clones			
Z-pBR322(Pst)/HcIF-SN280-AH3	15,000	150,000	1:10
Z-pBR322(Pst)/HcIF-SN35	5,000	100,000	1:20
"Weakly hybridizing" clones			
Z-pBR322(Pst)/HcIF-SN201	< 1		
Z-pBR322(Pst)/HcIF-SN206	5,000	3,000	1.7:1
Z-pBR322(Pst)/HcIF-SN208	30		
Z-pBR322(Pst)/HcIF-SN243	100		
Z-pBR322(Pst)/HcIF-SN244	10		
Z-pBR322(Pst)/HcIF-SN248	3		
Z-pBR322(Pst)/HcIF-SN249	10		

*"Weakly" and "strongly" hybridizing clones were identified with the use of the cDNA insert of Hif-4c as a probe (*6*). The *E. coli* HB101 containing the hybrid plasmid indicated were grown in tryptone medium with shaking, to an A_{650} of about 1 to 2. The cells were harvested, weighed, and resuspended in 1/100th or 1/20th of the original culture volume, and lysed by the lysozyme freeze-thaw method described in (*6*). The S30 supernatants were tested by the cytopathic effect reduction assay in microtiter plates. Extracts from control cells were negative (< 1 I.U./ml). Human CCL23 cells and BEK cells (Flow) were grown in Eagles minimum essential medium and 10 percent fetal calf serum. Exposure to IFN-containing extracts was for 24 hours; the cells were challenged with an appropriate dilution of Mengo virus and stained 24 hours later. Values were estimated visually relative to partially purified leukocyte IFN (preparation PIF) (*39*) of known titer. This preparation was about three times more active on human than on bovine cells.

Table 2. Comparison of IFN cDNA sequences. Abbreviations: S, silent sites; R, replacement sites.

	IFN-α1/IFN-α2			IFN-α1/IFN-β	IFN-α2/IFN-β
	Nucleotide differences*	Corrected divergence (%)†		Nucleotide differences*	Nucleotide differences*
		S	R		
5′ Noncoding				5/6 (83)	
Signal	5/51 (10)	25	11	38/63 (60)	23/45 (51)
Coding	52/498 (10)	32	9	269/492 (55)	259/492 (53)
3′ Noncoding	93/196 (47)	74		162/243 (67)	141/196 (72)

*Ratio of the number of nucleotide differences to the number compared; numbers in parentheses are percents. The nucleotide sequence of IFN-α1 is from Mantei *et al.* (*7*), and that of IFN-β from Taniguchi *et al.* (*8*). †Corrected for multiple events according to Perler *et al.* (*31*).

Table 3. Divergence of interferon polypeptides

IFN*	1	2	3	4	5	6
1 (Mouse C IFN)		(80)	(67)	(47)	(195)	(308)
2 (Human Le IFN-α1)	(50)		(17)19	(30)	(203)148	(195)
3 (Human Le IFN-α2)	(45)	(15)17		(23)	(259)138	(195)
4 (Human Ly IFN-α3)	(35)	(25)	(20)		(203)	(195)
5 (Human F IFN-β)	(75)	(76)68	(81)66	(76)		(133)
6 (Mouse A IFN)	(84)	(75)	(75)	(75)	(65)	

*The values in the lower left half of the matrix are the percent difference of the amino acid sequences compared; values in parentheses are comparisons of the 20 to 23 NH_2-terminal residues; those without parentheses are comparisons of the entire sequence of the mature polypeptide. The values in the top right half of the matrix were corrected for multiple mutations using table 36 in Dayhoff (*28*). The structures of 1, 4, and 6 were determined by protein sequencing (*4*, *27*); those of 2, 3, and 5 were deduced from the nucleotide sequences of the cDNA's in this article and in (*7*, *8*); Le, leukocyte; Ly, lymphoblastoid.

length of the mRNA may exceed 1000 nucleotides [exclusive of the poly(A) tail]. The corresponding values for IFN-α1 mRNA are 242 and about 900 nucleotides. The very substantial differences between the IFN-α1 and IFN-α2 mRNA's suggest that the corresponding genes are nonallelic.

Figure 3 shows an alignment of all or part of the amino acid sequences deduced or determined for human fibroblast (IFN-β), two human leukocytes (IFN-α1 and IFN-α2), a human lymphoblastoid (IFN-α3), and mouse IFN's A and C. The amino acids common to all sequences are noted in the bottom line. There is little homology among the first 28 to 30 amino acids (about 10 percent); between residues 28 and 80 (IFN-α numbering), there is almost 40 percent homology; between residues 80 and 122, there is about 15 percent homology; and in the last segment, there is 40 percent homology. Thus, there seem to be two domains that are evolutionarily conserved, as previously reported (*5*). The divergence of the different IFN's is summarized in Table 3. In the cases examined, the NH_2-terminal regions show divergence similar to that of the overall sequence.

As reported by Taira *et al.* (*27*), mouse C IFN is more similar to human type α-IFN (35 to 50 percent difference) than to mouse A IFN (84 percent difference), and mouse A IFN is more similar to human type β (fibroblast) IFN (67 percent difference) than to mouse C IFN. Therefore, mouse C IFN may be considered as being of the α-type, and mouse A IFN, of the β-type. The divergences of human α- and β-type IFN's lie between 63 and 80 percent, while the divergence of the corresponding mouse IFN's is 84 percent (comparison of 20 to 23 NH_2-terminal sequences in all cases). Therefore, the α-β divergence antedates the mouse-human divergence.

There are not sufficient comparisons available between human and mouse IFN sequences to allow us to estimate reliably the rate of fixation of mutations in the IFN proteins. If we use the values available, correct them for multiple exchanges with the use of Dayhoff's table (*28*) (IFN-α of human and mouse, average 65 percent; IFN-β, 133 percent) (Table 3), and assume that the human and mouse lines diverged 75 million years ago, we find a unit evolutionary period, that is, the time in million-year units required for the fixation of a 1 percent difference in amino acid sequence, of 1.2 for IFN-α and 0.5 for IFN-β. These values are significantly shorter than those of hemoglobins (3.3 to 3.7) and most other proteins, but they are of the same order of magnitude as those of immunoglobulins (0.7 to 1.7) or the highly mutable fibrinopeptides (1.1 to 1.7). This conclusion could be in error if the NH_2-terminal sequences of IFN's were substantially more variable than the entire polypeptide (for which there is no indication in Table 2), or if the IFN's we are comparing in human and mouse are not orthologous (*29*). For example, if a split of IFN-α into α_A and α_B had occurred long ago, and the mouse IFN-α examined was the α_A-type, whereas that of the human was the α_B-type, then the comparison would lead to an erroneously high estimate of the unit evolutionary period of IFN-α. To explain the data, this situation would also have to be true for the IFN-β comparison. On the other hand, it has been suggested that low unit evolutionary periods may occur in cases of multiple gene families or dispensable proteins (*30*), and the IFN's may qualify at least for the former category.

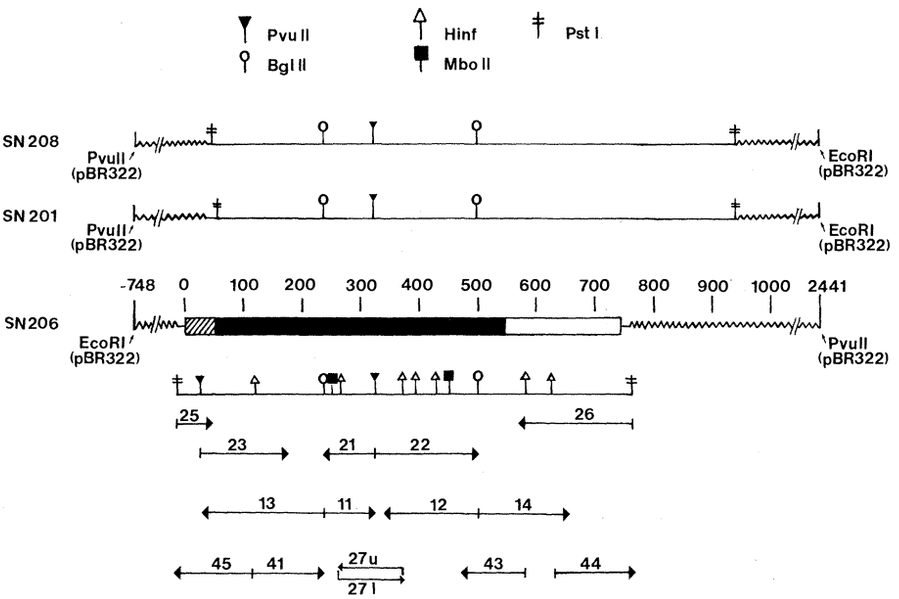

Fig. 1. Strategy for the determination of the nucleotide sequence of Hif-SN206 DNA. The restriction map was determined by double cleavage with Hinf I, Pst I, Bgl II, and by Smith-Birnstiel mapping (*34*) with fragments labeled at the two Bgl II positions. In the case of SN206, the map was subsequently refined with the use of the nucleotide sequence data shown in Fig. 2. The arrows indicate the sequences read off individual fragments; the numbers above the arrows designate the fragment (see below); and the short vertical lines represent the labeled origins. The black segment of the map indicates the coding sequence of the mature IFN; the cross-hatched segment indicates part of a putative signal sequence; straight lines next to the rectangle indicate homopolymeric deoxyguanylate-deoxycytidylate flanking regions; and wavy lines indicate pBR322. The numbers above the map indicate base pairs, measured from the 5′ terminus of the heteropolymeric part of the insert. Note that SN201 and SN208 have the orientation opposite to that of SN206 relative to pBR322. Plasmid DNA was prepared by method B described in Wilkie *et al.* (35). Restricted DNA (usually about 10 μg) was labeled at the 5′ terminus as described by Mantei *et al.* (*7*). Labeled fragments were cleaved with a second restriction enzyme; the products, separated by electrophoresis through a 5 percent polyacrylamide gel in tris-borate-EDTA buffer (*36*), were extracted from the gel and purified as described in (*37*). The various fragments for sequencing were prepared as described below. (The numbers indicate the nominal fragment chain length in base pairs; the ends are designated by the enzyme that generated them; the labeled end is indicated by an asterisk; and the boldface numbers refer to the arrows.) Fragments **25** and **26**, cleavage of Hif-SN206 with Pst I, labeling, cleavage with Bgl II, isolation of **25** Pst*-Bgl-257, and **26** Pst*-Bgl-279. Fragments **21**, **22**, and **23**, cleavage of Hif-SN206 with Pvu II, labeling, cleavage with Bgl II, isolation of **21** Pvu*-Bgl-88, **22** Pvu*-Bgl-176, and **23** Pvu*-Bgl-214. Fragments **11**, **12**, **13**, and **14**, cleavage of Hif-SN206 with Bgl II, labeling, cleavage with Pst I, isolation of **14** Bgl*-Pst-279, and a mixture of **13**, and the Bgl*-Bgl*-264 precursor to **11** and **12**. Cleavage of the comigrating fragments with Pvu II, isolation of **13** Bgl*-Pst-257, **11** Bgl*-Pvu-88, and **12** Bgl*-Pvu-176. Fragments **27**L, **27**U, **41**, **43**, **44** and **45**, cleavage with Hinf I, labeling, isolation of precursor (P) fragments **27**P Hinf*-Hinf*-113, **28**P-Hinf* Hinf*-146, **30**P Hinf*-Hinf*-159, **31**P Hinf*-Hinf*-397, **32**P Hinf*-Hinf*-1522. Fragment **41**, cleavage of **28**P with Mbo II, isolation of **41** Hinf*-Mbo-112; **43**, cleavage of **30**P with Mbo II, isolation of **43** Hinf*-Mbo-126; **44**, cleavage of **31**P with Pst I, isolation of **44** Hinf*-Pst-151; **45**, cleavage of **32**P with Pst I, isolation of **45** Hinf*-Pst-139. Fragments **27**U and **27**L were obtained by strand separation; **27**P was passed through a Chelex-100 column, adjusted to 0.3*M* NaOH, loaded on an 8 percent polyacrylamide gel (prerun for 1 hour at 150 V) in tris-borate-EDTA buffer (*36*), isolation of **27**U Hinf*-Hinf-113, and **27**L Hinf-Hinf*-113. The fragments were degraded according to Maxam and Gilbert (*25*) with the modifications described in protocols they gave us in September 1978. The products were fractionated on 0.1 by 25 by 36 cm 12 percent polyacrylamide (acrylamide-bisacrylamide in the ratio of 18 to 1) gels in 50 m*M* tris-borate, 1 m*M* EDTA (*p*H 8.3), with runs of 2, 8, 18, and 25 hours at 900 V, after 12 hours of prerun at 700 V.

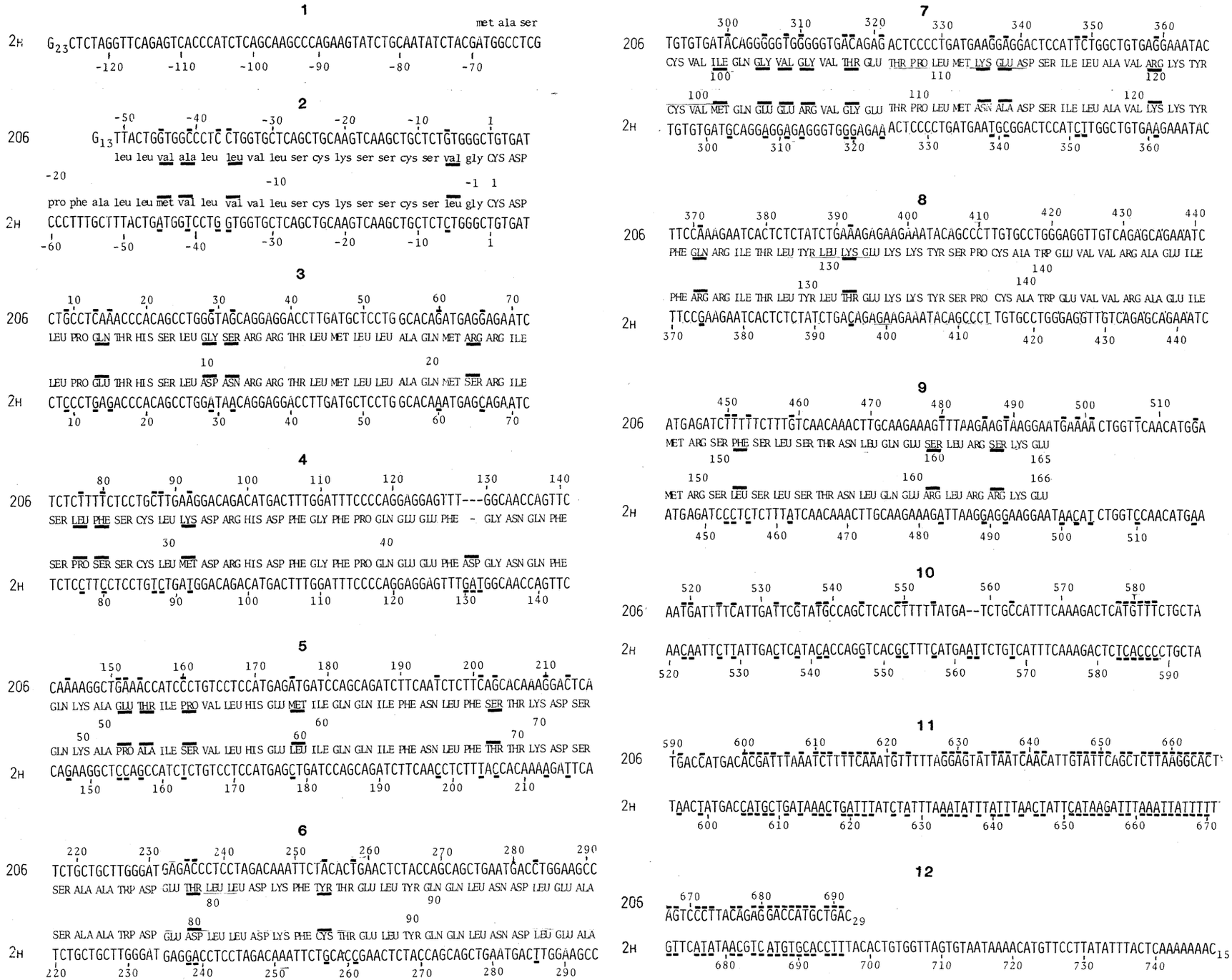

Fig. 2. The nucleotide sequence of Hif-SN206 and Hif-2h cDNA. The sequence of Hif-SN206 was determined as described in the legend to Fig. 1; the sequence of Hif-2h is taken from Mantei *et al.* (*7*). The amino acid sequence was deduced from the nucleotide sequence; lowercase letters indicate the putative signal sequence. The black horizontal bars above or below the sequences indicate amino acids and nucleotides that differ between the two sequences. The dashes within the sequences represent gaps (and not missing nucleotides) that were introduced to match the sequences optimally. Abbreviations: A, adenine; G, guanine; T, thymine; C, cytosine; Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Glu, glutamic acid; Gln, glutamine; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine; and Val, valine.

Fig. 3. Comparison of interferon polypeptide sequences. The sequences are written in the one-letter notation recommended by the IUPAC-IUB Commission on Biochemical Nomenclature (*28*): A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; and Y, tyrosine. Mouse A and C IFN's are from Taira *et al.* (*27*); lymphoblastoid IFN, from Zoon *et al.* (*4*) and Hunkapiller and Hood (*38*); Hu IFN-α1 from Mantei *et al.* (*7*); and Hu IFN-β from Taniguchi *et al.* (*8*). Dashes indicate gaps introduced to maximize the homologies; a question mark indicates an unidentified residue.

If the unit evolutionary periods we have derived for IFN's (0.5 to 1.2) are taken at face value, they suggest that the three human type α-IFN's diverged about 9 to 35 million years ago, and that human α- and β-IFN's diverged 62 to 300 million years ago. These values are lower than those obtained if the (higher) unit evolutionary period of globins is applied (*5*).

We may also attempt to estimate the time of divergence of the human α-type IFN's from the differences in the nucleotide sequence. Perler *et al.* (*31*) and Lomedico *et al.* (*32*) have calculated the unit evolutionary period for nucleotides (the time in million-year units required for the fixation of 1 percent nucleotide changes between two lines) as being 10 for replacement sites (that is, positions within codons that, on mutation, change the cognate amino acid) and 0.7 for silent sites (that is, positions that, on mutation, do not cause an amino acid change).

The values for the divergence between IFN-α1 and IFN-α2 for replacement and silent sites, calculated and corrected for multiple mutations at the same site, as described by Perler *et al.* (*31*), were 8.8 percent and 32 percent, respectively. This would mean that the two IFN's diverged between 22 and 88 million years ago. Since the divergence of the silent sites seems to provide a less reliable clock than that of the replacement sites (*31*), the lower value, 22 million years, should be the more likely one.

The finding that there are no less than three different IFN-α genes explains, at least in part, the heterogeneity of leukocyte and lymphoblastoid IFN preparations (*4, 20*); of course, different degrees of posttranslational modifications may contribute to the heterogeneity. Studies on human chromosomal DNA indicate the existence of not less than eight IFN-related sequences (*33*). It will be of interest to ascertain how many of these are expressed and to determine whether the cognate proteins have different biological functions.

References and Notes

1. W. E. Stewart, II, *et al.*, *Nature (London)* **286**, 110 (1980).
2. E. A. Havell, Y. K. Yip, T. Vilček, *J. Gen. Virol.* **38**, 51 (1977); K. Paucker, *Tex. Rep. Biol. Med.* **35**, 23 (1977).
3. E. Knight, Jr., M. W. Hunkapiller, B. D. Korant, R. W. F. Hardy, L. E. Hood, *Science* **207**, 525 (1980).
4. K. C. Zoon, M. E. Smith, P. J. Bridgen, C. B. Anfinsen, M. W. Hunkapiller, L. E. Hood, *ibid.*, p. 527.
5. T. Taniguchi, N. Mantei, M. Schwarzstein, S. Nagata, M. Muramatsu, C. Weissmann, *Nature (London)* **285**, 547 (1980).
6. S. Nagata, H. Taira, A. Hall, L. Johnsrud, M. Streuli, J. Ecsödi, W. Boll, K. Cantell, C. Weissmann, *ibid.* **284**, 316 (1980).
7. N. Mantei, M. Schwarzstein, M. Streuli, S. Panem, S. Nagata, C. Weissmann, *Gene* **10**, 1 (1980).
8. T. Taniguchi, S. Ohno, Y. Fujii-Kuriyama, M. Muramatsu, *ibid.*, p. 11.
9. R. Derynck, J. Content, E. DeClercq, G. Volckaert, J. Tavernier, R. Devos, W. Fiers, *Nature (London)* **285**, 542 (1980).
10. E. F. Wheelock, *Science* **149**, 310 (1965); J. A. Green, S. R. Cooperband, S. Kibrick, *ibid.* **164**, 1415 (1969).
11. H. M. Johnson and S. Baron, *Cell. Immunol.* **25**, 106 (1976); J. Stabo, I. Green, L. Jackson, S. Baron, *J. Immunol.* **112**, 1589 (1974); M. P. Langford, J. A. Georgiades, G. J. Stanton, F. Dianzani, H. M. Johnson, *Infect. Immun.* **26**, 36 (1979).
12. W. E. Stewart and J. Desmyter, *Virology* **67**, 68 (1975); J. Desmyter and W. E. Stewart, *ibid.* **70**, 451 (1976); J. Vilček, E. A. Havell, S. Yamazaki, *Ann. N.Y. Acad. Sci.* **284**, 703 (1977); K. Paucker, B. J. Dalton, E. T. Törmä, C. A. Ogburn, *J. Gen. Virol.* **35**, 341 (1977); K. Berg and I. Heron, *J. Immunol.* **11**, 489 (1980).
13. S. Bose, D. Gurari-Rotman, U. T. Ruegg, L. Corley, C. B. Anfinsen, *J. Biol. Chem.* **251**, 1659 (1976).
14. E. T. Törmä and K. Paucker, *ibid.*, p. 4810; K. E. Mogensen, L. Pyhälä, E. Törmä, K. Cantell, *Acta Pathol. Microbiol. Scand. Sect. B* **82**, 305 (1974).
15. W. E. Stewart, II, L. S. Lin, M. Wiranowska-Stewart, K. Cantell, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4200 (1977).
16. P. J. Bridgen, C. B. Anfinsen, L. Corley, S. Bose, K. C. Zoon, U. T. Ruegg, C. E. Buckler, *J. Biol. Chem.* **252**, 6585 (1977).
17. M. N. Thang, D. C. Thang, M. K. Chelbi-Alix, B. Robert-Galliot, M. J. Commoy-Chevalier, C. Chany, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3717 (1979).
18. W. J. Jankowski, M. W. Darey, J. A. O'Malley, E. Sulkowski, W. A. Carter, *J. Virol.* **16**, 1124 (1975).
19. J. K. Chen, W. J. Jankowski, J. A. O'Malley, E. Sulkowski, W. A. Carter, *ibid.* **19**, 425 (1976); C. B. Anfinsen, S. Bose, L. Corley, D. Gurari-Rotman, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3139 (1974).
20. M. Rubinstein *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 640 (1979).
21. S. L. Lin and W. E. Stewart, II, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **37**, 1441 (1978).
22. E. A. Havell, S. Yamazaki, J. Vilček, *J. Biol. Chem.* **252**, 4424 (1977).
23. K. C. Chadha, P. M. Grob, R. L. Hamill, E. Sulkowski, *Arch. Virol.* **64**, 109 (1980).
24. K. Henco and M. Mishina, personal communication.
25. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
26. N. J. Proudfoot and G. G. Brownlee, *Nature (London)* **263**, 211 (1976).
27. H. Taira, R. J. Broeze, B. M. Jayaram, P. Lengyel, M. W. Hunkapiller, L. E. Hood, *Science* **207**, 528 (1980).
28. M. O. Dayoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C., 1978), vol. 5, suppl. 3.
29. W. M. Fitch and E. Margoliash, *Evol. Biol.* **4**, 67 (1970).
30. A. C. Wilson, W. S. Carlson, T. J. White, *Annu. Rev. Biochem.* **46**, 573 (1977).
31. F. Perler, A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, J. Dodgson, *Cell* **20**, 555 (1980).
32. P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, T. Tizard, *ibid.* **18**, 545 (1979).
33. S. Nagata, N. Mantei, C. Weissmann, *Nature (London)*, in press.
34. H. O. Smith and M. L. Birnstiel, *Nucleic Acids Res.* **3**, 2387 (1976).
35. N. M. Wilkie, J. B. Clements, W. Boll, N. Mantei, D. Lonsdale, C. Weissmann, *ibid.* **7**, 859 (1979); Eco RI was a gift from W. Boll and Bsp I, from A. Kiss; all other enzymes were purchased from New England Biolabs.
36. A. C. Peacock and C. W. Dingman, *Biochemistry* **6**, 1818 (1967).
37. W. Müller, H. Weber, F. Meyer, C. Weissmann, *J. Mol. Biol.* **124**, 343 (1978).
38. M. W. Hunkapiller and L. E. Hood, personal communication.
39. K. Cantell and S. Hirvonen, *J. Gen. Virol.* **39**, 541 (1978). Preparation PIF was a gift from K. Cantell.

40. The authors thank P. Dierks for supplying [γ-^{32}P]ATP and F. Meyer and A. Levanon for advice concerning the sequencing. Supported by Biogen S.A. and the Kanton of Zürich.

27 June 1980