PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Structure and Expression of Human IFN-<latex>α</latex> Genes [and Discussion]

C. Weissmann, S. Nagata, W. Boll, M. Fountoulakis, A. Fujisawa, J.-I. Fujisawa, J. Haynes, K. Henco, N. Mantei, H. Ragg, C. Schein, J. Schmid, G. Shaw, M. Streuli, H. Taira, K. Todokoro and U. Weidle

Phil. Trans. R. Soc. Lond. B 1982 **299**, 7-28
doi: 10.1098/rstb.1982.0102

Phil. Trans. R. Soc. Lond. B **299**, 7–28 (1982)
Printed in Great Britain

Structure and expression of human IFN-α genes

By C. Weissmann, S. Nagata†, W. Boll, M. Fountoulakis, A. Fujisawa, J.-I. Fujisawa, J. Haynes, K. Henco‡, N. Mantei, H. Ragg, C. Schein, J. Schmid, G. Shaw, M. Streuli, H. Taira§, K. Todokoro and U. Weidle

Institut für Molekularbiologie I, Universität Zürich, Hönggerberg, 8093 Zürich, Switzerland

Copy DNA (cDNA) was prepared from induced leucocyte poly(A) RNA and cloned in *Escherichia coli*. IFN-α cDNA clones were isolated by subculture cloning with the use of a translation hybridization assay. Definitive identification of the clones was based on the production of an interferon-like protein by the transformed bacteria. Different IFN-α cDNAs, with characteristic target cell specificities, were identified. The cloned cDNAs typically encode a mature polypeptide of 166 (or, for IFN-α2, 165) amino acids and a signal sequence of 23 amino acids.

A human chromosomal library was screened with IFN cDNA and 17 distinct IFN-α-related sequences were isolated and identified, of which 7 proved to be non-allelic authentic genes and 4 pseudogenes; 6 sequences remain to be elucidated. Taking into account the work of Goeddel and his colleagues, 13 non-allelic authentic genes and 6 pseudogenes can be distinguished. In addition, 9 genes believed to be allelic to the 13 authentic genes have been sequenced. The IFN-α genes may be classified into two major subfamilies, which diverged at least 33 Ma ago, but perhaps much earlier, if sequence rectification occurred. At least one IFN-α gene appears to have resulted by a recombinational event between members of the subfamily I and II. IFN-β is distantly related to IFN-α's and may have diverged from a common ancestor at least 500 Ma ago. Both IFN-α and IFN-β genes differ from most other genes of higher organisms by being devoid of introns. The mouse was found to possess an IFN-α gene family of a size similar to that of man; the murine genes also do not have introns.

IFN-α genes devoid of their signal sequence were joined to prokaryotic promoters to produce the mature interferons in *E. coli* in high yield. IFN-α2, purified to homogeneity, has been crystallized by T. Unge and B. Strandberg (Uppsala). Hybrid genes consisting of IFN-α1 and IFN-α2 segments were constructed and expressed in *E. coli*; the target cell specificities of such hybrids were dependent on the arrangement of the segments and were different from those of either parent.

The chromosomal gene for HuIFN-α1 was introduced into mouse L cells to study the mechanism of its expression. Correct transcription was only detected after induction (with Newcastle disease virus); expression was transient, with the same kinetics as those of the endogenous mouse IFN mRNA.

Natural murine IFNs and human IFN-β and IFN-γ are glycosylated. Because *E. coli* cells transformed with the genes of eukaryotic glycoproteins are not expected to yield correctly glycosylated polypeptides, we prepared lines of hamster cells permanently transformed with hybrid plasmids, which contained an IFN gene linked to the SV40 early promoter, as well as dihydrofolate reductase as a selective marker. After intracellular amplification of the introduced genes, cell lines were obtained which constitutively produced IFN at about 40 000 units ml^{-1} and could be propagated for at least several months.

† Present address: The Institute of Medical Science, University of Tokyo, P.O. Takanawa, Tokyo, Japan.
‡ Present address: BASF, Abteilung Biotechnologie, 6700 Ludwigshafen, F.R.G.
§ Present address: Department of Pure and Applied Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153, Japan.

Cloning human IFN-α cDNA

When we undertook to clone the gene for IFN-α (or leucocyte IFN, as it was then designated) in 1977, we faced two major problems. First, no information on the structure of the protein was available that would allow physical identification or synthesis of part or all of the coding sequence, and second, the level of IFN mRNA in interferon-producing cells (which could serve as starting material for the preparation of IFN complementary DNA (cDNA) clones) was estimated to be low, around 0.1–0.01 %. It was therefore necessary to design a cloning strategy different from those employed previously for somatostatin (Itakura *et al.* 1977), growth hormone (Seeburg *et al.* 1977; Harpold *et al.* 1978), insulin (Ullrich *et al.* 1977; Crea *et al.* 1978; Chan *et al.* 1979; Villa-Komaroff *et al.* 1978) or parathyroid hormone (Kronenberg *et al.* 1979). We decided to forego attempts at extensive purification of the IFN mRNA, and invest our major effort in screening a large number of cDNA clones by an assay ultimately based on the biological detection of interferon. Poly(A) RNA from interferon-producing leucocytes was fractionated on a sucrose density gradient and interferon mRNA was measured by injecting samples of each fraction into *Xenopus* oocytes and determining the interferon activity excreted by the oocytes (Reynolds *et al.* 1975; Colman & Morser 1979) by an antiviral assay (Stewart & Sulkin 1966). The 12 *S* poly(A) RNA fraction, which showed the highest interferon mRNA activity, was used to generate double-stranded cDNA, and this was joined to the plasmid pBR322 at the *Pst*I site by established methods (Villa-Komaroff *et al.* 1978) (figure 1). As the *Pst*I site is located in the β-lactamase gene of the plasmid, insertion of an appropriate cDNA into this position could give rise to a fused protein consisting of a part of β-lactamase linked to part or all of the interferon (or preinterferon) molecule (cf. Villa-Komaroff *et al.* 1978).

Ten thousand clones of transformed bacteria were picked individually into microtitre wells. To identify the desired plasmid, we used a modification of the hybridization–translation assay (Harpold *et al.* 1978), the principle of which is outlined in figure 2. A mixture of several different hybrid plasmids, one of which may contain the IFN cDNA sequence, is denatured and bound to a solid support. Poly(A) RNA from IFN-producing cells, which contains some IFN mRNA, is hybridized to the filter-bound DNA; if the DNA mixture contains IFN cDNA, IFN mRNA will be bound to it. After the filter has been washed, the bound RNA is recovered by elution under denaturing conditions and injected into oocytes to determine the presence of IFN mRNA. This procedure allows the detection of IFN cDNA in a mixture of hybrid DNAs. The principle of subculture cloning was applied to isolate a single hybrid plasmid capable of binding IFN mRNA, as shown schematically in figure 3. Twelve pools of 512 bacterial clones each were prepared and the DNA assayed in the fashion described above. A positive pool was identified, and the 512 clones composing it were repooled in eight subgroups of 64 clones each. A positive subgroup was again determined by the hybridization–translation assay, and the cognate clones repooled in eight groups of eight. Finally, a single hybrid DNA clone was isolated that had the property of hybridizing to IFN mRNA. In reality, the assays were not always consistently positive (or negative), and at each level more than one pool was positive, indicating that the IFN mRNA was considerably more frequent than 1 in 5000, our worst-case assumption. Although a positive hybridization–translation assay constituted an important criterion for a plasmid containing IFN cDNA, the identification provided by this assay was not definitive, because any cDNA fortuitously sharing some sequences with IFN mRNA, perhaps even in a non-coding region, would give a positive response. Conclusive identification would be provided

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

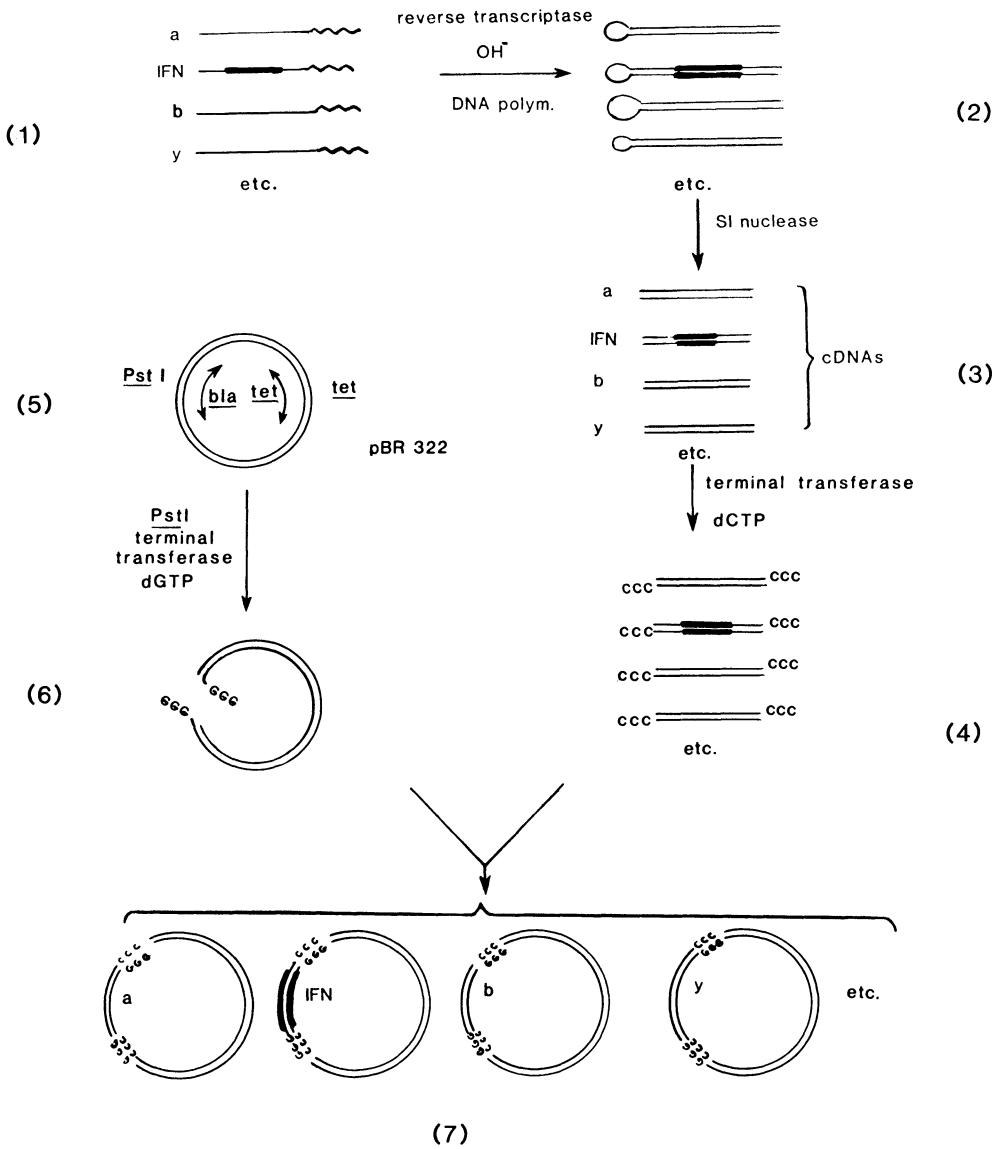


FIGURE 1. Shotgun cloning of cDNA sequences derived from leucocyte poly(A). Poly(A) RNA from Sendai virus-induced leucocytes (which contain a variety of mRNAs a, b, . . ., y, etc., and a small amount of IFN mRNA) was centrifuged through a sucrose gradient and the 12*S* fraction (1) copied into single-stranded cDNA by using reverse transcriptase. After removal of the RNA by alkali treatment, this DNA was converted to double-stranded DNA by using DNA polymerase, which attaches to the 3′ end of the template and thereby generates a DNA loop (2). After cleavage with nuclease S_1, which is specific for single-stranded DNA (3), the 3′ ends were elongated with dCMP residues, by using terminal transferase and dCTP (4). Plasmid pBR322 (*bla*: β-lactamase gene; *tet*, tetracycline resistance gene) (5) was cleaved with endonuclease *Pst*I, the 3′ ends were elongated with dGMP residues, by using terminal transferase and dGTP (6), and hybridized with the dCMP-elongated cDNA. The hybrids (7) were introduced into *E. coli*, and 10000 tetracycline resistant bacterial colonies were picked into the wells of microtitre plates.

by the demonstration that the cDNA sequence can give rise to a biologically active protein by expression in *Escherichia coli* or in a eukaryotic cell. Usually this requires linking a complete coding region to an appropriate prokaryotic promoter and translation initiation region. However, we decided to try a short cut and assay extracts of bacteria containing putative interferon cDNA plasmids without further manipulation. The first clone identified had a

cDNA insert of only 320 nucleotides, and was almost certainly too short to encode a functional interferon molecule, since the mRNA, as judged by its sedimentation behaviour, was about 1000 nucleotides in length. We therefore used that cDNA as a radioactive probe to identify some 200 additional clones by colony hybridization (Grunstein & Hogness 1975). A number of these bacterial clones, some of which contained cDNAs of the correct length, were grown to small cultures, the bacteria harvested, lysed and the extracts tested for antiviral activity. Gratifyingly, a number of samples gave positive results, and the antiviral activity was shown to be due to an interferon-like molecule by specific neutralization with anti-IFN-α antibodies and

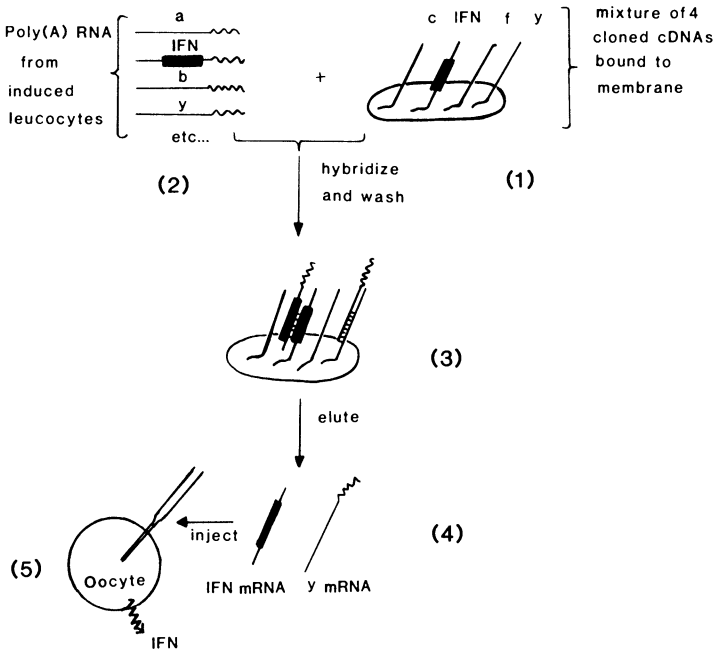


FIGURE 2. An assay for hybrid plasmids containing interferon cDNA. Hybrid plasmid DNA from a pool of clones (see legend to figure 1) is linearized, denatured and attached to a nitrocellulose or a derivatized cellulose filter (Nagata *et al.* 1980*a*). Poly(A) RNA from induced leucocytes (2) is hybridized to the filter-bound cDNA. If the mixture of cDNAs contains part or all of an interferon cDNA (or a sequence related to it), the interferon mRNA present in the poly(A) RNA will be retained on the filter (3) and can be eluted from it after the hybrids are denatured (4). Other mRNAs related to cDNA species bound to the filter will likewise be recovered, but do not disturb the assay. The poly(A) RNAs recovered from the filter are injected into *Xenopus* oocytes; if the sample contains interferon mRNA, the oocytes excrete interferon into the medium (5) (Colman & Morser 1979).

other physicochemical criteria (Nagata *et al.* 1980*a*). It would seem that biologically active molecules arose as a rare event by abnormal initiation, perhaps at the junction of the GC tails and the cDNA, as in the dihydrofolate reductase clones described by Chang *et al.* (1978).

Whatever the mechanism by which it came about, the expression of typical IFN-α antiviral activity allowed the rapid and incontrovertible identification of a functional IFN-α cDNA.

THE STRUCTURE OF IFN-α cDNA

Nucleotide sequence analysis of the cDNA of one of the IFN-producing clones, Hif-2h (figure 4), allowed the deduction of the amino acid sequence encoded within the open reading frame. It consisted of 189 amino acids; a comparison with the sequence of the first twenty

amino acids of natural IFN purified from lymphoblastoid cells (Zoon *et al.* 1980), which had just become available, showed that the first amino acid of natural IFN-α corresponded to the 24th codon of the cloned sequence, suggesting that IFN, like so many other secreted proteins, was synthesized as a pre-protein, and that an amino-terminal segment of 23 residues was cleaved off during maturation to yield a mature protein of 166 amino acids, which we called IFN-α1 (Mantei *et al.* 1980). Since the sequence determined by Zoon *et al.* (1980) and that

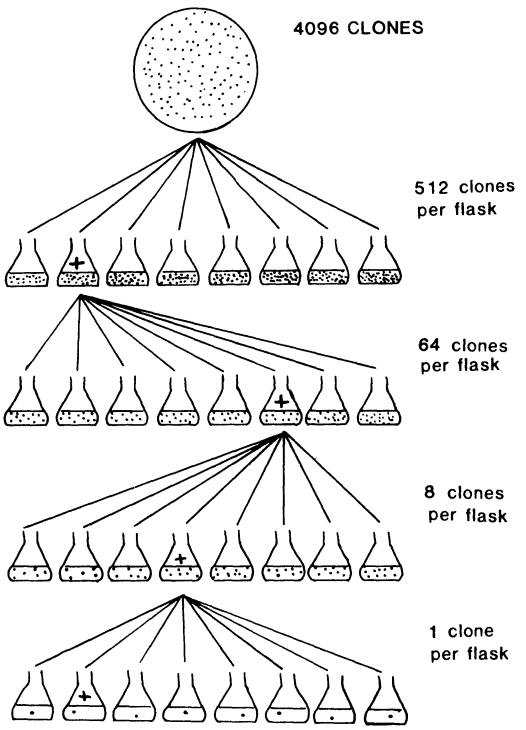


FIGURE 3. Isolation of a IFN cDNA-containing bacterial clone by subculture cloning (scheme). Bacterial clones were stamped from microtitre wells onto agar plates. In the first round of assays, 12 groups of 512 clones were used to seed 2 l cultures and plasmid DNA from the 12 groups was prepared. The DNA was assayed as described in figure 2 for the presence of sequences hybridizing to IFN-α mRNA, and a positive group was identified. The 512 clones corresponding to one positive group were assembled from the corresponding microtitre wells into 8 groups of 64, and plasmid was prepared from each group and assayed. This procedure was continued until a single positive clone was identified.

deduced by our group differed in 5 of 20 positions, we suspected there might exist more than one type of IFN-α. Indeed, a second IFN-producing clone picked from our cDNA clone bank, IFN-α2, encoded a mature IFN-α, containing 165 (rather than 166) amino acids, that proved to differ in 20% of its amino acid residues from IFN-α1 (Streuli *et al.* 1980) and differed also from the partial sequence described by Zoon *et al.* (1980). It thus became clear that there existed at least three different genes for interferons of the α type, and that the multiple peaks of IFN-α activity resolved by h.p.l.c. (Rubinstein *et al.* 1979) might correspond to products of different genes and not to multiple forms of one IFN species differing merely in length and in degree of glycosylation, as was commonly held. It later emerged that all human IFN-α species examined so far are in fact not glycosylated (Allen & Fantes 1980; M. Rubinstein, personal communication). The finding that IFN-α1 and IFN-α2 had about the same specific activity on bovine cells, whereas IFN-α1 had one-tenth of the specific activity of IFN-α2 on

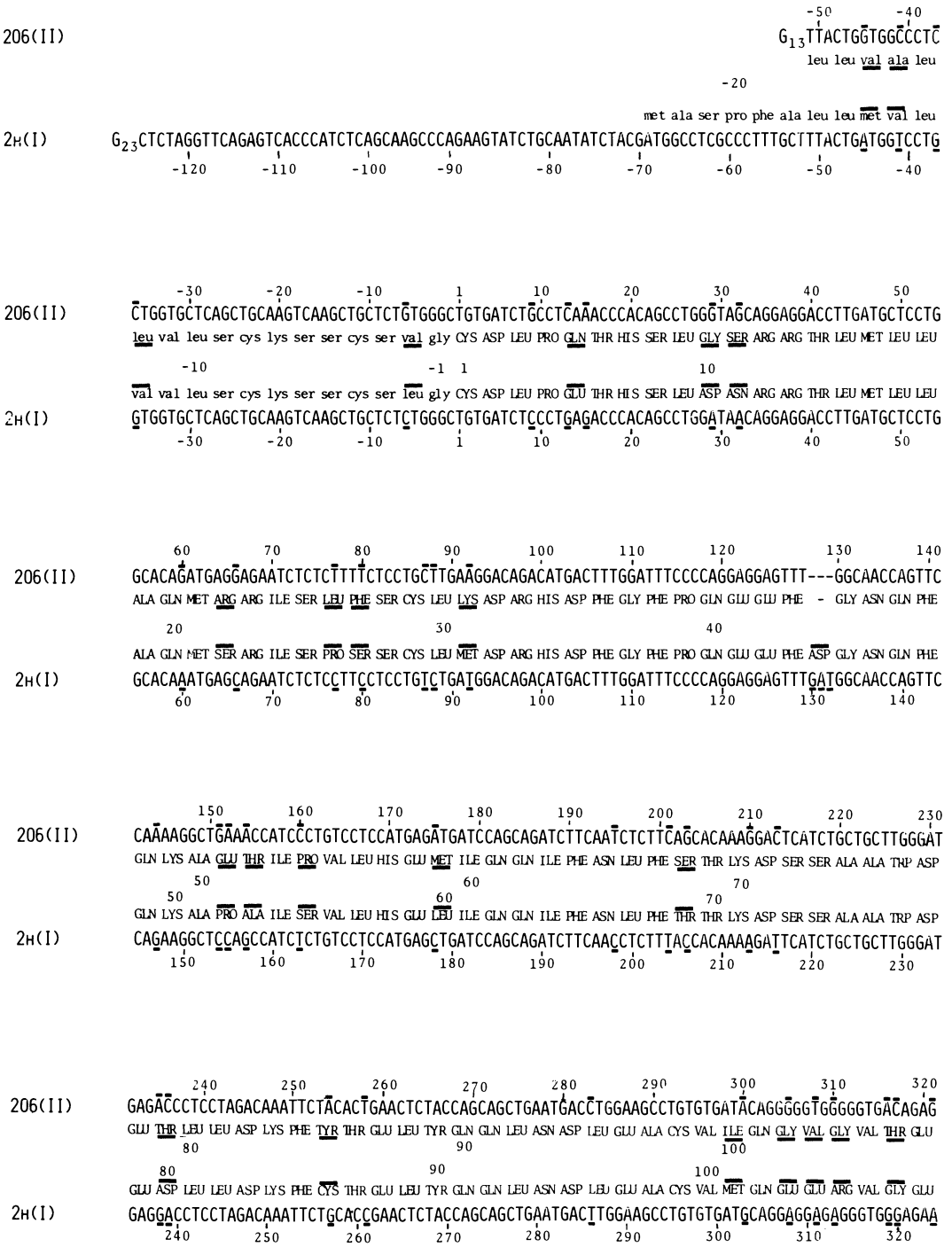


FIGURE 4. The nucleotide sequence of *Hif*-SN206 (IFN-α2) and *Hif*-2h (IFN-α1) cDNA. The sequence of *Hif*-SN206 was determined by Streuli *et al.* (1980), that of *Hif*-2h is taken from Mantei *et al.* (1980). The amino acid sequence was deduced from the nucleotide sequence; lower case letters indicate 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 that were introduced to match the sequences optimally.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

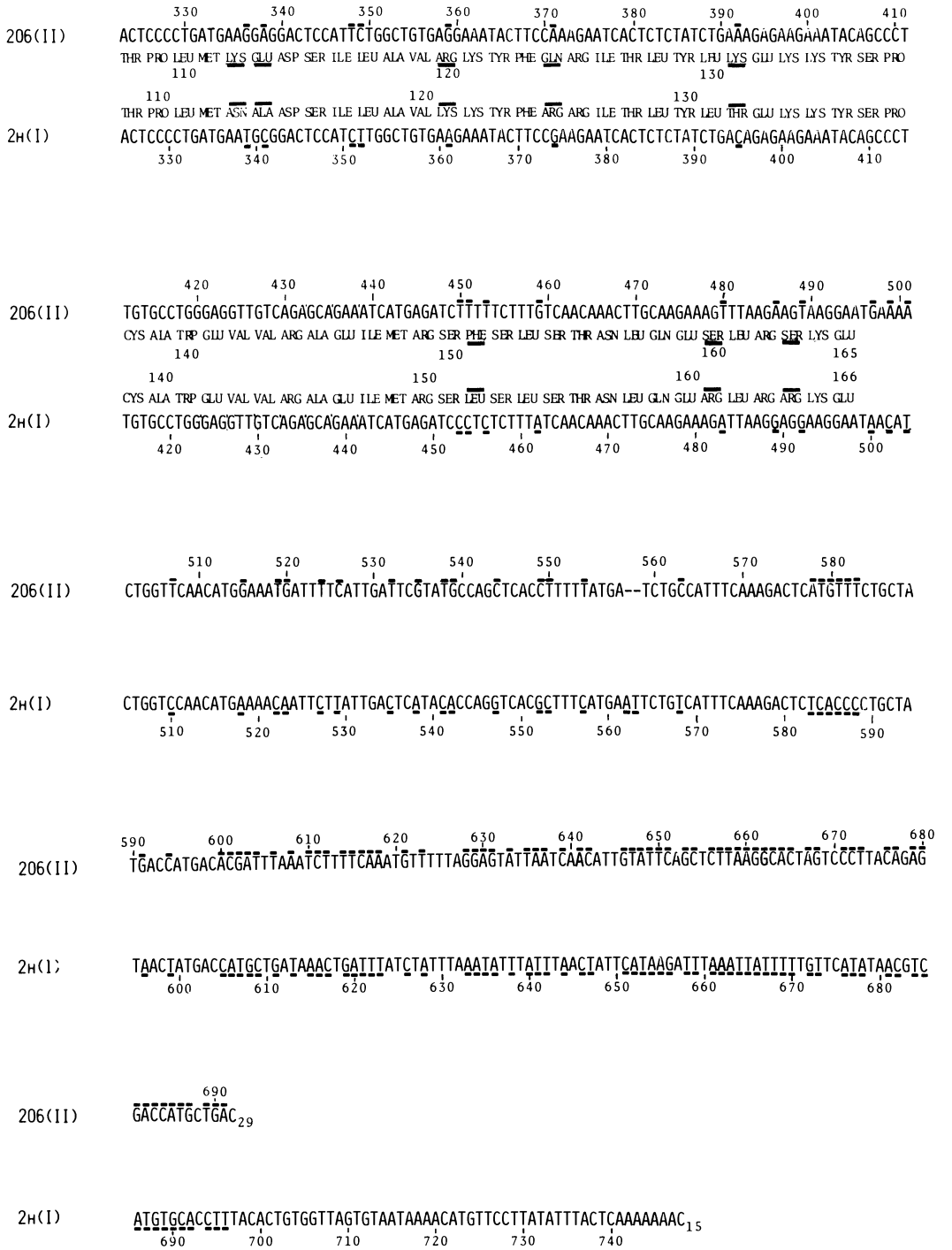


FIGURE 4. (*cont.*)

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

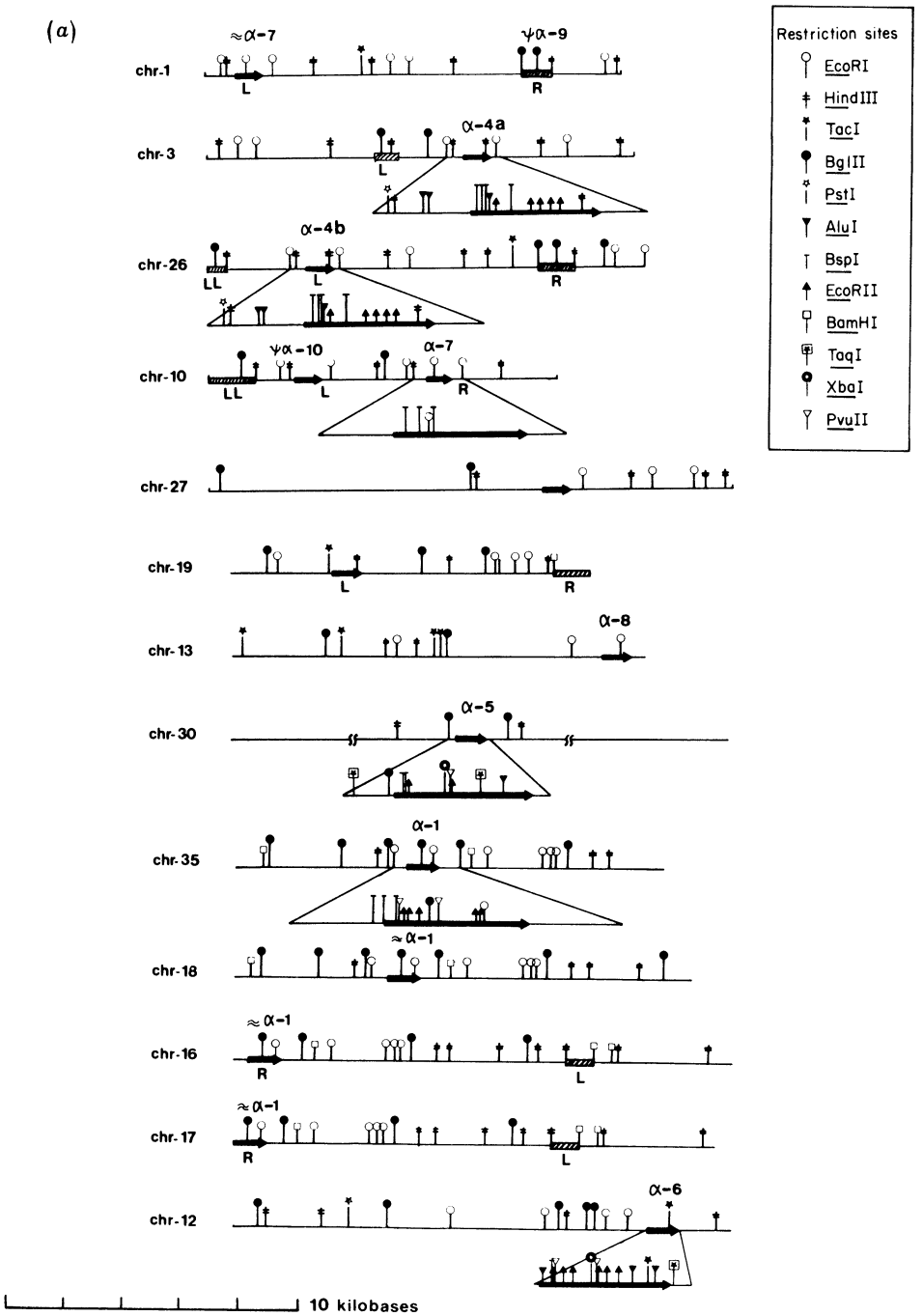


FIGURE 5*a*. For description see opposite.

human cells (Streuli *et al.* 1980; Stewart *et al.* 1980) demonstrated a distinct difference in the target cell specificity of the two IFN species, and raised the possibility – so far unsupported – that the IFN-α species might have different targets within one and the same organism. In any event it could be shown that at least one of the species, IFN-α1, had a number of the properties previously ascribed to the impure mixture of natural interferons, for example, activation of Natural Killer cells, inhibition of Daudi cell growth, enhancement of antibody-dependent cell-mediated cytotoxity, and suppression of antigen and mitogen-induced leucocyte inhibition (Masucci *et al.* 1980).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

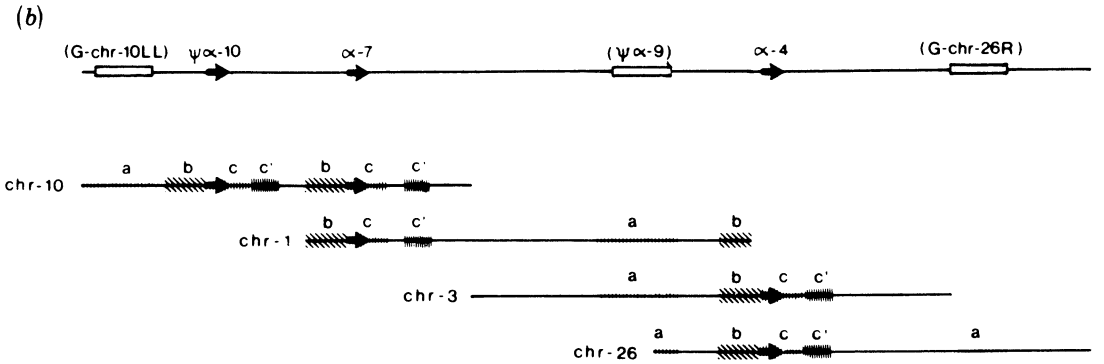


FIGURE 5. Location of IFN-α genes and pseudogenes on cloned fragments of human chromosomal DNA. (*a*) The designation at the left of each line identifies the clone. Arrows indicate the orientation of the gene or pseudogene. Genes designated with α followed by a number have been sequenced; ψ indicates a pseudogene; ≈ means that the gene is identical or very similar (allelic) to the gene with the same designation. (*b*) From the data of (*a*) and heteroduplex mapping a partial linkage map was constructed. The 36 kilobase-pair DNA segment comprises six genes or pseudogenes: α4 and α7 are sequenced genes. ψ-α10 and ψ-α9 are sequenced pseudogenes. G-chr-10LL and G-chr-26R have not yet been sequenced. (From Brack *et al.* (1981).)

THE CHROMOSOMAL INTERFERON GENES

In order to isolate the chromosomal interferon genes, a human gene library cloned in phage λ by Lawn *et al.* (1978) was screened with a ^{32}P-labelled IFN-α1 cDNA probe.

As the human genome has about 3×10^9 base pairs and the average size of the cloned DNA fragments was about 16000 base pairs, it was expected that one phage in about 200000 would contain an interferon gene, if there were one gene per haploid genome. In fact, 16 clones of 240000 screened contained DNA hybridizing to IFN-α1 cDNA, suggesting that there might be 10–15 interferon-related genes per haploid genome (Nagata *et al.* 1980*b*). Restriction and sequence analysis of the cloned DNAs, as well as electron microscopic studies (Brack *et al.* 1981) led to the maps shown in figure 5*a*, in which at least 15 *different* IFN-α genes and pseudogenes are entered. From the sequence analyses of our clones (K. Henco, J. Schmid, J.-I. Fujisawa, A. Schamböck, T. Kovacic, M. Pasek, H. Hochstedt & J. Brosius) and those of Genentech (Lawn *et al.* 1981*a*, *b*; D. Goeddel, personal communication), as well as from expression experiments, there exist at least 13 distinct functional interferon genes, which we consider to be non-allelic (table 1) because they are flanked by very dissimilar DNA sequences or are arranged in tandem, or both. In addition, several variants of these genes, believed to be alleles, were identified. A careful examination of the chromosomal IFN-α1 gene showed that it was devoid of introns (Nagata *et al.* 1980*b*); no evidence for introns was found in any of the other IFN-α genes examined, nor in the only IFN-β gene identified so far (Ohno & Taniguchi 1981; Maeda *et al.* 1981; Tavernier *et al.* 1981; Lawn *et al.* 1981*c*). Probably ten or more IFN-α genes are expressed in man, as judged by the fact that they give rise in *E. coli* to biologically active interferons (Nagata *et al.* 1980*b*; S. Nagata, M. Mishina, W. Boll, J.-I. Fujisawa & C. Weissmann unpublished results), and that cDNAs (which are of course derived from mRNAs) corresponding to many of the genes have been isolated (Goeddel *et al.* 1980). Allen & Fantes (1980) have, by amino acid sequence analysis, shown that interferon preparations from lymphoblastoid cells contain at least five different interferon species, of which one corresponds to IFN-α2, and another most probably to IFN-α1. Rubinstein *et al.* (1979) have separated about ten distinct IFN species, of which two were identified with IFN-α2 (Levy *et al.* 1982). Although the amino acid sequences assembled by Allen & Fantes (1980) from partial sequence data agreed with the

complete structures predicted from nucleotide sequence determination, the IFN-α species analysed by Levy *et al.* (1982) lacked the 10 carboxyterminal amino acid residues, most probably as a consequence of (physiological or artefactual?) post-synthetic proteolytic cleavage.

Beside the genes that have an intact coding sequence, we have identified a number of IFN-α pseudogenes, i.e. DNA sequences that are clearly related to functional genes but whose coding sequences contain mutations that preclude the expression of a full-length interferon-like protein.

TABLE 1. INTERFERON GENES AND PSEUDOGENES

(=, Identical; ≈, allelic or very similar.)

(1) *Genes closely related to IFN-α1*

(*a*) Sequenced and correlated with Genentech sequences

Zürich		Genentech Inc.		Dr Pestka
IFN-α1		≈ IFN-D		
IFN-α2	= λα2	≈ IFN-A	=	IV
IFN-α5		= IFN-G		
IFN-α6	= λ5	≈ IFN-K		
IFN-α7		= IFN-J(λ1J)	≈	I–II
IFN-α8		≈ IFN-B		

(*b*) Sequenced but not correlated

IFN-α13†
IFN-α4a ≈ IFN-α4b
IFN-F
IFN-C
λ2h ≈ IFN-H
λ2C$_1$
λ4a

(*c*) Not sequenced

G-chr-19 (≈ IFN-α6)
G-chr-27 (≈ λ2h)

Altogether at least 13 non-allelic and 7 allelic genes.

(2) *Genes distantly related to IFN-α1*

(*a*) Sequenced

ψIFN-E
ψIFN-αM
ψIFN-α10 ≈ ψλ1‡
ψIFN-α9
ψIFN-α11
ψIFN-α12

(*b*) Not sequenced

G-chr-10
G-chr-26R
G-chr-19R
G-chr-16

Altogether at least 6 distinct pseudogenes.

Symbols: =, identical; ≈ allelic or very similar; †, similar to but non-allelic with IFN-α1; ‡, similar to IFN-C.

These mutations range from a single nucleotide substitution introducing a premature termination signal in the coding region, to multiple, extensive insertions and deletions that mutilate the gene to virtual unrecognizability (S. Nagata, A. Fujisawa, J. Haynes, K. Todokoro, unpublished results). Such pseudogenes were found as weakly hybridizing regions within the chromosomal DNA clones described above. Many additional weakly hybridizing regions have been identified, but not sequenced; these may represent either additional pseudogenes, or genes distantly related to IFN-α genes.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

All IFN-α genes and the IFN-β gene are located on chromosome 9 (Owerbach *et al.* 1981; Slate *et al.* 1982); several genes and pseudogenes are located within a few kilobases of each other (figure 5*b*) (Brack *et al.* 1981), whereas others are less closely linked.

The evolution of the IFN genes

A comparison of the IFN-α1 and IFN-β (fibroblast) genes had revealed a distinct, albeit distant, nucleotide and amino acid sequence relationship (Taniguchi *et al.* 1980). Assuming a mutation fixation rate similar to that of globin, we estimated that the two genes had diverged not less than some 500–1000 Ma ago, and must thus be as old as the vertebrates.

When did the IFN-α gene family arise? The two most dissimilar members sequenced so far, IFN-α1 and IFN-α7, have a divergence of 11.7% (calculated on the basis of replacement sites, and corrected for multiple mutations at the same site (cf. Perler *et al.* 1980)). Since the mutation fixation rate at replacement sites for IFN-α (as determined in a comparison of human and mouse IFN-α's (G. Shaw, N. Mantei & C. Weissmann, in preparation) is about 0.35 Ma^{-1} (compared with the value of 0.18 determined for β-globin (Perler *et al.* 1980)), the earliest IFN-α gene duplication dates back at least 33 Ma. Earlier estimates were about 22–88 Ma (Streuli *et al.* 1981) or 26 Ma (Miyata & Hayashida 1982). However, as different members of a gene family may be rectified (i.e. rendered more similar to each other (Hood *et al.* 1975)) by unequal crossing over (Smith 1976) or gene conversion (Radding 1978; Baltimore 1981), the family may be far older. In fact, since the mouse (G. Shaw, N. Mantei, H. Taira & P. Lengyel, unpublished results) and the rat (R. Dijkema, personal communication) also have a large IFN-α gene family, it is conceivable that the IFN-α gene divergence even precedes the mammalian radiation (75 Ma ago).

Figure 6 shows the amino acid sequences of the human IFN-α's determined so far. They have been assembled into three major groups; the members of group I and group II each have distinctive features in common, while the members of the intermediate group have features of both group I and II.

Table 2 shows the distinctive amino acids at 10 positions of groups I and II. IFN-H has the pattern of group I until between positions 79 and 83, whereupon it resembles group II. IFN-α8 has the pattern of group I at the termini, but the middle segment, from between position −8 and −13 to between 79 and 83, resembles a member of group II. IFN-F seems to pertain to group II except for a region around amino acid 70 and the carboxy-terminal region. The nucleotide sequence of IFN-F and IFN-H suggests that these genes may have come about by unequal crossing over or by gene conversion; no clear statement can be made for IFN-α8, which may represent a third subfamily.

Figure 6 further shows that pairs of IFN-α sequences such as α1 and D, α2 and A, λ2h and H, α4a and α4b differ by only 1 or 2 amino acids. We believe that these pairs represent allelic variants; however, only in α4a and α4b is this assumption supported by a study of the cognate chromosomal DNA segments, which shows that about 3.4 and 5 kilobases of DNA flanking the gene on the 5′ and 3′ sides, respectively, form perfect heteroduplexes (Brack *et al.* 1981). In all other cases, the questionable argument in favour of an allelic relation rests on the low degree of divergence of nucleotide sequence within the region of the gene, in particular of the 3′ non-coding region.

Only a single IFN-β gene has been conclusively identified in man (Lawn *et al.* 1981*c*;

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

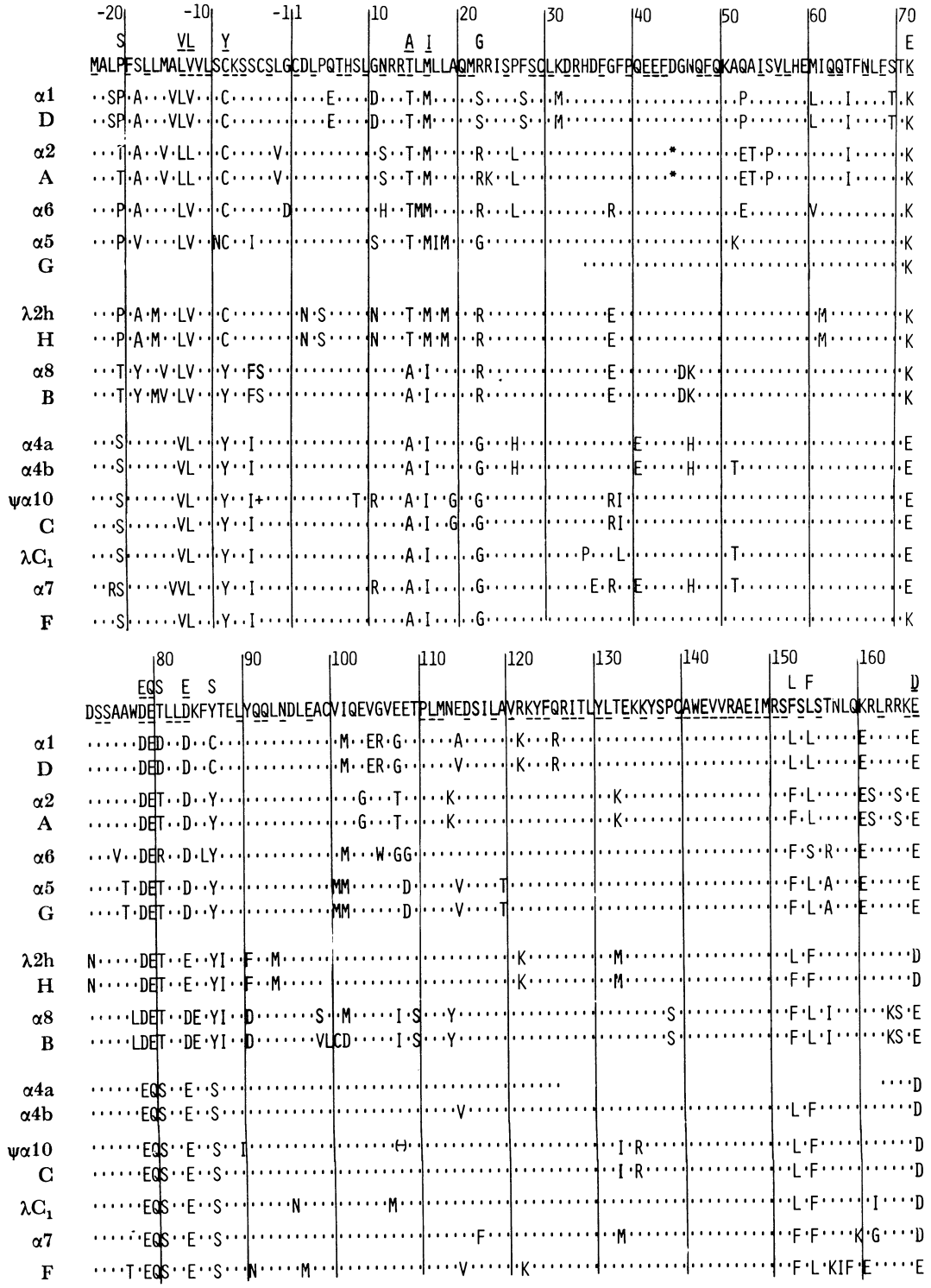


FIGURE 6. Amino acid sequences of IFN-α's. All sequences have been derived from the nucleotide sequences of cloned cDNAs or chromosomal DNAs. Sequences designated with letters are from work by the Genentech group (Goeddel *et al.* 1981; Lawn *et al.* 1981 *a*, *b*); the sequences designated with arabic numerals were determined by Mantei *et al.* (1980) (IFN-α1), Streuli *et al.* (1980) (IFN-α2), A. Schamböck (IFN-α4a), K. Henco (IFN-α4b), J. Fujisawa (IFN-α5), J. Schmid (IFN-α6), T. Kovacic (Biogen S. A., Geneva) (IFN-α7), M. Pasek (Biogen S. A., Geneva) (IFN-α8), and H. Hochstadt and J. Brosius (Harvard University Biological Laboratories) (ψIFN-α10). Top line, consensus sequence (underlined amino acids are common to all IFN-α's. *, No amino acid (deleted triplet); +, termination codon; −, frame-shift (due to point deletion) (the reading frame has been corrected to allow comparison).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Tavernier *et al.* 1981; Maeda *et al.* 1981; Ohno & Taniguchi 1981); suggestions of other IFN-β genes (Weissenbach *et al.* 1980; Sehgal & Sagar 1980) still lack substantiation. This means that either the IFN-β gene never underwent duplication, or that if it did, all but one of the IFN-β genes were eliminated or degenerated to pseudogenes. The number of IFN-β genes in other vertebrates has yet to be determined.

The relation of IFN-γ to the other two IFN types is not clear. A limited homology between IFN-α and IFN-γ (Gray *et al.* 1982) has been pointed out by Epstein (1982). The chromosomal IFN-γ gene has three introns (D. Goeddel, personal communication) in contrast to the IFN-α and IFN-β genes, stressing the dissimilarity between IFN-α and IFN-β on the one hand and IFN-γ on the other.

TABLE 2. CHARACTERISTIC AMINO ACID POSITIONS IN THE TWO MAJOR IFN-α SUBFAMILIES

subfamily		−13	−8	14	16	70	78	79	83	154	160
I		L	C	T	M	K	D	E	D	L†	E
intermediate	H	L	C	T	M	K	D	E	e	f	d
	α8	L	y	a	i	K	D	E	D	L	E
	F	v	y	a	i	K	e	q	e	L	E
II		v	y	a	i	e	e	q	e	f	d

Subfamily I comprises interferons α1(D), α2(A), α5(G) and α6. Subfamily II consists of α4a(α4b), α10(C), λC_1, α7 and F. The amino acids common to all members of a subfamily but not to all IFN-α's are given in capital letters (subfamily I) and lower case letters (subfamily II). Negative numbering, signal sequence; positive numbering, mature sequence.

† α6 has S at this position.

EXPRESSION OF HUMAN INTERFERON IN *E. COLI*

The strains of *E. coli* containing the IFN-α1 cDNA in the *Pst*I site of the plasmid PBR322 had relatively low levels of interferon activity, around 20000 units l^{-1}.‡ We believe that translation of the active polypeptide did not initiate at the β-lactamase ribosome binding site because changing the reading frame within the β-lactamase coding region (by using a set of plasmids prepared by K. Talmadge (1980)) did not affect the level of expression of IFN activity (Nagata *et al.* 1980*a*). Probably the active product was preinterferon, initiated with low efficiency at the IFN initiation triplet, and not a fused protein. If the preinterferon had the same specific activity on human cells as the mature IFN-α1 (about 2×10^7 U/mg, one-tenth that of IFN-α2 (Streuli *et al.* 1980)), then the transformed *E. coli* would have accumulated about 50 molecules per cell. The yield was increased about 1000-fold by fusing a DNA fragment containing the lac promoter and the beginning of the β-galactosidase gene (Roberts *et al.* 1979) to the leader sequence of the IFN-α1 (A. Hall & C. Weissmann, unpublished results). The interferon polypeptide produced by this and similar constructions contained a number of amino acids pertaining to the signal sequence; correct cleavage between signal and mature sequence as reported for rat pre-proinsulin (Talmadge *et al.* 1980) was not observed. Therefore more elaborate constructions were carried out, in which the signal sequence of the IFN-α2 gene was replaced by a DNA fragment containing the lac promoter and extending to the initiator AUG (figure 7). A similar construction involving the β-lactamase promoter and protein initiator region proved to be even more effective, and was adopted for the production of IFN-α2. Under large-scale (up to 30000 l) fermentation conditions, concentrations of $3–5 \times 10^8$ units of IFN per litre were attained (i.e. about 2 mg of IFN per litre, since the specific

‡ Units with reference to a secondary NIH standard.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

activity of IFN-α2 on human cells is about 2×10^8 U mg^{-1}). Interferon was purified by a procedure involving ammonium sulphate precipitation, chromatography on matrex blue, DEAE-Sepharose, Sephadex G-100 and chromatofocusing. The final preparations were homogeneous, as judged by SDS-polyacrylamide gel electrophoresis of 50 μg of protein (figure 8*a*), and were crystallized by T. Unge and B. Strandberg (Uppsala) (figure 8*b*). Purified IFN-α2 was found to be non-toxic in the rhesus monkey at many times the dosage

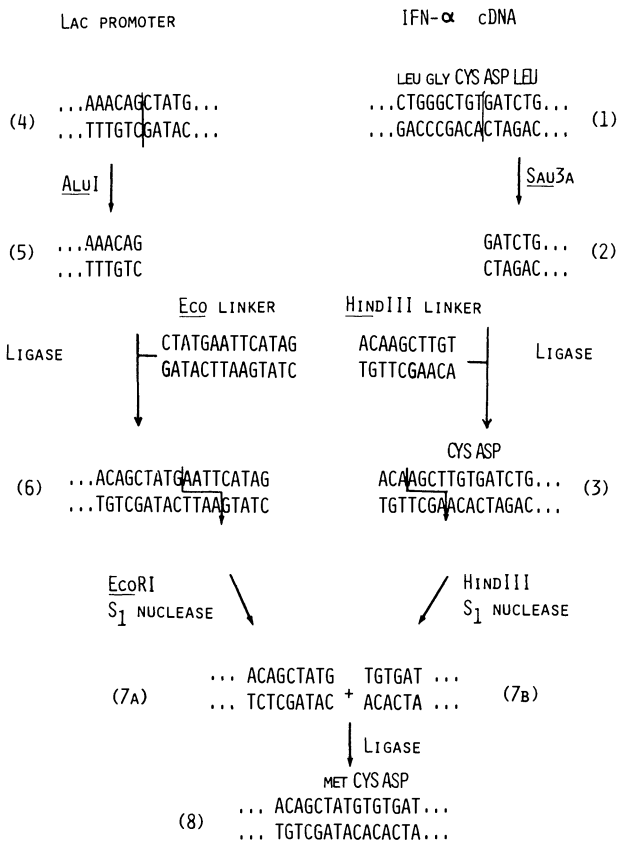


FIGURE 7. Construction of a hybrid plasmid encoding methionyl-IFNα2. The coding sequence of IFN-α2 (1) was cleaved at the *Sau*3a site (2) located between the codon for the first and second amino acid of the mature interferon sequence, and a synthetic oligonucleotide (*Hin*d III linker) was ligated to the cleaved end, reconstructing the first (cysteine) codon and adding a *Hin*d III cleavage site (3). A bacterial promoter-containing DNA fragment derived from the lac operon (4) was cleaved at an *Alu*I site preceding the AUG initiator triplet by two nucleotides (5), and a synthetic oligonucleotide (*Eco*RI linker) was ligated to the cut end reconstructing the original sequence to the end of the AUG triplet and adding an *Eco*RI site (6). The interferon sequence (3) was digested with *Hin*d III and the single-strand specific nuclease S_1 to generate a flush end immediately preceding the cysteine codon (7b). The promoter sequence (6) was digested with *Eco*RI and S_1 nuclease to give a flush end immediately after the initiator AUG codon. The two ends (7a) and (7b) were joined with DNA ligase to yield a sequence encoding mature interferon preceded by a methionine residue. When expressed in *E. coli*, the interferon molecules lose the methionine residue by post-translational cleavage.

required to fully prevent lesion formation after inoculation with vaccinia (Schellekens *et al.* 1981), and is currently being tested in phase I and phase II trials in man. Applied as an intranasal spray and followed by rhinovirus challenge it has proved to be remarkably effective in protecting human volunteers against colds (Scott *et al.* 1982).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hybrid interferons and their target cell specificity

As mentioned above, IFN-α1 and IFN-α2 have distinct target specificities: although both IFNs show about the same specific antiviral activity on bovine MDBK cells, IFN-α2 is about seven times more active on human WISH cells than IFN-α1, and IFN-α1 is 30 times more active on mouse L929 cells than IFN-α2 (figure 9). We prepared a number of hybrid IFNs, among them some consisting of the amino terminal third of IFN-α1 and the carboxy terminal two-thirds of IFN-α2 (IFN-α1(B)α2), or vice versa (IFNα2(B)α1), in the expectation that the location of the polypeptide domain (idiotope) responsible for binding to the receptor might be identified. For example, if the idiotope were located in the amino terminal third of the molecule,

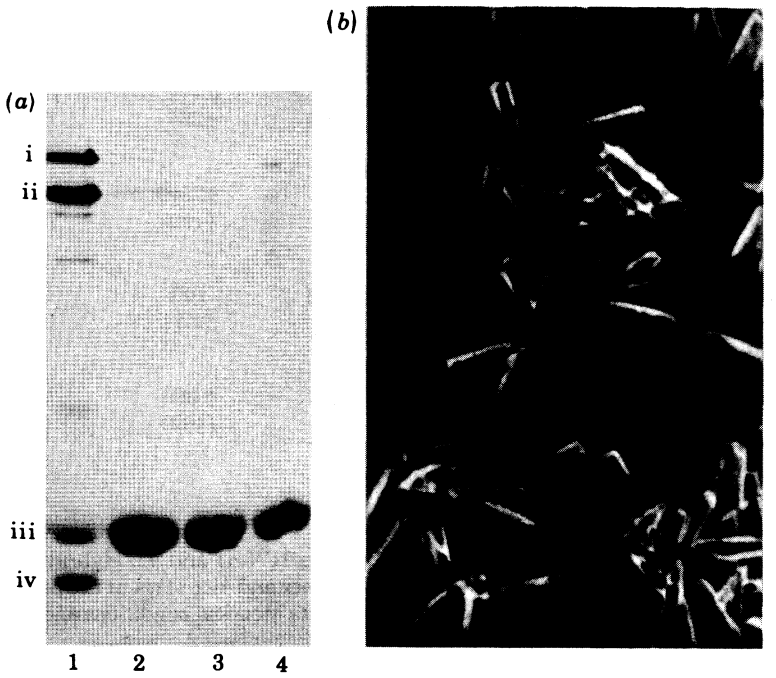


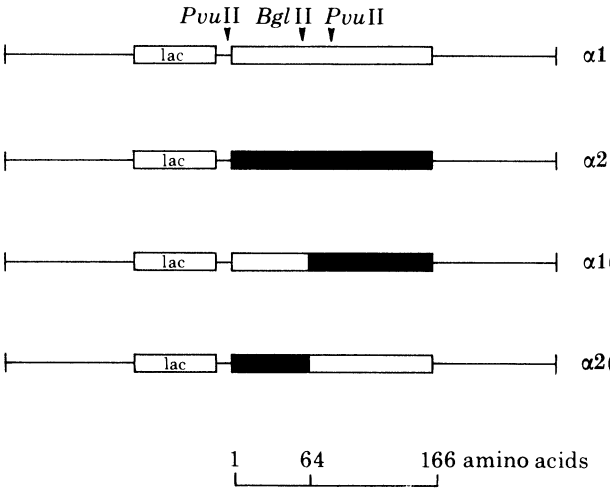
Figure 8. Purified IFN-α2 from *E. coli*. Interferon was purified from extracts of *E. coli* by using chromatography on Matrex blue, Sephadex, DEAE-Sepharose and chromatofocusing. The specific activity of the preparation was 2×10^8 units mg^{-1} (M. Fountoulakis, J. Ecsödi, C. Schein & C. Weissmann). (*a*) Sodium dodecyl sulphate–polyacrylamide gel electrophoresis. Lane 1, molecular mass markers, from top to bottom: (i) bovine serum albumin (68000), (ii) catalase (60000), (iii) β-lactoglobulin (18400), (iv) cytochrome *c* (13000). Lanes 2–4, purified IFN-α2 at 50, 30 and 20 μg, respectively. (*b*) Photomicrograph of IFN-α2 crystals prepared by T. Unge, S. Lövgren and B. Strandberg (The Wallenberg Laboratory, Uppsala).

then IFN-α1(B)α2 should show the target specificity of IFN-α1, namely relatively low activity on human cells and high activity on mouse cells. In fact we found the apparently paradoxical result that in regard to the activity on human cells, the hybrids behaved as though the amino-terminal third specified target cell specificity, whereas in regard to the activity on mouse cells, the carboxy-terminal two-thirds seemed to be determining the specificity. In addition, the specific activity of a hybrid interferon on a particular cell line could be higher or lower than that of either parent (figure 9).

Clearly, a simple model invoking a single idiotope located entirely within either the carboxy-terminal or the amino-terminal half of the molecule would be hard put to explain either of the above findings. We have therefore proposed (Streuli *et al.* 1981) that IFN possesses two distinct

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

idiotopes, one located in the amino-proximal third, the other in the carboxy-terminal two-thirds (or, as judged from the properties of hybrids consisting of half-molecules of two IFNs (Streuli *et al.* 1981), in the carboxy-terminal half) of the polypeptide; these could be spatially adjoining or remote. We postulate that each idiotope interacts with a cognate area on the receptor and that the efficiency of triggering of the antiviral response is determined by the quality of the fit of the idiotope–receptor pairs. The model shown in figure 10 illustrates how either the amino-proximal or the carboxy-proximal idiotope can co-determine target specificity, and how IFN hybrids can have specific activities higher or lower than that of either parent.



	IFN units ml^{-1} (relative specific activities)		
	human WISH	bovine MDBK	mouse L929
α1	3 500 (20)	17 000 (100)	1 800 (10)
α2	25 000 (190)	13 000 (100)	60 (0.5)
α1(B) α2	100 (1.7)	6 000 (100)	9 (0.15)
α2(B) α1	7 000 (100)	7 000 (100)	12 000 (170)

FIGURE 9. Target cell specificity of IFN-α1, α2 and hybrids between IFN-α1 and α2. Hybrid genes were prepared by joining the 5′ moiety of one gene with the 3′ moiety of another and vice versa at the position corresponding to amino acid 64. The interferons were expressed in *E. coli* under identical conditions and assayed on different target cells. (From Streuli *et al.* (1981).)

For example, consider the antiviral activity on mouse cells, and assume that the fit of the amino-proximal idiotope (A-idiotope) and the carboxy-proximal idiotope (C-idiotope) to their receptors is described by + and + + +, respectively, while the corresponding values for IFN-α2 are + and 0, respectively. The overall fit of IFN-α1 and IFN-α2 would be + + + + and +, respectively, and would result in a triggering efficiency that is some function of the overall fit (e.g. logarithmic, so that each additional + leads to a tenfold higher antiviral effect). A hybrid molecule with the A-idiotope of IFN-α1 and the C-idiotope of IFN-α2 would have a fit of +, i.e. less than either parental molecule, and the reciprocal hybrid one of + + + + +, more than either parental IFN. The characteristic behaviour on mouse cells would appear to be due to the carboxy-terminal moiety of the IFNs. The results found on human and bovine cells can be explained in like fashion.

IFN-α's and IFN-β show two major blocks of sequence homology, between amino acids 28 to 80 and 115 to 151, respectively (Taniguchi *et al.* 1980). This led us to the suggestion that these two regions might correspond to the idiotopes described above, in which case the receptors for the two interferons might be the same or closely related (Streuli *et al.* 1981); this was recently reported to be so (Branca & Baglioni 1981). Sternberg & Cohen (1982; see also

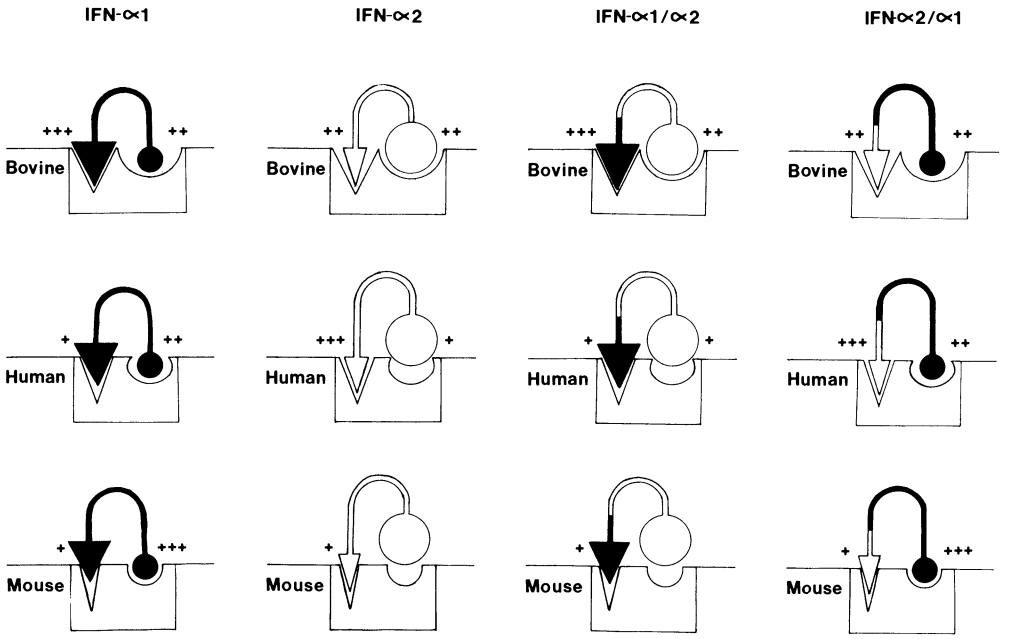


FIGURE 10. A model for the interaction of interferon and its receptor. IFN is postulated to have two binding sites, or one binding site consisting of two idiotopes, located in the amino-proximal (triangles) and carboxy-proximal (circles) moiety of the polypeptide, respectively. Black, IFN-α1; white, IFN-α2. Thin black lines indicate the profiles of the receptors, which differ for different species. The quality of the idiotope–receptor fit is indicated by the number of + signs. Triggering efficiency is a function of the two idiotope–receptor interactions.

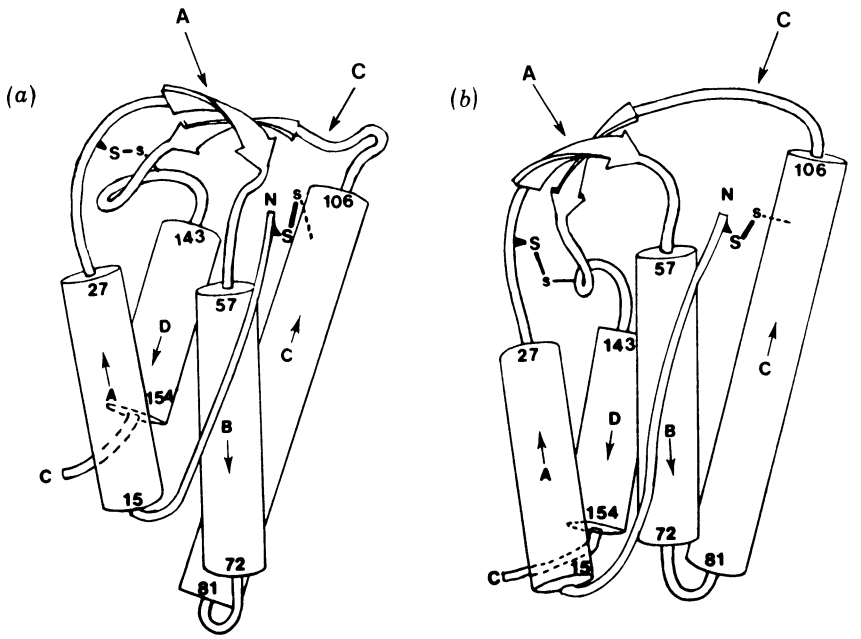


FIGURE 11. Predicted interferon structures: (*a*) and (*b*) show schematically two possible generalized IFN structures. Four α-helices (cylinders) form a right-handed bundle. The helix termini are numbered according to the IFN-β sequence; the disulphide bridges in IFN-α1 and possible β-strand regions (light, flat arrows) are indicated. (From Sternberg & Cohen (1982).) Black arrows indicate the possible positions of the amino-proximal (A) and carboxy-proximal (C) idiotopes of IFN.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Sternberg & Cohen, this symposium) have predicted a structure for interferons in which both the regions 37–48 and 105–119 lie at the extremity of a loop, on the same side of the molecule, in a fashion reconcilable with our model (figure 11).

Expression of cloned interferon genes in eukaryotic cells

We have taken three approaches in attempting to express cloned IFN-α genes in eukaryotic cells.

(*a*) *Thymidine kinase gene-linked transformation*

It has been shown that when thymidine kinase (TK)-negative mouse fibroblasts are exposed to concatenates (Mantei *et al.* 1979) or a mixture of hybrid plasmids (Wigler *et al.* 1979) containing the Herpes simplex TK gene and the rabbit β-globin gene, respectively, in the presence of calcium phosphate, most of the TK^+ transformants have also acquired one or more copies of the rabbit globin gene (Mantei *et al.* 1979; Wigler *et al.* 1979). Most of the globin DNA-containing cell lines produce correctly initiated rabbit β-globin mRNA (along with some incorrectly initiated transcripts), although the endogenous mouse β-globin gene is not expressed, as normally occurs in mouse fibroblasts (Mantei *et al.* 1979). Thus the exogenous β-globin gene is not, or only incompletely, under physiological control.

When similar transformations were carried out with the human chromosomal IFN-α1 gene, none of 12 transformed mouse clones produced detectable levels of correctly initiated IFN-α mRNA, although transcripts initiated far upstream from the gene were found. Thus, in contrast to the rabbit β-globin gene, the human IFN-α gene was not illegitimately expressed. Upon treatment with Newcastle disease virus, the transformed fibroblasts produced both mouse IFN-α mRNA and correctly initiated human IFN-α1 mRNA with the same kinetics: little if any RNA was found 4 h after infection; it rose to a maximum after 11 h, decreased after 20 h and none was detected after 27 h. We estimated that human IFN-α1 mRNA reached only about one-fiftieth of the level of the mouse IFN mRNA, perhaps because of incompatibility between the human and the mouse system, or because the site of integration of the DNA plays a role in expression (Mantei & Weissmann 1982). Similar results have been reported for the IFN-β system by Ohno & Taniguchi (1982).

(*b*) *Constitutive expression of cloned interferon genes*

Whereas the human IFN-α's examined so far do not carry carbohydrate residues (Allen & Fantes 1980; M. Rubinstein, personal communication), there is strong evidence that murine IFN-α's are glycosylated (Fujisawa *et al.* 1978), as are human IFN-β (Knight 1976; Havell *et al.* 1977; Tan *et al.* 1979) and IFN-γ (Yip *et al.* 1981). Most probably, eukaryotic proteins produced in *E. coli* will be glycosylated either incorrectly or not at all. It is therefore desirable to be able to prepare glycosylated IFNs and other glycoproteins in eukaryotic cells transformed by cloned genes. To obtain constitutive expression, we linked the coding sequence of murine pre-IFN-α1 (G. Shaw & S. Nagata, unpublished results) to the SV40 early promoter, using the plasmid pKCR constructed by O'Hare *et al.* (1981) as vector. The coding sequence is followed by an intron-containing *Bam-Pvu*II DNA segment from the rabbit β-globin gene and tandem polyadenylation sites from the rabbit β-globin gene and from the SV40 early region. When monkey (Cos) cells (Gluzman 1981) were transformed with this plasmid (which was first cleaved with endonuclease *Tha*I to excise the pBR322 sequences, leaving the eukaryotic and

viral moiety intact) by the calcium phosphate method (Wigler *et al.* 1978), antiviral activity was found in the cell supernatant beginning on the second day, and reached a maximum of 14000 units ml^{-1} between the third and tenth day, after which the activity decreased (H. Taira & G. Shaw, unpublished results). It is interesting to note that the uncleaved hybrid plasmid gave values of about one-third of this, which may be due to the inhibitory effects of certain pBR322 sequences on the replication of the hybrid plasmid in the transformed cell (Lusky & Botchan 1981). The interferon produced by this transient expression system could be concentrated and

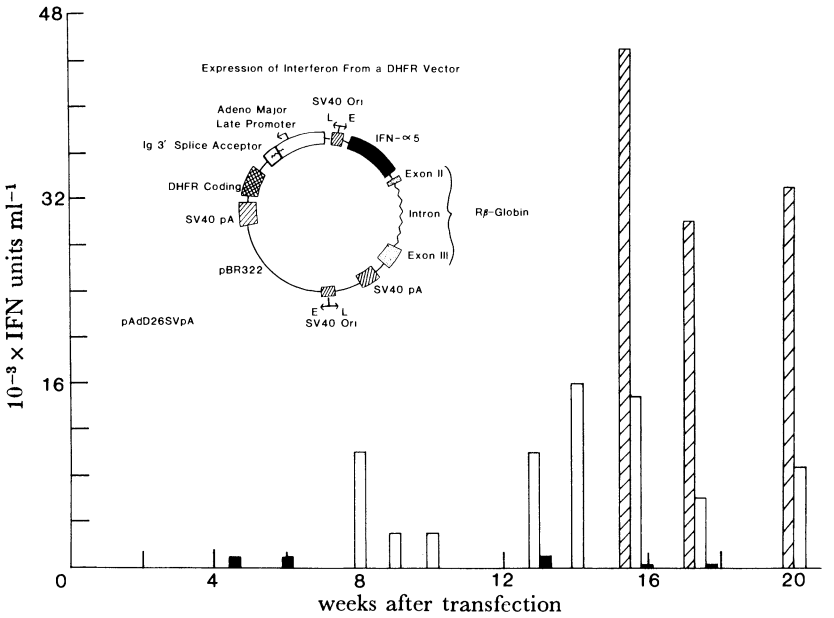


FIGURE 12. Constitutive production of IFN-α by transformed cell lines. Dihydrofolate reductase (DHFR)-negative hamster cells (provided by L. Chasin) were transformed with the plasmid depicted in the inset, which contains the mouse DHFR cDNA sequence under the control of the adeno major late promoter (provided by R. Kaufman, S. Latt & P. Sharp, who also devised the approach to expression of cloned cDNA described in this figure) and the IFN-α5 gene under the control of the SV40 early promoter. The transformed cells were cultured in selective medium (lacking adenosine, deoxyadenosine and thymidine). Several surviving clones were grown to confluency and the IFN activity accumulating in the supernatant in 24 h was determined (solid columns, clone α5-2N). Cells derived from clone α5-2N were exposed to 0.05 μM methotrexate and clones growing up in selective medium were tested for IFN production as above (open columns, average of three cell lines). Cells resistant to 0.05 μM methotrexate were selected for resistance to 1 μM methotrexate and tested as above (hatched columns, average of four cell lines).

purified 1000-fold by adsorption to and elution from a controlled pore glass column as described by Taira *et al.* (1980). It was fully neutralized by antibody to natural murine interferon (obtained from M. Aguet) and is currently being further characterized.

(*c*) *The dihydrofolate reductase* (*DHFR*) *amplification system*

Although the approach described above allows the preparation of small amounts of cloned gene products, it is not convenient for the production of more substantial quantities, mainly because of the transient nature of the synthesis.

It has been shown by Alt *et al.* (1978) that when cell cultures are exposed to appropriate levels of methotrexate, an inhibitor of DHFR, clones of surviving cells are found to have amplified their DHFR gene (and regions adjoining it) up to several hundredfold. R. Kaufman,

S. Latt & P. Sharp (personal communication) have constructed a plasmid containing both the mouse DHFR cDNA under the control of the adeno major late promoter, and the SV40 T antigen gene. They showed that DHFR-negative cells transformed by the hybrid plasmid and subjected to methotrexate selection contained a large number of integrated plasmids and produced a high level of T antigen.

We constructed a hybrid that contained the DHFR moiety of Kaufman's plasmid, as well as the IFN-α5 gene placed under the control of the SV40 early promoter and followed by the rabbit β-globin large intron and the SV40 polyadenylation sequence (see figure 12, inset). This plasmid was introduced into DHFR$^-$ hamster cells (prepared by L. Chasin) and DHFR$^+$ colonies were selected. When grown to confluency, the cells excreted interferon at about $0.5–1.0 \times 10^3$ units ml^{-1} d^{-1}. A first round of selection at 0.05 μM methotrexate yielded cell cultures producing $10–20 \times 10^3$ units ml^{-1} d^{-1}. After a second round of selection, at 1 μM methotrexate, cultures producing $30–50 \times 10^3$ units ml^{-1} d^{-1} were obtained. This may be compared with the value of $5–20 \times 10^3$ units ml^{-1} of mixed interferons produced by induced, cultured lymphoblastoid cells (Rubinstein *et al.* 1979). As shown in figure 12, the production of interferon by the different cell lines remained more or less constant over weeks and months, respectively. We believe that this approach will make it possible to produce a glycosylated protein constitutively at levels not attainable by conventional means.

This work was supported by the Kanton of Zürich, the Schweizerische Nationalfonds and Biogen S. A. We thank Ms R. M. von Rotz for her invaluable contributions and patience in preparing the typescript for this paper.

References

Allen, G. & Fantes, K. H. 1980 *Nature, Lond.* **287**, 408–411.
Alt, F. W., Kellems, R. E., Bertino, J. R. & Schimke, R. T. 1978 *J. biol. Chem.* **253**, 1357–1370.
Baltimore, D. 1981 *Cell* **24**, 592–594.
Brack, C., Nagata, S., Mantei, N. & Weissmann, C. 1981 *Gene* **15**, 379–394.
Branca, A. A. & Baglioni, C. 1981 *Nature, Lond.* **294**, 768–770.
Chan, S. J., Noyes, B. E., Agarwal, K. L. & Steiner, D. F. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 5036–5040.
Chang, A. C. Y., Nunberg, J. H., Kaufman, R. J., Erlich, H. A., Schimke, R. T. & Cohen, S. N. 1978 *Nature, Lond.* **275**, 617–624.
Colman, A. & Morser, J. 1979 *Cell* **17**, 517–526.
Crea, R., Kraszweski, A., Hirose, T. & Itakura, K. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 5765–5769.
Epstein, L. B. 1982 *Nature, Lond.* **295**, 453–454.
Fujisawa, J., Iwakura, Y. & Kawade, Y. 1978 *J. biol. Chem.* **253**, 8677–8679.
Goeddel, D. V., Leung, D. W., Dull, T. J., Gross, M., Lawn, R. M., McCandliss, R., Seeburg, P. H., Ullrich, A. Yelverton, E. & Gray, P. W. 1981 *Nature, Lond.* **290**, 20–26.
Goeddel, D. V., Yelverton, E., Ullrich, A., Heyneken, H. L., Miozzari, 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, P. C. & Pestka, S. 1980 *Nature, Lond.* **287**, 411–417.
Gluzman, Y. 1981 *Cell* **23**, 175–182.
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. & Goeddel, D. V. 1982 *Nature, Lond.* **295**, 503–508.
Grunstein, M. & Hogness, D. S. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 3961–3965.
Harpold, M. M., Dobner, P. R., Evans, R. M. & Bancroft, F. C. 1978 *Nucl. Acids Res.* **5**, 2039–2053.
Havell, E. A., Yamazaki, S. & Vilček, J. 1977 *J. biol. Chem.* **252**, 4425–4427.
Hood, L., Campbell, J. H. & Elgin, S. C. R. 1975 *A. Rev. Genet.* **9**, 305–353.
Itakura, K., Hirose, T., Crea, R., Riggs, A. D., Heyneker, H. L., Bolivar, F. & Boyer, H. W. 1977 *Science, Wash.* **198**, 1056–1063.
Knight, E., Jr 1976 *Proc. natn. Acad. Sci. U.S.A.* **73**, 520–523.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Kronenberg, H. M., McDevitt, B. E., Majzoub, J. A., Nathans, J., Sharp, P. A., Potts, J. T. & Rich, A. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 4981–4985.
Lawn, R. M., Adelman, J., Dull, T. J., Gross, M., Goeddel, D. V. & Ullrich, A. 1981*a* *Science, Wash.* **212**, 1159–1162.
Lawn, R. M., Adelman, J., Franke, A. E., Houck, C. M., Gross, M., Najarian, R. & Goeddel, D. V. 1981*c* *Nucl. Acids Res.* **9**, 1045–1052.
Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. 1978 *Cell* **15**, 1157–1174.
Lawn, R. M., Gross, M., Houck, C. H., Franke, A. E., Gray, P. W. & Goeddel, D. V. 1981*b* *Proc. natn. Acad. Sci. U.S.A.* **78**, 5435–5439.
Levy, W. P., Rubinstein, M., Shively, J., Valle, U. D., Lai, C.-Y., Moschera, J., Brink, L., Gerber, L., Stein S. & Pestka, S. 1982 (In the press.)
Lusky, M. & Botchan, M. 1981 *Nature, Lond.* **293**, 79–81.
Maeda, S., McCandliss, R., Chiang, T.-R., Costello, L., Levy, W. P., Chang, N. T. & Pestka, S. 1981 In *Developmental biology using purified genes* (ed. D. D. Brown), pp. 85–96. New York: Academic Press.
Mantei, N., Boll, W. & Weissmann, C. 1979 *Nature, Lond.* **281**, 40–46.
Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S. & Weissmann, C. 1980 *Gene* **10**, 1–10.
Mantei, N. & Weissmann, C. 1982 *Nature, Lond.* **297**, 128–132.
Masucci, M. G., Szigeti, R., Klein, E., Klein, G., Gruest, J., Montagnier, L., Taira, H., Hall, A., Nagata, S. & Weissmann, C. 1980 *Science, Wash.* **209**, 1431–1435.
Miyata, T. & Hayashida, H. 1982 *Nature, Lond.* **295**, 165–168.
Nagata, S., Mantei, N. & Weissmann, C. 1980*b* *Nature, Lond.* **287**, 401–408.
Nagata, S., Taira, H., Hall, A., Johnsrud, L., Streuli, M., Ecsödi, J., Boll, W., Cantell, K. & Weissmann, C. 1980*a* *Nature, Lond.* **284**, 316–320.
O'Hare, K., Benoist, C. & Breathnach, R. 1981 *Proc. natn. Acad. Sci. U.S.A.* **78**, 1527–1531.
Ohno, S. & Taniguchi, T. 1981 *Proc. natn. Acad. Sci. U.S.A.* **78**, 5305–5309.
Ohno, T. & Taniguchi, T. 1982 *Nucl. Acids. Res.* **10**, 967–977.
Owerbach, D., Rutter, W. J., Shows, T. B., Gray, P., Goeddel, D. V. & Lawn, R. M. 1981 *Proc. natn. Acad. Sci. U.S.A.* **78**, 3123–3127.
Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. 1980 *Cell* **20**, 555–566.
Radding, C. M. 1978 *A. Rev. Biochem.* **47**, 847–880.
Reynolds, F. H., Jr, Premkumar, E. & Pitha, P. M. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 4881–4885.
Roberts, T. M., Kacich, R. & Ptashne, M. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 760–764.
Rubinstein, M., Rubinstein, S., Familetti, P. C., Müller, R. S., Waldman, A. A. & Pestka, S. 1979 *Proc. natn. Acad. Sci. U.S.A.* **76**, 640–644.
Schellekens, H., de Reus, A., Bolhuis, R., Fountoulakis, M., Schein, C., Ecsödi, J., Nagata, S. & Weissmann, C. 1981 *Nature, Lond.* **292**, 775–776.
Scott, G. M., Phillpotts, R. J., Wallace, J., Tyrrell, D. A. J., Gauci, C. L. & Greiner, J. 1982 (Submitted.)
Seeburg, P. H., Shine, J., Martial, J. A., Baxter, J. D. & Goodman, H. M. 1977 *Nature, Lond.* **270**, 486–494.
Sehgal, P. B. & Sagar, A. D. 1980 *Nature, Lond.* **288**, 95–97.
Slate, D. L., D'Eustachio, P., Pravtcheva, D., Cunningham, A. C., Nagata, S., Weissmann, C. & Ruddle, F. H. 1982 *J. exp. Med.* **155**, 1019–1024.
Smith, G. P. 1976 *Science, Wash.* **191**, 528–535.
Sternberg, M. J. E. & Cohen, F. E. 1982 *Int. J. Biol. Macromol.* **4**, 137–144.
Stewart, W. E. II, Sarkar, F. H., Taira, H., Hall, A., Nagata, S. & Weissmann, C. 1980 *Gene* **11**, 181–186.
Stewart, W. E. II & Sulkin, S. E. 1966 *Proc. Soc. exp. Biol. Med.* **123**, 650–653.
Streuli, M., Hall, A., Boll, W., Stewart, W. E. II, Nagata, S. & Weissmann, C. 1981 *Proc. natn. Acad. Sci. U.S.A.* **78**, 2848–2852.
Streuli, M., Nagata, S. & Weissmann, C. 1980 *Science, Wash.* **209**, 1343–1347.
Taira, H., Broeze, R. J., Jayaram, B. M., Lengyel, P., Hunkapiller, M. W. & Hood, L. E. 1980 *Science, Wash.* **207**, 528–530.
Talmadge, K., Kaufman, J. & Gilbert, W. 1980 *Proc. natn. Acad. Sci. U.S.A.* **77**, 3988–3922.
Tan, Y. H., Baraket, F., Berthold, W., Smith-Johannsen, H. & Tan, C. 1979 *J. biol. Chem.* **254**, 8067–8073.
Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. & Weissmann, C. 1980 *Nature, Lond.* **285**, 2848 -2852.
Tavernier, J., Derynck, R. & Fiers, W. 1981 *Nucl. Acids Res.* **9**, 461–471.
Ullrich, A., Shine, J., Chirgwin, J., Pictet, R., Tischer, E., Rutter, W. J. & Goodman, H. M. 1977 *Science, Wash.* **196**, 1313–1319.
Villa-Komaroff, L., Efstratiadis, A., Broome, S., Lomedico, P., Tizard, R., Naber, S. P., Chick, W. L. & Gilbert, W. 1978 *Proc. natn. Acad. Sci. U.S.A.* **75**, 3727–3731.
Weissenbach, J., Chernajovsky, Y., Zeevi, M., Shulman, L., Soreq, H., Nir, U., Wallach, D., Perricaudet, M., Tiollais, P. & Revel, M. 1980 *Proc. natn. Acad. Sci. U.S.A.* **77**, 7152–7156.
Wigler, M., Pellicer, A., Silverstein, S. & Axel, R. 1978 *Cell* **14**, 725–731.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Wigler, M., Sweet, R., Sim, G. K., Wold, B., Pellicer, A., Lacy, E., Maniatis, T., Silverstein, S. & Axel, R. 1979 *Cell* **16**, 777–785.
Yip, Y. K., Pang, H. L., Urban, C. & Vilček, J. 1981 *Proc. natn. Acad. Sci. U.S.A.* **78**, 1601–1605.
Zoon, K. C., Smith, M. E., Bridgen, P. J., Anfinsen, C. B., Hunkapiller, M. W. & Hood, L. E. 1980 *Science, Wash.* **207**, 527–528.

Discussion

In reply to questions, Professor Weissmann said that the human interferon genes present in mouse cells did not produce detectable interferon. Human interferon mRNA could be detected by hybridization in about 20–30 % of the clones, but only about 10 interferon mRNA molecules per cell were formed, the lower limit of detection being about 1 molecule per cell. There was thus too little mRNA present for the interferon to be detected. He suggested that the reason that so few mRNA molecules were formed was because transcription took place in a heterologous cell, and attempts to carry out these experiments in homologous cells were now being made, using an interferon marked by structural changes to distinguish between mRNA made by using the genomic DNA or the introduced DNA. He also stated, in response to another question, that the inserted pieces of DNA were not very large – with about 1 kilobase at the 5′ end of the gene and 1.5 kilobases at the 3′ end. The terminal methionine of the interferons made in bacteria was sometimes removed and sometimes not.

The structure of the genes was then discussed. Professor Weissmann thought that they should be regarded as a family of genes, some of which were arising and others disappearing. There was no clear understanding of why there were no introns, but it was suggested that this might be why interferon is still synthesized when host synthesis is shut off. The significance of the conserved sequence is not understood but is currently being investigated, as is the presence of interferon genes in primitive organisms such as prochordates.

It seems likely that different IFN genes have different functions, which should be explored, but there was some worry that hybrid genes might carry new antigenic determinants and so be immunogenic. Nevertheless 10 genes are expressed in man and are not antigenic. It seems probable that most individuals carry the same range of 14 genes, though little is known yet about the distribution of genes through the population.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

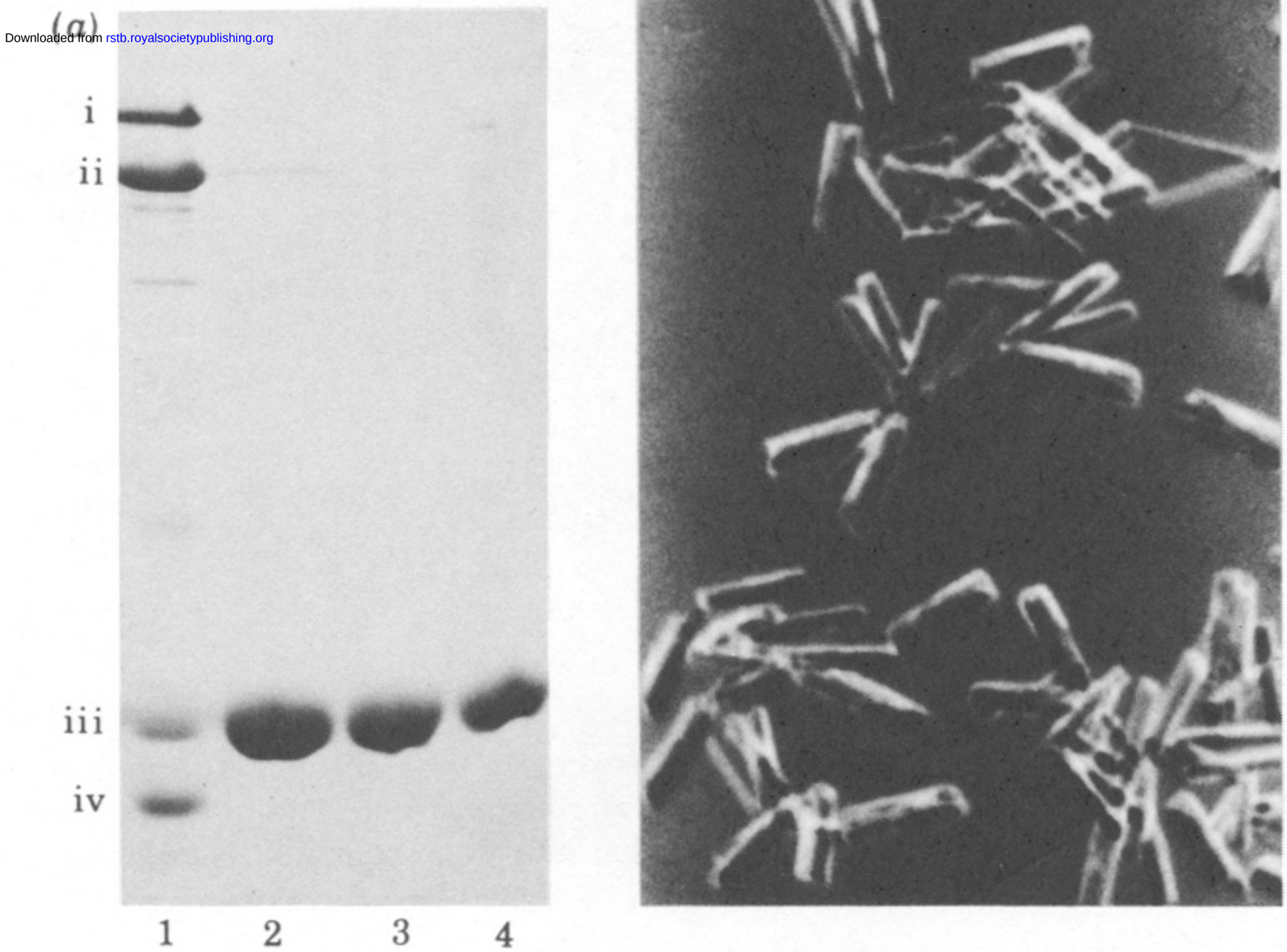


FIGURE 8. Purified IFN-α2 from *E. coli*. Interferon was purified from extracts of *E. coli* by using chromatography on Matrex blue, Sephadex, DEAE-Sepharose and chromatofocusing. The specific activity of the preparation was 2×10^8 units mg^{-1} (M. Fountoulakis, J. Ecsödi, C. Schein & C. Weissmann). (*a*) Sodium dodecyl sulphate–polyacrylamide gel electrophoresis. Lane 1, molecular mass markers, from top to bottom: (i) bovine serum albumin (68000), (ii) catalase (60000), (iii) β-lactoglobulin (18400), (iv) cytochrome *c* (13000). Lanes 2–4, purified IFN-α2 at 50, 30 and 20 μg, respectively. (*b*) Photomicrograph of IFN-α2 crystals prepared by T. Unge, S. Lövgren and B. Strandberg (The Wallenberg Laboratory, Uppsala).