Biochem. J. (1991) 276, 511–518 (Printed in Great Britain)

Natural human interferon-α2 is *O*-glycosylated

Günther R. ADOLF,* Inge KALSNER,* Horst AHORN,* Ingrid MAURER-FOGY* and Kari CANTELL†
* Ernst-Boehringer-Institut für Arzneimittelforschung, Bender + Co Ges mbH, Dr. Boehringer-Gasse 5-11, A-1121 Vienna, Austria, and † National Public Health Institute, SF-00300 Helsinki, Finland

Natural human interferon α_2 (IFN-α2) was isolated from a preparation of partially purified human leucocyte IFN by monoclonal-antibody immunoaffinity chromatography. The purified protein had a specific activity of 1.5×10^8 i.u./mg; it was estimated to constitute 10–20% of the total antiviral activity of leucocyte IFN. *N*-Terminal amino-acid-sequence analysis identified the subspecies IFN-α2b and/or IFN-α2c, whereas IFN-α2a was not detectable. The structure of natural IFN-α2 was found to differ from that of its recombinant (*Escherichia coli*-derived) equivalent. First, reverse-phase h.p.l.c. showed that natural IFN-α2 was significantly more hydrophilic then expected. Secondly, the apparent molecular mass of the natural protein determined by SDS/PAGE was higher than that of recombinant IFN-α2; incubation under mild alkaline conditions known to eliminate *O*-linked carbohydrates resulted in a reduction of the apparent molecular mass to that of the recombinant protein. On sequence analysis of proteolytic peptides, Thr-106 was found to be modified. These results suggested that Thr-106 of natural IFN-α2 carries *O*-linked carbohydrates. Reverse-phase h.p.l.c. as well as SDS/PAGE of natural IFN-α2 showed that glycosylation is heterogeneous. For characterization of the carbohydrate moieties, the protein was treated with neuraminidase and/or *O*-glycanase and analysed by gel electrophoresis; in addition, glycopeptides obtained by proteinase digestion and separated by h.p.l.c. were characterized by sequence analysis and m.s. Further information on the composition of the glycans was obtained by monosaccharide analysis. The results indicate that natural IFN-α2 contains the disaccharide galactosyl-*N*-acetylgalactosamine (Gal-GalNAc) linked to Thr-106. In part of the molecules, this core carbohydrate carries (α-)*N*-acetylneuraminic acid, whereas a disaccharide, probably *N*-acetyl-lactosamine, is bound to Gal-GalNAc in another proportion of the protein. Further glycosylation isomers are present in small amounts. As IFN-α2 is the only IFN-α species with a threonine residue at position 106, it may represent the only *O*-glycosylated human IFN-α protein.

INTRODUCTION

At least 15 functional genes in the human genome code for proteins of the IFN-α family. These molecules are closely related in structure, with amino acid sequence similarities generally in the region of about 90% (Henco *et al.*, 1985). IFN-α proteins consist of 166 amino acids (with the exception of IFN-α2, which has 165 amino acids) and characteristically contain four conserved cysteine residues which form two disulphide bridges. IFN-α species are slightly acidic in character; with one exception (IFN-α14), the proteins lack a recognition site for asparagine-linked glycosylation. Two other human IFN species, namely IFN-ω1 and IFN-β, are more distantly related to IFN-α; these proteins are *N*-glycosylated. IFN-α, -β and -ω, collectively referred to as 'class I IFNs', bind to the same high-affinity cell membrane receptor. Originally discovered on the basis of their antiviral activity, class I IFNs have since been shown to exert a multitude of biological effects; several preparations of natural IFN-α, natural IFN-β and recombinant IFN-α2 have been registered for therapeutic application in a range of viral and neoplastic diseases.

The biological significance of the multitude of IFN-α species is still unclear. After viral infection, many, if not all, of the IFN-α genes are expressed at least at the mRNA level; quantitative differences have been described (Goeddel *et al.*, 1981; Hiscott *et al.*, 1984). Analysis of IFN-α produced by leukaemic human leucocytes and transformed human cell lines has also demonstrated the presence of multiple IFN proteins (Allen & Fantes, 1980; Rubinstein *et al.*, 1981; Allen, 1982; Hobbs & Pestka, 1982). Although the various species of IFN-α generally have very similar biological activities, quantitative differences, e.g. in their specific antiviral activity or their ability to stimulate natural killer cells, have been reported (see e.g., Fish *et al.*, 1983; Weck *et al.*, 1981; Ortaldo *et al.*, 1984).

In a series of experiments aiming at the isolation and characterization of IFN species present in human leucocyte IFN, we have attempted to utilize the superior selectivity of monoclonal antibodies (MAbs) not only to purify IFN from non-IFN contaminants, but also to isolate single species of IFN. In a previous study we used this approach to isolate natural IFN-ω1 and have defined some of the properties of this molecule (Adolf *et al.*, 1990). We now describe the purification and some characteristics of natural IFN-α2. Unexpectedly, we have found that natural IFN-α2 carries *O*-linked carbohydrate residues.

EXPERIMENTAL

IFN bioassay

The antiviral activity of the IFN preparations was determined in cytopathic-effect-reduction assays performed in microtitre plates, using human A549 lung-carcinoma cells challenged with encephalomyocarditis virus. Details of this procedure have been described (Adolf, 1987, 1990). In each bioassay, all titrations were performed in duplicate. A laboratory reference preparation of recombinant human IFN-α2c was included in each assay; the activity of this preparation had previously been assigned by comparison with the international reference preparation for recombinant human IFN-α2, Gxa01-901-535. All IFN activities observed were corrected with regard to the assigned potency of this reference preparation.

IFN e.l.i.s.a.

An e.l.i.s.a. was set up that employs two neutralizing murine IgG MAbs to IFN-α and an IFN-α2c laboratory reference

Abbreviations used: IFN, interferon; MAb, monoclonal antibody.

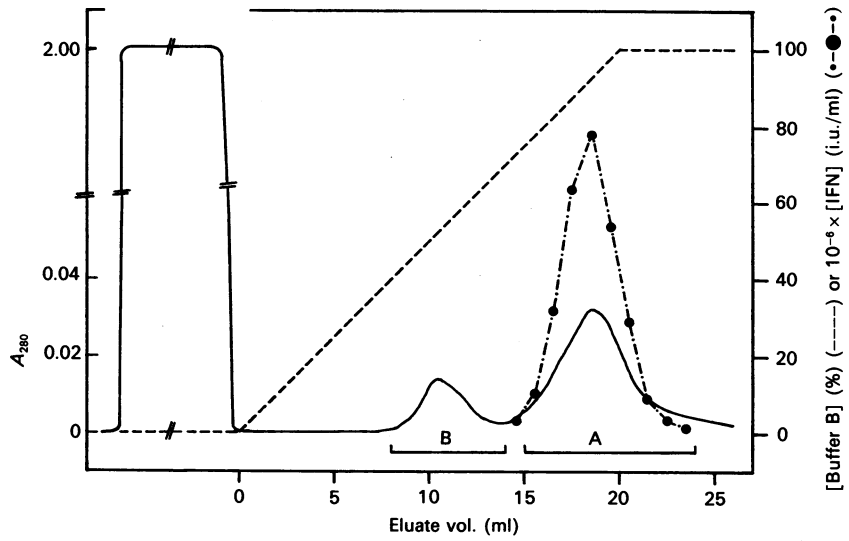

Fig. 1. MAb-affinity chromatography of human leucocyte IFN

Partially purified leucocyte IFN was applied to the column; IFN bound was eluted in a pH gradient. IFN activity in the fractions was determined by e.l.i.s.a., using recombinant IFN-α2c as a standard. The flow-through and eluate fractions designated B also contained immunoreactive material; however, this activity could not be quantified (see the Results section).

preparation (see above) as a standard. Development of the antibodies and some of their properties have been described (Adolf *et al.*, 1982, 1987). Antibody EBI-1 was used for coating of the assay plates; antibody EBI-10 covalently coupled to horseradish peroxidase was added together with the sample. *o*-Phenylenediamine and $NaBH_4$ were used as substrates for the enzyme; the reaction was stopped by addition of H_2SO_4, and the absorbance of the resulting product was determined ($A_{492} - A_{690}$).

Purification of natural human IFN-α2

An affinity column was prepared by coupling 12 mg of MAb EBI-10 [purified from mouse ascites by $(NH_4)_2SO_4$ precipitation and protein G-affinity chromatography using standard procedures] to 1 g of CNBr-activated Sepharose 4B according to the recommendations of the manufacturer (Pharmacia). The final bed volume of the column was approx. 3 ml. A preparation of partially purified human leucocyte IFN (Cantell *et al.*, 1981*a*,*b*), depleted of its IFN-ω1 component (Adolf *et al.*, 1990), containing about $(2–3) \times 10^6$ i.u./ml at a total protein concentration of 2 mg/ml, was applied to the column at a flow rate of 1 ml/min; in two separate runs, 200 and 350 ml were applied. The column was then washed with 0.1 M-sodium phosphate buffer, pH 7.5 (buffer A), and eluted with a linear gradient formed of buffer A and buffer B (0.1 M-sodium citrate, pH 2.1) in a Pharmacia f.p.l.c. system (flow rate 1 ml/min). Fractions were assayed for IFN activity by e.l.i.s.a. Appropriate fractions of both runs were pooled, neutralized with 1 M-NaOH, and re-applied to the same column re-equilibrated with buffer A (flow rate 0.25 ml/min). The same elution program was then used again. Appropriate fractions were again pooled, neutralized, and stored frozen in aliquots until further analysis.

SDS/PAGE, h.p.l.c. techniques and amino-acid-sequence analysis

SDS/PAGE and reverse-phase h.p.l.c. were used to analyse preparations of purified IFN-α2; all methods were previously described in detail (Adolf *et al.*, 1990). Determination of the *N*-terminal amino acid sequence was performed in an Applied Biosystems model 477A pulsed liquid-phase sequencer; amino acid derivatives were analysed on-line by reverse-phase h.p.l.c. (Adolf *et al.*, 1990). Highly purified recombinant (*E. coli*-derived) human IFN-α2c used for comparison was kindly provided by Dr. G. Bodo.

Mapping of proteolytic peptides

Affinity-purified IFN-α2 was further purified, denatured and desalted by reverse-phase h.p.l.c. as described (Adolf *et al.*, 1990). The peak fractions were collected and dried in a Speed-Vac concentrator. Portions [29 μg (peak 1) and 100 μg of protein (peak 2) respectively] were redissolved in 0.1 ml of 1% NH_4HCO_3; 0.5 and 1 μg respectively of trypsin (sequencing grade, Boehringer Mannheim) in 3 and 6 μl respectively of 0.01% trifluoroacetic acid was added, and the reaction mixture was incubated at 37 °C. After 6 h incubation, the same amount of trypsin was added again, and incubation was continued for further 18 h. The reaction mixtures were reduced for 2 h at room temperature by the addition of 10 μl of 0.5 M-dithiothreitol and 100 μl of 6 M-urea before analysis. Reverse-phase h.p.l.c. was performed on a DeltaPak C_{18} column [Waters; 3.9 mm × 150 mm, particle size 5 μm; pore diameter 10 nm (100 Å)] at 30 °C, using the following solvents: solvent A, 0.1% trifluoroacetic acid in water; solvent B, 0.1% trifluoroacetic acid in acetonitrile. The following gradient program was used (flow rate 1 ml/min): 0–55 min, 0–55% B (linear gradient); 55–70 min, 55% B. Peptides were detected by their absorption at 214 and 280 nm. The resulting maps were compared with a map of recombinant (*E. coli*)-derived IFN-α2c prepared similarly. Peptides derived from natural IFN-α2 that differed in their elution behaviour from their recombinant counterparts were collected and subjected to *N*-terminal sequencing, or were cleaved further with *Staphylococcus aureus* V8 proteinase (endopeptidase Glu-C; sequencing grade; Boehringer-Mannheim). Portions [0.88 μg (peak 1/I), 2.6 μg (peak 2/Ib) and 1.5 μg (peak 2/Ic)] of the peptides were each dissolved in 0.1 ml of 25 mM-phosphate buffer, pH 7.8. Proteinase dissolved in water was added (17.5, 52.5 and 29 ng respectively), and the reaction mixtures were incubated at 37 °C. After 6 h, the same amounts of proteinase were added again, and the incubation was continued for another 18 h. The samples were then submitted to reverse-phase h.p.l.c. analysis as described above. Appropriate fractions were collected and subjected to *N*-terminal sequencing.

Deglycosylation of IFN-α2

Purified, denatured and desalted IFN-α2 was treated with *Vibrio cholerae* neuraminidase (50 munits/ml; 18 h at 37 °C in 20 μl of 50 mM-sodium acetate (pH 5.5)/4 mM-$CaCl_2$) and/or endo-α-*N*-acetylgalactosaminidase (100 munits/ml, 18 h at 37 °C in the same buffer). Both enzymes were obtained from Boehringer-Mannheim. Chemical cleavage was achieved by incubation in 0.1 M-NaOH for 20 h at room temperature.

^{252}Cf plasma-desorption m.s.

Mass spectra of tryptic peptides were obtained on a BIO-ION 20 time-of-flight mass spectrometer (BIO-ION, Nordic AB, Uppsala, Sweden). Samples were dissolved in aq. 0.1% trifluoroacetic acid and deposited on nitrocellulose-coated targets (BIO-ION). Spectral accumulation times ranged between 0.5 and 12 h, depending on the yield. The primary ion start rate was 1000 fission fragments/s. Spectra were obtained at an acceleration voltage of 17 kV. Mass assignments were calculated from peak centroids, using H^+ and Na^+ for calibration.

Monosaccharide analysis

Separation of monosaccharides was carried out on a Dionex BioLC system consisting of a Gradient Pump Module (GPM II), an Eluent Degas Module (EDM II), a Basic Chromatography Module, and a CarboPac PA-1 column (4.6 mm × 250 mm) in series with a CarboPac PA guard column. The system was equipped with a pulsed amperometric detector (PAD II) containing a gold electrode to which potentials of $E_1 = 0.05$ V, $E_2 = 0.6$ V and $E_3 = -0.6$ V were applied for duration times $T_1 = 300$ ms, $T_2 = 120$ ms and $T_3 = 60$ ms respectively. All solvents were degassed by saturation with helium and stored in closed pressurized vessels with helium gas sparging. For determination of neutral monosaccharides, total hydrolysis of peak 1 protein was performed by incubating the sample with 1 M-trifluoroacetic acid for 2 h at 100 °C. After drying the sample in a SpeedVac concentrator it was redissolved in 0.2 ml of distilled water and injected. Separation of monosaccharides was achieved by isocratic elution with 0.02 M-NaOH (flow rate 1 ml/min). For determination of sialic acid, tryptic peptides were incubated in 0.05 M H_2SO_4 for 1 h at 80 °C, neutralized with 0.4 M-NaOH and injected. The column was eluted with 0.15 M-sodium acetate in 0.1 M-NaOH (flow rate 1 ml/min). For detection of both neutral monosaccharides and sialic acids, 0.3 M-NaOH was added to the column effluent at a flow rate of 0.5 ml/min.

RESULTS

Purification of natural human IFN-α2

A preparation of human leucocyte IFN, derived from Sendai-virus-induced human peripheral blood leucocytes and partially purified according to the procedure developed by Cantell *et al.* (1981*a*,*b*), was used as the starting material for isolation of IFN-α2. IFN-ω1 present in this preparation (10–15% of the total IFN activity) has been selectively removed in previous experiments by affinity chromatography on MAb OMG-4 (Adolf *et al.*, 1990). The flowthrough of the OMG-4 column had a specific antiviral activity of (1–2) × 10^6 i.u./mg; IFN-α, with a specific activity of about 2 × 10^8 i.u./mg, thus constitutes at most 1% of the total protein. To purify IFN-α2 from contaminant proteins and, at the same time, from other IFN-α species, we again relied on the high selectivity of MAbs. One of a series of antibodies we have developed, EBI-10 (IgG1), was previously shown to be quite specific for IFN-α2; whereas this antibody was able to inhibit the antiviral activity of recombinant IFN-α2 at concentrations below 1 μg/ml, other recombinant IFN-α species were neutralized only at 1000-fold higher concentrations (IFN-α1, IFN-αF) or not at all (IFN-αB, IFN-αC; Adolf, 1987). To demonstrate that EBI-10 will recognize all three known variants of IFN-α2, i.e. IFN-α2a, -2b and -2c (see below), an e.l.i.s.a. using EBI-10 as the detector antibody (see below) was carried out. All three IFN-α2 variants gave parallel titration curves and almost identical quantitative responses (results not shown).

An immunoaffinity column was prepared by coupling antibody EBI-10 to CNBr-activated Sepharose 4B. The leucocyte IFN preparation was applied to the column; about 30% of the antiviral activity was bound (mean; two different preparations). The column was eluted using a linear buffer gradient formed by 0.1 M-sodium phosphate, pH 7.5, and 0.1 M-sodium citrate, pH 2.1. This buffer system results in a slightly concave pH gradient (results not shown) and was expected to resolve bound IFN-α species in the order of the affinity for the antibody. Two protein peaks were observed in the eluate (Fig. 1); fractions were analysed for their IFN-α content by a two-site e.l.i.s.a. that uses antibody EBI-10 as well as a second MAb, EBI-1 (murine IgG1), which also shows high affinity for IFN-α2 (Adolf *et al.*, 1982; Adolf, 1987). Recombinant IFN-α2c was used as a standard. Whereas the sample as well as fractions of the peak eluted at low pH (Fig. 1, peak A) gave titration curves parallel with that of the standard, the flow-through and fractions of the first peak (B) gave curves with different slope; these samples could therefore not be quantified by the e.l.i.s.a. (Fig. 2) and were submitted to biological assays (Table 1). The low pH required for elution of the IFN in peak A as well as the results of the e.l.i.s.a. indicated that IFN-α2 was at least a major component of peak A. To ensure that all immunoreactive IFN-α had been bound by the antibody, the flow-through was passed over the column for a second time and bound protein was again eluted as described

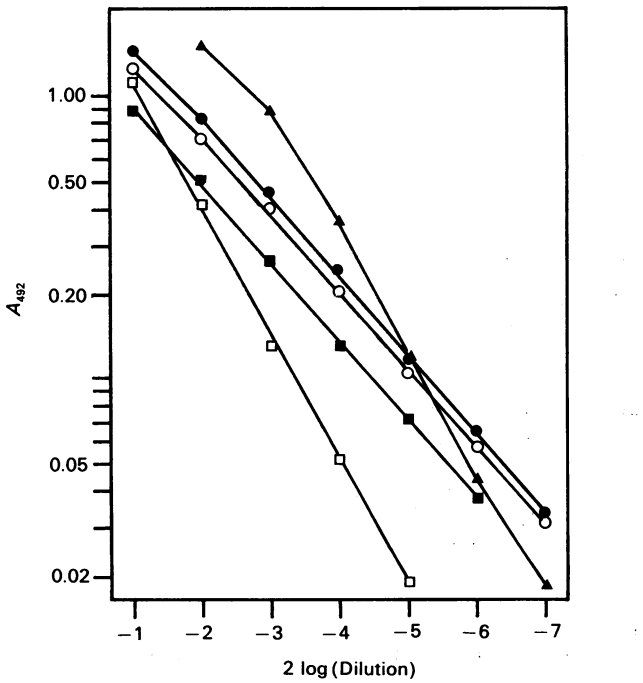

Fig. 2. E.l.i.s.a. for human IFN-α

IFN species resolved by MAb-affinity chromatography were analysed in a sandwich e.l.i.s.a. employing two MAbs. Appropriately diluted fractions were applied to the assay plate and diluted in serial 2-fold steps. Symbols: ●, a reference preparation of recombinant human IFN-α2c; ○, leucocyte IFN applied to the column; □, flowthrough; ■, a sample of eluate A; ▲, a sample of eluate B.

Table 1. Purification of natural IFN-α2

Stage	Volume (ml)	Protein (mg/ml)	10^{-6} × Antiviral activity (i.u./ml)*	10^{-6} × Antiviral activity (total i.u.)	Recovery (%)
P-IF†	550	1.7	2.8	1540	100
First cycle					
Flow-through	550	1.7	2.2	1216	79
Eluate A	18	0.08	9.6	172	11
Eluate B	17	0.05	4.3	73	4.8
Second cycle					
Eluate A	8	0.1	12	96	6.2
Eluate B	4	0.1	13	54	3.5

* Mean values for five independent bioassays.
† P-IF (partially purified human leucocyte IFN) after removal of IFN-ω1.

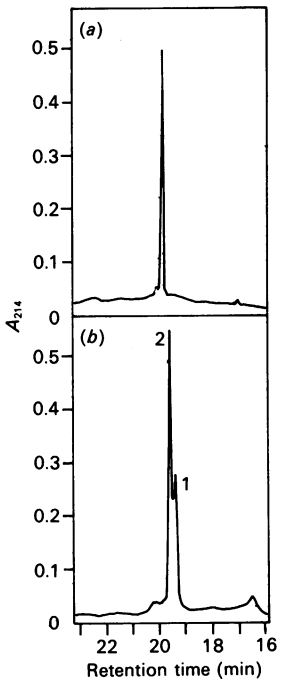

Fig. 3. Reverse-phase h.p.l.c. analysis of IFN-α2

(*a*) *E. coli*-derived recombinant human IFN-α2c; (*b*) natural IFN-α2 (eluate A of the affinity column).

above. The material eluted amounted to less than 10% of the IFN activity bound in the first run.

Column fractions representing peaks A and B respectively were independently pooled, neutralized, and each pool was subjected to a second cycle of affinity chromatography on the same column. In both cases more than 95% of the IFN activity was bound; elution was observed at the same position in the gradient as in the first cycle of chromatography (results not shown). The original sample of the leucocyte IFN, flow-through and pooled peak fractions of both chromatographic cycles were analysed for protein content by Coomassie Blue dye-binding assays, and for IFN activity in an antiviral bioassay. The results of one experiment are summarized in Table 1.

Identification of affinity-purified protein as IFN-α2

IFN-α purified by affinity chromatography was analysed first by reverse-phase h.p.l.c. Protein derived from eluate peak A showed two incompletely resolved peaks, 1 and 2, with a mass ratio of about 1:3 (Fig. 3*b*); peak 1 represents a more hydrophilic protein fraction. Both peak fractions were collected, rechromatographed, and subjected to *N*-terminal amino-acid-sequence analysis. The following sequence was obtained in both fractions (the cysteine residues given in parentheses were not identified, but are inferred on the basis of the conserved IFN sequences):

```
 1                5                10                15
(Cys)-Asp-Leu-Pro-Gln-Thr-His-Ser-Leu-Gly-Ser-Arg-Arg-Thr-Leu-
16               20                25                30
Met-Leu-Leu-Ala-Gln-Met-Arg-Arg-Ile-Ser-Leu-Phe-Ser-(Cys)-Leu-...
```

By comparison with published sequences, both proteins were identified as IFN-α2.

In both peak fractions (1 and 2) amino acid 23 was unequivocally identified as arginine, indicating that IFN-α2a, the IFN-α2 variant with lysine in position 23 (= LeIF A; Goeddel *et al.*, 1981) was not present in the leucocyte IFN preparation in detectable quantities (< 10% of total IFN-α2).

The specific antiviral activity of natural IFN-α2 with respect to the international reference preparation for IFN-α2, Gxa01-901-535, based on a determination of the protein content of the sample by its absorption at 214 nm (Adolf *et al.*, 1990), was determined as 1.5×10^8 i.u./mg (mean; two independent preparations; five bioassays for each preparation).

When the retention time of natural IFN-α2 on the reverse-phase column was compared with that of *E. coli*-derived IFN-α2c, it was apparent that the natural protein was eluted significantly earlier (Fig. 3). The increased hydrophilicity of the natural protein, as well as its heterogeneity, must be due to some post-translational modification. Further analysis was thus required to define the structural differences between the natural and *E. coli*-derived proteins (see below).

Reverse-phase h.p.l.c. of eluate peak B resulted in a more complicated pattern of five incompletely resolved peaks (results not shown). Sequence analysis revealed that all these peaks represented IFN-α species, but none of them consisted of IFN-α2.

Heterogeneity of natural IFN-α2 is due to *O*-linked glycosylation

H.p.l.c.-purified IFN-α2 was further analysed by SDS/PAGE after reduction with dithiothreitol (Fig. 4). Under the conditions chosen, *E. coli*-derived IFN-α2c showed an apparent molecular mass of 17500 Da (molecular mass calculated from the amino acid sequence 19287 Da). H.p.l.c. peak 1 (about 25% of the total protein) gave a single, relatively broad, and thus possibly heterogeneous, band, with an apparent molecular mass of 20000 Da, whereas the major peak 2 was further resolved into two major components (20000 and 19000 Da respectively), as well as a minor one (21000 Da). These differences in molecular mass from that of the recombinant protein, the size heterogeneity, as well as the increased hydrophilicity, strongly indicated that natural IFN-α2 is glycosylated. As there is no recognition site for *N*-glycosylation present in the IFN-α2 structure, *O*-glycosylation was considered.

The *O*-glycosidic bond of glycoproteins can be opened by β-elimination under alkaline conditions. When natural IFN-α was incubated in 0.1 M-NaOH for 20 h at room temperature, the heterogeneity was eliminated, and the apparent molecular mass was reduced; protein derived from both peaks (1 and 2) resulted in a single band with an apparent molecular mass indistinguishable from that of alkali-treated *E. coli*-derived IFN-α2c (Fig. 4). Protein shortened by ten amino acids at the *C*-terminal end as described by Levy *et al.* (1981), which would result in higher electrophoretic mobility, was not detected.

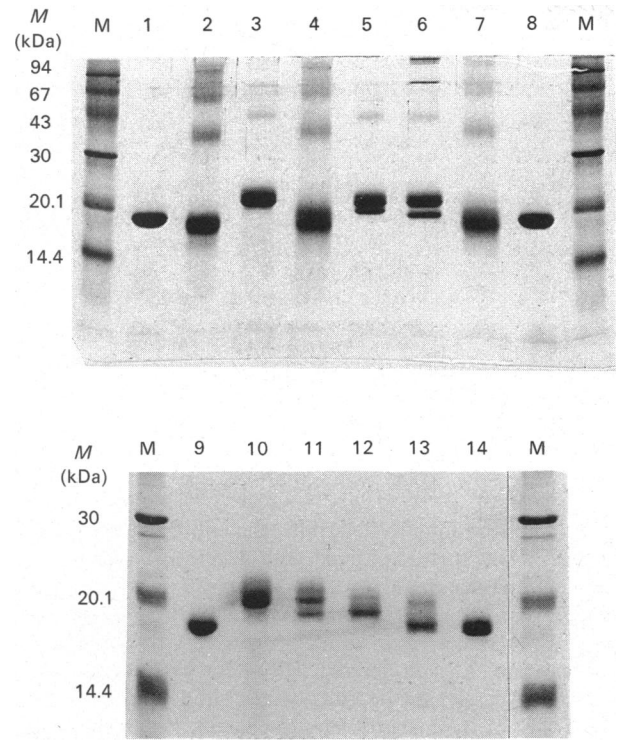

Fig. 4. SDS/PAGE of h.p.l.c.-purified IFN-α2

Proteins were reduced with dithiotreitol, separated on 15% -(w/v)-polyacrylamide gels and stained with Coomassie Blue. M, molecular-mass (*M*) markers. Lanes 1 and 2, *E. coli*-derived IFN-α2c: lane 1, untreated; lane 2, NaOH-treated. Lanes 3 and 4, natural IFN-α2 (h.p.l.c. peak 1): lane 3, untreated; lane 4, NaOH-treated. Lanes 5, 6 and 7 (natural IFN-α2; h.p.l.c. peak 2): lane 5, untreated; lane 6, treated with *O*-glycanase; lane 7, NaOH-treated. Lanes 8 and 9, *E. coli*-derived IFN-α2c, untreated. Lane 10, natural IFN-α2, h.p.l.c. peak 1, untreated. Lanes 11, 12 and 13, natural IFN-α2, h.p.l.c. peak 2; lane 11, untreated; lane 12, treated with neuraminidase; lane 13, treated with neuraminidase and *O*-glycanase. Lane 14, *E. coli*-derived IFN-α2c, untreated. Proteins with higher molecular mass most probably represent oligomeric forms of IFN-α2.

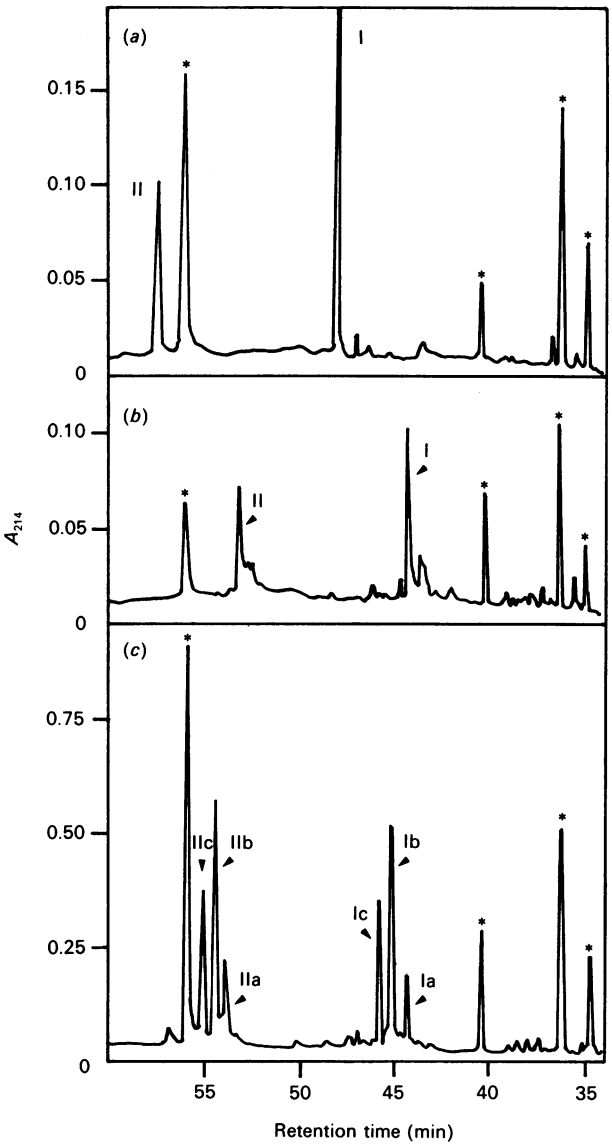

Fig. 5. Separation of tryptic peptides by reverse-phase h.p.l.c.

(*a*) *E. coli*-derived IFN-α2c; (*b*) natural IFN-α2, reverse-phase h.p.l.c. peak 1; (*c*) natural IFN-α2, reverse-phase h.p.l.c. peak 2 (see Fig. 3). Only the peptide map region containing the putative glycopeptides is shown. Stars indicate major peaks derived from unglycosylated peptides (retention time identical in all preparations). Peaks I and II represent peptides comprising amino acids 84–112 and 71–112 respectively (see also Table 2).

Carbohydrate analysis

Further evidence for *O*-glycosylation was obtained in experiments using neuraminidase and endo-α-*N*-acetylgalactosaminidase (*O*-glycanase; an enzyme that exclusively cleaves the peptide–carbohydrate bond of serine- or threonine-linked unsubstituted Gal-GalNAc; Fig. 4). Treatment with neuraminidase, or sequential treatment with both enzymes, had no detectable effect on the electrophoretic mobility of protein derived from peak 1 (results not shown). In contrast, neuraminidase treatment diminished or eliminated the 21 kDa and 20 kDa bands derived from peak 2, and simultaneously increased the 19 kDa component. Treatment with *O*-glycanase affected only the 19 kDa component, reducing its mass to about 17.5 kDa. Sequential treatment with both enzymes resulted in the disappearance of all three bands and in the appearance of a new band, again with a molecular mass of about 17.5 kDa. A small amount of 20 kDa protein, however, was unaffected by all enzyme reactions; this protein most probably represents a contamination of peak 2 with peak 1 protein, owing to incomplete resolution of the h.p.l.c. column (Fig. 3). These results suggest that natural IFN-α2 contains a core *O*-linked carbohydrate moiety, probably of the Galβ1-3GalNAc-(Ser/Thr) type (19 kDa component of peak 2, sensitive to *O*-glycanase), which, in a proportion of the molecules, is modified by one or two neuraminic acid residues (20 kDa and 21 kDa components respectively of peak 2; sensitive to neuraminidase) or by different, so far unknown, sugar residues (peak 1) that protect the peptide–carbohydrate bond from cleavage by *O*-glycanase. Further glycosylation isomers may be present as minor components, but are not detectable by SDS/PAGE (see below).

Additional information on the structure of the carbohydrate moieties was obtained by m.s. (plasma desorption, time-of-flight separation) of peptides generated by trypsin digestion and separated by reverse-phase h.p.l.c. Peptide maps obtained from both h.p.l.c. peaks (1 and 2) of natural IFN-α2 were compared with a corresponding map of *E. coli*-derived IFN-α2c; major differences were seen in map regions that, as determined by *N*-terminal sequence analysis, contain peptides comprising amino acid residues 84–112 and 71–112 (Fig. 5; peaks designated I and II respectively); compared with peptides derived from re-

Table 2. Molecular masses of some tryptic peptides of natural IFN-α2 determined by m.s.

Peak*	Peptide (amino acid position number range)	Molecular mass (a.m.u.)†			Proposed carbohydrate structure	
		Observed	Calculated (aglycone)	Difference	Structure	Mass (a.m.u.)‡
1/II	71–112	5485	4736	749	Gal-GalNAc-Gal-GlcNAc	752
2/Ib	84–112	3975	3304	671	NeuAc-Gal-GalNAc-	678
2/IIb	71–112	5448	4736	712	NeuAc-Gal-GalNAc-	678
2/IIc	71–112	5132	4736	396	Gal-GalNAc-	387

* Peak numbers according to Fig. 5.
† a.m.u., atomic mass units.
‡ Calculated mass including one Na^+ ion.

combinant IFN-α2c, the respective fragments of natural IFN-α2 were clearly more hydrophilic. Moreover, the map of h.p.l.c. peak 2 gave three major peaks for each peptide (Ia, Ib, Ic; IIa, IIb, IIc), in qualitative and quantitative accordance with the number and intensity of bands observed in SDS/PAGE. The molecular masses of three major peptides determined by mass spectrometry are given in Table 2. The observed differences between the masses calculated from the amino acid sequence and the experimentally determined values are in agreement with the proposed core carbohydrate structure, Gal-GalNAc, and its substitution by neuraminic acid in a proportion of the protein (peak 2). The presence of sialic acid was also confirmed in experiments using ion-exchange h.p.l.c. with pulsed amperometric detection (Dionex system) of monosaccharides derived from tryptic peptides. For this experiment, corresponding peptides were pooled (Ia + IIa, Ib + IIb and Ic + IIc respectively), sialic acids were liberated by acid hydrolysis and analysed. (α-)*N*-Acetylneuraminic acid was identified in peptides a and b, but was absent from peptides c, the most hydrophobic ones of the triplets. *N*-Glycolylneuraminic acid was not detected (results not shown).

The carbohydrate structure, which was insensitive to neuraminidase and *O*-glycanase (peak 1; about 25% of the total protein), was also analysed by m.s. of a major tryptic peptide (1/II; Table 2; Fig. 5). The observed mass indicates the presence of a tetrasaccharide, probably consisting of an *N*-acetyl-lactosamine (Gal-GlcNAc) disaccharide bound to the GalNAc molecule of the core structure (Kamerling *et al.*, 1990); however, other interpretations are possible. Analysis of the neutral carbohydrates of total peak 1 protein using the Dionex h.p.l.c. system after acid hydrolysis confirmed the presence of Gal, GalNAc and GlcNAc, and also indicated the presence of fucose as well as of an unidentified monosaccharide that was eluted at the position of the glucose marker; glucose, however, has never been found in *O*-linked glycans and thus is an unlikely component. Fucose or other monosaccharides may thus be present instead of *N*-acetyl-lactosamine in minor glycoforms of peak 1 protein; these protein species may thus account for the relatively broad and probably heterogeneous band observed in SDS/PAGE and for minor peaks of the peptide map that are eluted immediately before the main glycopeptide (Fig. 5*b*). The low amounts of these minor glycoforms, which together represent less then 10% of the total protein present in our IFN-α2 preparation, precluded further analysis.

Identification of the *O*-glycosylation site

To identify the amino acid residue(s) that is (are) glycosylated, relevant tryptic peptides (Fig. 5) were sequenced. As these peptides were too long to determine the entire sequence, part of the material was further cleaved with *Staphylococcus aureus* V8 proteinase and separated by reverse-phase h.p.l.c. (results not shown). Again, peptides derived from natural IFN-α2 that showed different retention times compared with corresponding peptides from *E. coli*-derived IFN-α2c were collected and sequenced. All relevant peptides were found to contain amino acids 97–112 of the mature IFN-α2 protein. Whereas all amino acid derivatives from *E. coli*-derived peptides were readily identified, the derivative of residue 106 (threonine) was undetectable in both h.p.l.c. peaks 1 and 2 of the natural protein. This result is consistent with glycosylation at this residue, as the corresponding phenyl isothiocyanate derivative cannot be extracted by the butyl chloride solvent used in the sequencing cycle.

DISCUSSION

We have utilized the unique specificity of a monoclonal antibody for the isolation of natural IFN-α2 from human leucocyte IFN. Protein of at least 95% purity was obtained after two sequential cycles of immunoaffinity chromatography. Bioassays confirmed the expected specific antiviral activity (1.5×10^8 i.u./mg). IFN-α2 therefore contributes about 15% of the total IFN activity of leucocyte IFN, taking losses incurred during purification into account. As suggested previously on the basis of mRNA frequencies (Goeddel *et al.*, 1981; Hiscott *et al.*, 1984), IFN-α2 thus represents a major constituent of this IFN preparation.

Amino-acid-sequence analysis indirectly showed that cysteine represents the *N*-terminus of natural IFN-α2, in accordance with earlier studies that used mixtures of IFN-α species for analysis (see, e.g., Allen & Fantes, 1980; Levy *et al.*, 1980; Zoon *et al.*, 1980). In contrast, natural IFN-ω1, isolated from the same batch of leucocyte IFN, was previously shown to be heterogeneous at the *N*-terminus; owing to variable cleavage of the precursor during biosynthesis, about 60% of IFN-ω1 molecules carry two additional amino acids (Adolf *et al.*, 1990). No evidence for blocked *N*-termini, as described by Rubinstein *et al.* (1981), was obtained for IFN-α2; further, protein clipped at the *C*-terminus (Levy *et al.* 1981) was not detectable.

Three variants of IFN-α2, differing in their amino acids at positions 23 and 34, are known: IFN-α2a (Lys-23, His-34; originally designated 'Le IF A'; Goeddel *et al.*, 1981), IFN-α2b (Arg-23, His-34; the prototype IFN-α2 described by Streuli *et al.*, 1980) and IFN-α2c (Arg-23, Arg-34; formerly IFN-α2"Arg"; Dworkin-Rastl *et al.*, 1982; Bodo & Maurer-Fogy, 1986). At amino acid position 23 of natural IFN-α2, only

arginine, but not lysine, could be identified. As a conservative estimate, at least 10 mol % of lysine would have been detected by our analytical procedure. It thus seems that IFN-α2a is at best a minor component of natural IFN-α2 and may be totally absent from this preparation. Amino acids beyond residue 30 of natural IFN-α2 could not be clearly identified during *N*-terminal sequencing, and we were thus not able to conclude whether IFN-α2b and/or IFN-α2c are represented in the preparation of leucocyte IFN we have analysed. cDNA coding for IFN-α2b was first derived from leucocytes isolated from the same donor population (blood donors at the Finnish Red Cross Blood Transfusion Service) that also provided leucocytes for the IFN preparation used as the starting material for the present study. IFN-α2a cDNA was isolated from a library prepared from Sendai-virus-induced KG-1 cells, a myeloid cell line derived from a patient with chronic myeloid leukaemia (Koeffler & Golde, 1978; Goeddel *et al.*, 1981); IFN-α2c cDNA was cloned from the African-Burkitt's-lymphoma cell line, Namalwa (Dworkin-Rastl *et al.*, 1982). IFN-α2a, -2b and -2c are often regarded as allelic variants, although evidence that at least IFN-α2a and -2b may be derived from different genes was presented (Pestka *et al.*, 1986). If this is correct, then IFN-α2a and IFN-α2c may represent allelic variants, and IFN-α2b may be the product of a separate gene; alternatively, these three proteins may be derived from allelic forms of a single gene. Our results thus can be interpreted to show that the IFN-α2a allele is rare in, or even absent from, at least the Finnish population, or that, for unknown reasons, the allele is not expressed. With the help of our EBI-10 MAb column it is feasible to determine whether the representation of IFN-α2 variant proteins is different in preparations of human leucocyte IFN derived from different donor populations. Our preliminary results indicate that virus-induced Namalwa cells, a source of 'lymphoblastoid' IFN (Allen & Fantes, 1982; Bodo, 1981), secrete IFN-α2b and IFN-α2c, but also do not produce detectable IFN-α2a (I. Maurer-Fogy & G. Bodo, unpublished work).

A number of earlier studies have attributed the size and charge heterogeneity of human leucocyte IFN observed in SDS/PAGE and isoelectric focusing respectively to heterogeneous glycosylation of a single polypeptide chain (reviewed by Stewart, 1979). Subsequently, cloning of IFN cDNA and genomic DNA revealed the unexpected multitude of IFN-α genes. With one exception, the corresponding proteins do not contain recognition sites (Asn-Xaa-Ser/Thr) for *N*-glycosylation (Henco *et al.*, 1985); only IFN-α14 (Le IF H, IFN-αN) contains two of these sequences. A related IFN protein, IFN-ω1, contains a single *N*-glycosylation site. We have previously shown that IFN-ω1 is a major component of human leucocyte IFN and is, indeed, *N*-glycosylated (Adolf *et al.*, 1990). Recently, we have obtained evidence that IFN-α14 is also present in human leucocyte IFN, although only in small quantity, and is also glycosylated (G. R. Adolf, I. Kalsner, H. Ahorn, I. Maurer-Fogy & K. Cantell, unpublished work). IFN-α14 and/or IFN-ω1 may thus account for the glycosylated IFN species with unexpectedly high molecular mass that were observed in preparations of human 'lymphoblastoid' (Namalwa) IFN and leucocyte IFN produced by leukaemic cells (Allen, 1982; Labdon *et al.*, 1984). Although it was suggested that *O*-glycosylated species of IFN-α may also exist (Labdon *et al.*, 1984; Kojima *et al.*, 1989; Zoon *et al.*, 1989), this has not been demonstrated directly.

Analysis of natural IFN-α2 by reverse-phase h.p.l.c. showed that the natural protein can be resolved into two components, both more hydrophilic than *E. coli*-derived IFN-α2. SDS/PAGE revealed that the protein is also heterogeneous in molecular mass, resulting in three bands, all of them with lower electrophoretic mobility than the equivalent *E. coli*-derived protein. The most straightforward explanation for the results of both analytical procedures is heterogeneous glycosylation of the natural protein. As IFN-α2 does not contain a recognition site for *N*-linked glycosylation, *O*-linked carbohydrates must be involved. This hypothesis was confirmed by cleavage of the putative peptide–carbohydrate bond with alkali; the resulting protein was homogeneous and showed the same molecular mass as the recombinant protein. Further comparison of natural and recombinant proteins after proteolytic cleavage, followed by separation and analysis of the resulting fragments, allowed us to define a candidate glycopeptide; sequence analysis of this peptide identified Thr-106 as the *O*-glycosylation site. A comparison of the amino acid sequences of all published IFN-α species (Henco *et al.*, 1985) revealed that this threonine residue is unique to IFN-α2; glycine, isoleucine or glutamic acid are present at the corresponding position (107) in all other proteins. This might indicate that IFN-α2 is the only *O*-glycosylated species of IFN-α, in agreement with previous reports that only two or three of the IFN-α species purified from human cells contain carbohydrates (Labdon *et al.*, 1984; Kojima *et al.*, 1989; Zoon *et al.* 1989). On the other hand, Thr-108 (109 in other IFN-α species) is strongly conserved; however, at least in natural IFN-α2, this residue is not glycosylated. The recognition sequence for *O*-glycosylation is not well defined; proline, serine and threonine residues are often observed in the vicinity of the *O*-glycosylation site. This holds true also for IFN-α2 (Thr-106, Thr-108 and Pro-109). It is not clear why Thr-106, but not Thr-108, is modified.

The carbohydrate component of natural IFN-α2 was analysed using a variety of techniques. Taken together, these experiments clearly demonstrate that natural IFN-α2 contains at least four major differently glycosylated protein species. For the majority of molecules (h.p.l.c. peak 2, containing about 75 % of the total protein), the most likely structure of the carbohydrate chain could be established on the basis of experiments using SDS/PAGE, glycosidases, tryptic peptide maps, m.s. and monosaccharide analysis. We propose that all molecules contain the core disaccharide, Gal-GalNAc, which is either unsubstituted or modified in a proportion of molecules by one or two (α-)*N*-acetylneuraminic acid residues. The carbohydrate composition of proteins represented in the minor h.p.l.c. peak (1) is less clear, but most probably is also heterogeneous. Its carbohydrate moiety is insensitive to neuraminidase and *O*-glycanase, and mass spectrometry of the only major tryptic glycopeptide indicated the presence of a tetrasaccharide. In accordance with these results, and also with published *O*-glycan structures (Kamerling *et al.*, 1990), we propose that the major protein species in peak 1 contains the same core carbohydrate as described above, Gal-GalNAc, substituted with the disaccharide Gal-GlcNAc. Although we do not have independent proof for this structure, our hypothesis is consistent with results of monosaccharide analysis of the total peak 1 protein; this method also revealed that, in addition to Gal, GalNAc and GlcNAc, other neutral monosaccharides, including fucose, are present. Minor components of peak 1 may thus show further diversity in carbohydrate composition; their presence may explain the broad band observed in SDS/PAGE and minor peaks in the peptide map. In summary, we have suggested, on the basis of a variety of experimental approaches, carbohydrate structure assignments for IFN-α2 glycoforms that together account for at least 90 % of the naturally occurring protein. A complete elucidation of the carbohydrate structures was beyond the scope of the present work and will require larger quantities of protein.

To determine whether IFN-α2 produced by the Burkitt's-lymphoma cell line Namalwa is also *O*-glycosylated, experiments similar to those described above were carried out. From an IFN preparation produced and partially purified as described by Bodo

(1981), IFN-α2 was isolated by MAb-affinity chromatography and reverse-phase h.p.l.c. Preliminary results indicate that Namalwa-derived IFN-α2 is more hydrophilic and has a higher molecular mass than *E. coli*-derived IFN-α2 and thus most likely is also glycosylated; again, Thr-106 could be identified as the glycosylation site. However, only a single peak was observed upon reverse-phase h.p.l.c., and the pattern of glycosylation therefore must be different (I. Maurer-Fogy & G. Bodo, unpublished work).

The significance of our findings is not clear at present. Preparations of IFN-α2 produced in *E. coli* and thus devoid of *O*-glycosylation have been registered as drugs in many countries. Considering results obtained with IFN-β, IFN-ω1 and IFN-γ, it is unlikely that glycosylation affects the biological activity of IFN-α2 *in vitro* (see, e.g., Bocci, 1983); our antiviral bioassays have not shown a significant difference in the specific antiviral activities of the natural and recombinant IFN-α2 proteins. Senda *et al.* (1990) have recently reported on the three-dimensional structure of murine IFN-β determined by X-ray crystallography; it is thought that all class I IFNs share this basic framework. In this structure, residue 106 would be located in loop CD on the surface of the molecule, consistent with a glycosylation site. With respect to the putative receptor interaction site of IFN, this loop is situated on the opposite side of the molecule, and thus glycosylation is unlikely to affect binding to the receptor. However, the pharmacokinetics of glycosylated and unglycosylated IFNα2 may well be different, and the stability of the protein may be influenced by glycosylation, as is the case for human granulocyte colony-stimulating factor (Oh-eda *et al.*, 1990). Further, the immunogenicity of therapeutically applied *E. coli*-derived IFN-α2 might be affected by the lack of glycosylation. Gribben *et al.* (1990) have recently reported that four out of 16 patients receiving recombinant human granulocyte–macrophage colony-stimulating factor produced in yeast developed antibodies to this protein; these antibodies were found to react with epitopes that in endogeneous granulocyte–macrophage colony-stimulating factor are protected by *O*-linked glycosylation, but which are exposed in the recombinant factor. Induction of antibodies to recombinant *E. coli*-derived IFN-α2 after prolonged treatment of patients has been described (Figlin & Itri, 1988), and it has been speculated that natural IFN-α may be less immunogenic than the recombinant IFN-α2 proteins (Galton *et al.*, 1989). This topic is highly controversial, and the incidence of antibody induction certainly depends on a variety of factors; however, the difference in glycosylation should at least be considered. It will now be possible to compare the reactivity of relevant serum samples with recombinant and natural IFN-α2.

We thank T. Gramanitsch, J. Hoffmann, D. Jobstmann, I. Schweiger, and E. Traxler for excellent technical assistance, and Dr. G. Bodo for providing recombinant IFN-α2c. The antiviral bioassays were kindly performed by Dr. B. Frühbeis, M. Karner, and S. Korica. We are grateful to Dr. M. Schlüter, Thomae G.m.b.H, for allowing us to use his Dionex h.p.l.c. system.

REFERENCES

Adolf, G. R. (1987) J. Gen. Virol. **68**, 1669–1676
Adolf, G. R. (1990) Virology **175**, 410–417
Adolf, G. R., Bodo, G. & Swetly, P. (1982) J. Cell. Physiol. Suppl. **2**, 61–68
Adolf, G. R., Maurer-Fogy, I., Kalsner, I. & Cantell, K. (1990) J. Biol. Chem. **265**, 9290–9295
Allen, G. (1982) Biochem. J. **207**, 397–408
Allen, G. & Fantes, K. H. (1980) Nature (London) **287**, 408–411
Bocci, V. (1983) Trends. Biochem. Sci. **8**, 432–434
Bodo, G. (1981) Methods Enzymol. **78**, 69–75
Bodo, G. & Maurer-Fogy, I. (1986) in The Biology of the Interferon System 1985 (Stewart, W. E., II & Schellekens, H., eds.), pp. 59–64, Elsevier Scientific Publishing Co., Amsterdam
Cantell, K., Hirvonen, S., Kauppinen, H. L. & Myllylä, G. (1981*a*) Methods Enzymol. **78**, 29–38
Cantell, K., Hirvonen, S. & Koistinen, V. (1981*b*) Methods Enzymol. **78**, 499–512
Dworkin-Rastl, E., Dworkin, M. B. & Swetly, P. (1982) J. Interferon Res. **2**, 575–585
Figlin, R. A. & Itri, L. M. (1988) Semin. Haematol. **25**, suppl. 3, 9–15
Fish, E. N., Banerjee, K. & Stebbing, N. (1983) Biochem. Biophys. Res. Commun. **112**, 537–546
Galton, J. E., Bedford, P., Scott, J. E, Brand, C. M. & Nethersell, A. B. W. (1989) Lancet **ii**, 572–573
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 (London) **290**, 20–26
Gribben, J. G., Devereux, S., Thomas, N. S. B., Keim, M., Jones, H. M., Goldstone, A. H. & Linch, D. C. (1990) Lancet **335**, 434–437
Henco, K., Brösius, F. J., Fujisawa, A., Fujisawa, J.-I., Haynes, J. R., Hochstadt, J., Kovacic, T., Pasek, M., Schambeck, A., Schmid, J., Todokoro, K., Wälchli, M., Nagata, S. & Weissmann, C. (1985) J. Mol. Biol. **185**, 227–260
Hiscott, J., Cantell, K. & Weissmann, C. (1984) Nucleic Acids Res. **12**, 3727–3746
Hobbs, D. S. & Pestka, S. (1982) J. Biol. Chem. **252**, 4071–4076
Kamerling, J. P., Hard, K. & Vliegenthart, J. F. G. (1990) in From Clone to Clinic (Crommelin, D. J. A. & Schellekens, H., eds.), pp. 295–304, Kluwer, Amsterdam
Koeffler, H. P. & Golde, D. W. (1978) Science **200**, 1153–1154
Kojima, S., Futatsuki, T., Hatano, Y. & Koide, T. (1989) J. Interferon Res. **9**, suppl. 2, S178
Labdon, J. E., Gibson, K. D., Sun, S. & Pestka, S. (1984) Arch. Biochem. Biophys. **232**, 422–426
Levy, W. P., Shively, J., Rubinstein, M., DelValle, U. & Pestka, S. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 5102–5104
Levy, W. P., Rubinstein, M., Shively, J., DelValle, U., Lai, C. Y., Moschera, J., Brink, L., Gerber, L., Stein, S. & Pestka, S. (1981) Proc. Natl. Acad. Sci. U.S.A. **78**, 6186–6190
Oh-eda, M., Hasegawa, M., Hattori, K., Kuboniwa, H., Kojima, T., Orita, T., Tomonou, K., Yamazaki, T. & Ochi, N. (1990) J. Biol. Chem. **265**, 11432–11435
Ortaldo, J. R., Herberman, R. B., Harvey, C., Osheroff, P., Pan, Y. C. E., Kelder, B. & Pestka, S. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 4926–4929
Pestka, S., Hotta, K., Monahan, J., Collier, K. J., Rashidbaigi, A., Langer, J. & Kung, H. F. (1986) in The Biology of the Interferon System 1985 (Stewart, W. E., II & Schellekens, H., eds.), pp. 3–11, Elsevier Scientific Publishing Co., Amsterdam
Rubinstein, M., Levy, W. P., Moschera, J. A., Lai, C. Y., Hershberg, R. D., Bartlett, R. T. & Pestka, S. (1981) Arch. Biochem. Biophys. **210**, 307–318
Senda, T., Matsuda, S., Kurihara, H., Nakamura, K. T., Kawano, G., Shimizu, H., Mizuno, H. & Mitsui, Y. (1990) Proc. Jpn. Acad. **66**, Ser. B, 77–80
Stewart, W. E., II (1979) The Interferon System, pp. 179–182, Springer-Verlag, New York and Vienna
Streuli, M., Nagata, S. & Weissmann, C. (1980) Science **209**, 1343–1347
Weck, P. K., Apperson, S., May, L. & Stebbing, N. (1981) J. Gen. Virol. **57**, 233–237
Zoon, K. C., Smith, M. E., Bridgen, P. J., Anfinsen, C. B., Hunkapiller, M. W. & Hood, L. E. (1980) Science **207**, 527–528
Zoon, K., Miller, D., Bekisz, J. & Zur Nedden, D. (1989) J. Interferon Res. **9**, suppl. 2, S184

Received 25 September 1990/6 December 1990; accepted 4 January 1991