Structure and properties of human interferon- α from Namalwa lymphoblastoid cells

Geoffrey ALLEN

Department of Molecular Biology, The Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, U.K.

(Received 14 April 1982/Accepted 9 August 1982)

The chromatographic properties of human interferon- α from Namalwa lymphoblastoid cells on Sephadex G-75 are described. The interferons are separated into two groups of four, with apparent molecular weights 19 050 and 22 000. Some of the latter form dimers at high concentrations. Fractions containing interferon were studied by polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate. Seven of the components had apparent molecular weights in this system, after reduction, of between 18 400 and 20 900: one component is probably glycosylated and has an apparent molecular weight of 27 500. Amino acid sequences of peptides derived from interferon mixtures were determined and are related to published sequences deduced from the nucleotide sequences of cloned complementary DNA coding for interferon- α . The results show that the major interferon- α species isolated from Namalwa cells do not undergo C-terminal processing. Amino acid analyses of peptides are presented in Supplementary Publication SUP 50117 (28 pages), which has been deposited with the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 7BQ, U.K., from whom copies can be obtained on the terms indicated in *Biochem. J.* (1981) 193, 5.

Human leucocyte-type interferon, IFN- α (for nomenclature see Stewart *et al.*, 1980), is a mixture of several molecular species, as has been shown by various techniques, including SDS/polyacrylamide-gel electrophoresis (Stewart & Desmyter, 1975; Desmyter & Stewart, 1976; Havell *et al.*, 1977; Bridgen *et al.*, 1977; Chadha *et al.*, 1978; Grob & Chadha, 1979; Berg & Heron, 1980; Allen & Fantes, 1980), isoelectric focusing (Fantes, 1970; Mogensen *et al.*, 1974; Bose *et al.*, 1976; Törmä & Paucker, 1976; Havell *et al.*, 1977; Stewart *et al.*, 1977; Salit & Ogburn, 1980), high-pressure liquid chromatography (Rubinstein *et al.*, 1979), affinity chromatography (Chen *et al.*, 1976; Grob & Chadha, 1979; Thang *et al.*, 1979), gel chromatography (Chadha *et al.*, 1978) and hydroxyapatite chromatography (Paucker *et al.*, 1977; Törmä & Paucker, 1976). Until recently it was generally assumed that human IFN- α consisted of a population of molecules with a single amino acid

sequence (or perhaps two sequences) subjected to various degrees or types of modification, either glycosylation (Bose *et al.*, 1976; Bridgen *et al.*, 1977) or degradative alteration (Thang *et al.*, 1979; Chadha *et al.*, 1978). However, it was found that at least five different amino acid sequences were present in Namalwa IFN- α (Allen & Fantes, 1980), and that complementary DNA prepared from mRNA coding for IFN- α was a mixture containing at least ten different sequences (Nagata *et al.*, 1980, 1981; Goeddel *et al.*, 1981). I present below details of investigations into the structure of human IFN- α from Namalwa lymphoblastoid cells. Tables of amino acid analyses of peptides are contained in Supplementary Publication SUP 50117.

Methods

Interferon samples

Human IFN- α was produced (Finter, 1982) from cells of the lymphoblastoid line Namalwa (Strander *et al.*, 1975) cultured in RPMI 1640 medium (Wellcome Media Production Laboratory, Beckenham, Kent, U.K.) (Moore *et al.*, 1967) containing 7% (v/v) calf serum. The interferon was

Abbreviations used: IFN- α , human leucocyte-type interferon; IFN- α A and IFN- α B, interferon mixtures constituting the first and second peaks respectively eluted from a Sephadex G-75 column; SDS, sodium dodecyl sulphate.

purified by a multistep method (Fantes *et al.*, 1980). Antiviral activity was measured by the dye-uptake method of Finter (1969) with V3 monkey cells (Christofinis, 1970) challenged with Semliki Forest virus. The starting material used in the present work was provided by Dr. K. H. Fantes, and had specific activity ranging from 25×10^6 to 120×10^6 i.u./mg of protein. Some batches were sufficiently pure to yield essentially pure interferon when chromatographed on Sephadex G-75 (Fantes & Allen, 1981). Some batches were contaminated by other proteins and further purification was required; this was achieved by gel filtration on Sephadex G-75 at pH 8, followed by gel filtration on Sephadex G-100 (superfine grade) in 5% (v/v) formic acid. If required, the interferon protein was finally purified on Sephadex G-100 in 8M-urea and 5% (v/v) formic acid. Alternatively, final purification was achieved after reduction and *S*-carboxymethylation or after performic acid oxidation. The interferon used in the present work for primary structural investigations contained at least seven components, as shown by separation by a combination of gel filtration and SDS/polyacrylamide-gel electrophoresis. The total amount of purified interferon used was about 4 mg. Primary structural studies were performed on two groups of components, labelled A and B, which constituted the two peaks obtained after Sephadex G-75 chromatography (Fig. 2).

SDS/polyacrylamide-gel electrophoresis

Interferon samples were studied by SDS/polyacrylamide-gel electrophoresis with the Tris/glycine system of Laemmli (1970), with a 12.5% or 15% acrylamide separating gel 9–10 cm long, and a 5% acrylamide stacking gel 1.5–2.0 cm long, in 1 mm-thick slabs. A 22 cm-long exponential gradient gel (Knight, 1975, 1976) was also used in some experiments. Samples (2–10 μ g) were generally reduced at 100°C for 2 min in 0.1 M-dithiothreitol in sample buffer (Laemmli, 1970) before application to the sample slot, but some samples were studied without reduction. After electrophoresis, the gels were fixed in 15% (w/v) trichloroacetic acid for 15 min, stained in Coomassie Brilliant Blue R-250 at 60°C for 30 min or at 20°C overnight, and destained in several changes of 7% (v/v) acetic acid/20% (v/v) methanol. For the determination of antiviral activity in protein bands, the gels were not fixed, but were stained briefly (30 min at 20°C) in Coomassie Blue and destained (90 min) just sufficiently to reveal the protein bands. The bands were excised and rinsed with water (2 ml) for 2 min. For reduced samples, the gel slices were shaken at 37°C for 16 h with 0.5 ml of 8M-urea/21 mM-2-mercaptoethanol/50 mM-NaCl/10 mM-sodium phosphate buffer, pH 7.2, containing 0.1% (w/v) SDS. For unreduced samples, the gel slices were shaken

with 0.5 ml of 0.1 M-NaCl/20 mM-sodium phosphate buffer, pH 7.2, containing 0.1% SDS. The extracts were added to 2.0 ml of 1% (w/v) bovine serum albumin in 0.1 M-NaCl/20 mM-sodium phosphate buffer, pH 7.2, and the solutions were dialysed against three changes of 50 mM-NaCl/10 mM-sodium phosphate buffer, pH 7.2, over 24 h before the antiviral assay.

Reduction and carboxymethylation

Incompletely purified IFN- α A (2.5 mg) was dissolved in 0.65 ml of 6M-guanidinium chloride/0.3 mM-EDTA/0.15 M-Tris/HCl buffer, pH 8.3. The protein was reduced with dithiothreitol (1.5 mM) at 37°C under N₂ for 1 h. Sodium iodo[¹⁴C]acetate solution (Amersham International, Amersham, Bucks., U.K.; diluted to 2150 c.p.m./nmol) (50 μ l of 50 mM solution) was added, and the solution was incubated in the dark under N₂ at 37°C for 80 min. 2-Mercaptoethanol (10 μ l) was added. The solution was applied to a column (10 mm $\times$ 525 mm) of Sephadex G-100 (superfine grade) in 5% (v/v) formic acid at 20°C and eluted at 3 ml/h. Fractions were assayed by SDS/polyacrylamide-gel electrophoresis, and pure *S*-carboxymethylated IFN- α A was dried down from appropriate fractions.

IFN- α B (1.1 mg) was similarly reduced and carboxymethylated. The protein was desalted on a column (6 mm $\times$ 60 mm) of Sephadex G-25 (fine grade) in 50 mM-NH₄HCO₃/0.01% (v/v) thiodiglycol.

Performic acid oxidation

Performic acid oxidation of interferon samples (5 mg/ml) was performed at 0°C as described by Hirs (1967).

Amino acid analyses

Amino acid analyses were performed on a Rank-Hilger Chromaspek instrument with a fluorimeter attachment for detection with the *o*-phthalaldehyde/2-mercaptoethanol reagent (Roth & Hampai, 1973). Interferon samples were hydrolysed in 6M-HCl/0.1% (w/v) phenol *in vacuo* at 110°C for 24 h and 72 h. Peptides were hydrolysed only for 20–24 h. Tryptophan was estimated in interferon spectrophotometrically (Edelhoch, 1967) and in peptides by the Ehrlich test or by fluorescence after performic acid oxidation. *S*-[¹⁴C]Carboxymethylcysteine was determined by liquid-scintillation counting of radioactivity with Fiso fluor 1 liquid scintillant (Fisons) and a Packard Tri-Carb liquid-scintillation counter. Approximate compositions of short peptides were determined by dansylation of the peptide hydrolysate (0.1–0.5 nmol) and identification of dansyl-amino acids by t.l.c.

Proteinase digestions

(1) Performic acid-oxidized IFN- α A (0.16 mg) was dissolved in 0.2 ml of 50 mM-NH₄HCO₃/0.01% thiodiglycol, and was digested with 3.0 μ g of trypsin [1-chloro-4-phenyl-3-tosylamidobutan-2-one ('TPCK')-treated; Worthington Biochemical Corp., Freehold, NJ, U.S.A.] at 37°C for 70 min. The digest was freeze-dried, and peptides were separated by thin-layer peptide 'mapping'. Performic acid-oxidized IFN- α B (0.16 mg) was similarly digested.

(2) Reduced S-[¹⁴C]carboxymethylated IFN- α A (0.92 mg) in 0.2 ml of 8 M-urea/0.5 M-NH₄HCO₃ was digested with occasional vortex-mixing at 20°C for 2 h with trypsin (18 μ g), added rapidly in 1.0 ml of water. A further 9 μ g of trypsin in 9 μ l of water was added and digestion continued for 1 h.

(3) Performic acid-oxidized IFN- α B (0.9 mg) was dissolved in 8 M urea/50 mM-NH₄HCO₃ (40 μ l). The solution was diluted with 50 mM-NH₄HCO₃ (0.8 ml), and trypsin (10 μ g) was added. The suspension was incubated at 37°C with occasional swirling and after 1 h it became clear. A further 10 μ g of trypsin was added and digestion was continued for 2 h at 37°C. The digest was applied directly to a Sephadex G-50 column for separation of peptides.

(4) Performic acid-oxidized IFN- α A (0.3 mg) was dissolved in 80 μ l of 8 M-urea/30 mM-NH₄HCO₃. Then 50 mM-NH₄HCO₃ (0.2 ml) was added, followed by *Staphylococcus aureus* (V8) proteinase (5 μ g; Miles Laboratories, Stoke Poges, Slough, Berks., U.K.) in 5 μ l of water. After 1 h at 37°C a further 5 μ g of proteinase was added and digestion was continued for a further 2 h. The digest was chromatographed on a column of Sephadex G-50 (superfine grade), and peptides eluted close to the void volume were further digested with 10 μ g of α -chymotrypsin [7-amino-1-chloro-3-tosylamidoheptan-2-one ('TLCK')-treated; Worthington Biochemical Corp.] in 50 mM-NH₄HCO₃ for 2 h at 37°C.

(5) Reduced S-[¹⁴C]carboxymethylated IFN- α B (1.0 mg) was digested with 20 μ g of α -chymotrypsin for 2 h at 37°C in 50 mM-NH₄HCO₃/0.01% thiodiglycol (1.0 ml).

Isolation of peptides

Peptides from digests (1), and the chymotryptic digest of large staphylococcal-proteinase-digest peptides (4), were isolated by thin-layer peptide 'mapping' (Heiland *et al.*, 1976), with electrophoresis at pH 4.4 followed by chromatography in butan-1-ol/acetic acid/pyridine/water (15:3:10:12, by vol.) on 20 cm $\times$ 20 cm plastic-backed cellulose MN 300 layers (Machery-Nagel, Düren, W. Germany). For analytical purposes, peptides were detected with the ninhydrin/Cd stain, the phenanthraquinone stain (for arginine residues) and by radioautography for

S-[¹⁴C]carboxymethylcysteine residues, whereas for preparative purposes dilute fluorescamine (0.001%) in 1% (v/v) pyridine in acetone was used. Full details of these methods have been given elsewhere (Allen, 1981). Peptide spots were scraped off and eluted with 0.1 M-NH₃ and 20% (v/v) pyridine.

Larger quantities of digests were initially separated on columns (1 cm $\times$ 195 cm) of Sephadex G-50 (superfine grade) in 50 mM-NH₄HCO₃/0.01% thiodiglycol. Peptides were detected by absorbance at 220 nm. Fractions of volume about 1.5 ml were collected at 5 ml/h, and were assayed by liquid-scintillation counting of radioactivity. Fractions comprising peptide peaks were pooled and freeze-dried. Further separation was by thin-layer peptide 'mapping'.

Determination of peptide sequence

Peptides were sequenced manually, by using either the micro dansyl-Edman method (Bruton & Hartley, 1970) or the dimethylaminoazobenzene isothiocyanate method (Chang *et al.*, 1978) with slight modifications described elsewhere (Allen, 1981). Some peptides were sequenced by both methods. S-[¹⁴C]Carboxymethylcysteine residues were identified by release of radioactivity from the peptides (liquid-scintillation counting), and by t.l.c. with the dimethylaminoazobenzene isothiocyanate method. N-Terminal sequences of IFN- α A and IFN- α B (4 nmol) were also determined by the dimethylaminoazobenzene isothiocyanate method.

Results

The chromatography of purified Namalwa interferon (64 μ g) on a column of Sephadex G-75 is shown in Fig. 1. Blue Dextran was added as an internal standard; no absorbance was detected at the void volume if Blue Dextran was omitted (cf. Fig. 2). Both 206 nm-absorbing material and the interferon activity were eluted as a double peak. The specific activity of interferon was calculated to be $1.8 \pm 0.5 \times 10^8$ i.u./mg (Fantes & Allen, 1981). The column was calibrated with bovine serum albumin (M_r 68 000), ovalbumin (M_r 44 000), soya-bean trypsin inhibitor (M_r 20 100) and myoglobin (M_r 16 890) (all from Sigma Chemical Co., Poole, Dorset, U.K.). The two peaks, IFN- α A and IFN- α B, had apparent M_r 22 000 and 19 050 respectively.

When a concentrated sample (0.8 mg in 2.0 ml) of the same preparation of interferon was chromatographed under identical conditions, the first peak was broadened and shifted to higher apparent molecular weight (Fig. 2). The scatter in the antiviral assay results reflects the imprecision of the screening assays used. The leading edge of peak A corresponded to an apparent M_r 38 900. As shown by

Fig. 1. Chromatography of Namalwa interferon on Sephadex G-75

Purified Namalwa interferon (64 μ g) in 2.0 ml of 0.14 M NaCl/20 mM sodium phosphate buffer, pH 7.2, was chromatographed on a column (1.0 cm $\times$ 198 cm) of Sephadex G-75 (superfine grade) in 50 mM NH_4HCO_3 /0.01% thiodiglycol at 20°C. Blue Dextran was added as an internal standard. The flow rate was 4.1 ml/h, and 1.03 ml fractions were collected. The eluate was monitored with a Uvicord S instrument (2.5 mm pathlength) at 206 nm (—). Interferon activity in selected fractions was determined (●).

Fig. 2. Chromatography of interferon at a high concentration on Sephadex G-75

Namalwa interferon (the same batch as was used for Fig. 1) was concentrated from 35 ml to 1 ml in an Amicon ultrafiltration cell with a UM-2 membrane at 4°C. The interferon (800 μ g), together with 1 ml of 50 mM NH_4HCO_3 used to rinse the membrane, was chromatographed as described in Fig. 1. The flow rate was 3.5 ml/h, and 0.85 ml fractions were collected. The eluate was monitored with a Uvicord S instrument (2.5 mm pathlength) at 206 nm (—). Samples were taken from appropriate fractions for SDS/polyacrylamide-gel electrophoresis (Fig. 3) and for antiviral assay (○).

SDS/polyacrylamide-gel electrophoresis (Fig. 3), the same components were present in fractions across the peak. This result indicated that a reversible

Fig. 3. SDS/polyacrylamide-gel electrophoresis of Sephadex column fractions of interferon

Samples (50 μ l) of fractions from the eluate of the Sephadex G-75 column shown in Fig. 2 were freeze-dried. The samples were subjected to electrophoresis through a 12.5% acrylamide gel after reduction as described in the Methods section. The protein bands were fixed and stained with Coomassie Brilliant Blue. The following samples were loaded track 1, standards, 5 μ g each of phosphorylase (M_r 97 000), bovine serum albumin (M_r 68 000), ovalbumin (M_r 45 000), soya-bean trypsin inhibitor (M_r 20 100), lysozyme (M_r 14 500) and insulin B-chains (M_r 3400); tracks 2–15, fractions 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 125, 126 and 128; track 16, interferon before separation on Sephadex G-75 (5 μ g).

dimerization of components of IFN- α A was occurring and this was confirmed by rechromatography of a sample (0.85 ml containing 25 μ g of protein) from the leading edge of peak A (fraction 96): the diluted protein now had apparent M_r 21 900.

When intermediate quantities (0.1–0.3 mg) of interferon were chromatographed (Fantes & Allen, 1981), the results were consistent with a reversible dimerization of components of IFN- α A, with a dissociation constant of the order of 1 μ M.

As shown in Fig. 3, on SDS/polyacrylamide-gel electrophoresis IFN- α A gave three main bands, M_r 20 900, 20 100 and 18 600, and IFN- α B gave four main bands, M_r 20 900, 20 400, 19 100 and 18 400, with the use of the standards shown. These values are $\pm$ 500, although the differences between different species are reproducible. A minor band, M_r 27 500, was also detected; a minor impurity in fraction 96 was also seen. Material used for amino acid sequence analysis gave only the seven major bands. Other preparations of interferon yielded a greater proportion of the 27 500- M_r band (Fantes & Allen, 1981), as shown in Fig. 4, track 1.

When the interferon was not reduced before SDS/polyacrylamide-gel electrophoresis, the distri-

Fig. 4. SDS/polyacrylamide-gel electrophoresis of interferon mixtures

Samples of interferon after separation by Sephadex G-75 chromatography were subjected to SDS/polyacrylamide-gel electrophoresis (15% acrylamide) as described in the Methods section, with or without reduction. A 3–10 μ g portion of each sample was applied. Track 1, IFN- α A, reduced; track 2, IFN- α B, reduced; track 3, IFN- α A, not reduced; track 4, IFN- α B, not reduced; track 5, total IFN- α , not reduced; track 6, standards, reduced: bovine serum albumin (M_r 68 000), ovalbumin (M_r 45 000), soya-bean trypsin inhibitor (M_r 20 100) and lysozyme (M_r 14 500); track 7, fraction of IFN- α A containing a high proportion of the 27 500- M_r component, reduced; track 8, IFN- α B, reduced; track 9, as track 6. Tracks 7–9 were run on a separate gel slab. Protein bands were stained with Coomassie Blue.

bution of bands was different, as shown in Fig. 4, tracks 3–5.

Each of the eight reduced proteins (Fig. 4, tracks 1 and 2) eluted from gel slices had antiviral activity, although the recovery of activity was low (Table 1). The activity recovered from bands of unreduced protein was much higher. Assuming a specific activity of 1.8×10^8 i.u./mg of protein, and equal activities for all components, about 1.8×10^5 i.u. per gel slice was expected. A clear separation of bands from unreduced samples was not obtained, but activity was distributed over the width of the partially resolved bands, totalling 11% of the expected recovery for IFN- α A and 47% for IFN- α B.

In an attempt to resolve more effectively the various interferon species, the long-gradient gel system (Knight, 1975; Berg & Heron, 1980) was used. The bands were on the whole more widely separated, particularly the unreduced components, but no further resolution into different components was seen. The use of a different buffer system (Chua & Bennoun, 1975) also gave a very similar pattern

of bands (results not shown). Isoelectric focusing in polyacrylamide gels revealed that more components were present in Namalwa interferon than were separated by SDS/polyacrylamide-gel electrophoresis. IFN- α B contained about 12 components with isoelectric points in the range 4.5–7, and IFN- α A gave only about three different bands with isoelectric points between 5 and 6 (results not shown).

Amino acid analyses of IFN- α A and IFN- α B (Table 2) are similar to each other and in reasonable agreement with the calculated compositions for two species of known sequence, IFN- α 1 (Mantei *et al.*, 1980) and IFN- α 2 (Streuli *et al.*, 1980). These analyses were performed on interferon samples which lacked the 27 500- M_r component, and no glucosamine or galactosamine (less than 0.2 residue per molecule) was detected. Amino acid analyses of three samples of interferon isolated after Sephadex G-75 chromatography from fractions in which the 27 500- M_r component was prominent (e.g. Fig. 4, track 7) revealed the presence of glucosamine, 0.5 ± 0.07 residue per molecule (uncorrected). Since the species of lower molecular weight contain undetectable amounts of glucosamine, it is likely that the 27 500- M_r species is the sole glycosylated species. The estimated proportion of this species in the mixtures analysed was 25–40%, as judged by Coomassie Blue staining. After correction for 30% destruction of glucosamine during hydrolysis, these results indicate that the 27 500- M_r component contains about 2 glucosamine residues per molecule.

N-Terminal sequences of IFN- α A and IFN- α B were reported earlier (Allen & Fantes, 1980) and are shown in Fig. 6. The suggestion that the *N*-terminal cysteine residue was involved in disulphide bonding (Allen & Fantes, 1980) has been confirmed, at least for one species (Wetzel, 1981).

Tryptic-digest peptide 'maps' of performic acid-oxidized IFN- α A and IFN- α B are shown in Fig. 5. The appearance of the 'map' of IFN- α A peptides was consistent with the protein being homogeneous, or with only limited microheterogeneity. The number of spots was consistent with cleavage at all arginine and lysine residues. However, the 'map' of IFN- α B peptides contained more spots, of various intensities, than would be consistent with a single peptide sequence of M_r 20 000. Most of the strong spots were different from those in the 'map' of IFN- α A peptides, except for those due to free lysine and arginine and one major peptide. Some weak spots in the 'map' of IFN- α B peptides corresponded to stronger spots in that of IFN- α A peptides. Peptide extracted from the thin layers were subjected to partial sequence analysis; larger quantities of tryptic peptides were isolated by gel filtration followed by peptide 'mapping' of pooled fractions.

The separation of peptides from the proteinase

Table 1. *Recovery of antiviral activity after SDS/polyacrylamide-gel electrophoresis of interferon samples*
 Samples of IFN- α A and IFN- α B (each about 5 μ g of protein) were subjected to electrophoresis in 15% acrylamide gels, and protein bands were eluted for determination of antiviral activity, as described in the text. The interferon bands are labelled in order of decreasing apparent molecular weight. The expected activity is based on the assumption that all components have the same specific activity.

Interferon component	M_r	IFN units recovered	Approx. % of expected activity
(i) Reduced interferon			
IFN- α A1	27 500	3060	1.7
IFN- α A2	20 900	3510	2.0
IFN- α A3	20 100	3940	2.2
IFN- α A4	18 600	2430	1.4
IFN- α B1	20 900	3940	2.2
IFN- α B2	20 400	6400	3.6
IFN- α B3	19 100	22 180	12.3
IFN- α B4	18 400	26 070	14.5
Control slice		<450	
(ii) Non-reduced interferon			
IFN- α A(1)		23 770	13.2
IFN- α A(2)		6400	3.6
IFN- α A(3)		32 820	18.2
IFN- α A(4)		17 220	9.6
IFN- α B		335 800	46.6

Table 2. *Amino acid analyses of interferon samples*

Values are adjusted to a total of 166 residues, except for the calculated composition of IFN- α 2 (165 residues). IFN- α A' contains as a major component a species of M_r 27 500, which is absent from those preparations of IFN- α A or IFN- α B analysed in the present work. Results for IFN- α A and IFN- α B (averages of five determinations) are corrected for destruction of serine and threonine and for incomplete hydrolysis after 24 h; those for IFN- α A' (three determinations) are uncorrected. N.D., Not determined. Errors are $\pm 5\%$ for most residues, but $\pm 10\%$ for proline, cysteine, methionine, tyrosine, histidine, arginine, tryptophan and glucosamine.

	Composition (residues/molecule)				
	IFN- α A	IFN- α B	IFN- α A'	Calc. for IFN- α 1 (Mantei <i>et al.</i> , 1980)	Calc. for IFN- α 2 (Streuli <i>et al.</i> , 1980)
Asx	13.1	14.3	14.6	17	12
Thr	9.4	8.8	8.9	9	10
Ser	13.2	13.0	12.8	13	14
Glx	24.6	25.5	22.9	25	26
Pro	7.5	7.4	7.2	6	5
Gly	6.4	5.5	6.8	3	5
Ala	9.2	9.4	11.3	10	8
Cys	4.5*	4.8*	N.D.	5	4
Val	7.4	7.7	7.9	6	7
Met	5.0†	5.2†	2.4	6	5
Ile	7.0	7.4	6.9	7	8
Leu	18.9	19.1	20.4	22	21
Tyr	3.9	3.3	6.0	4	5
Phe	8.5	8.2	9.6	8	10
His	3.2	3.5	3.6	3	3
Lys	10.7	10.0	10.9	8	10
Arg	11.0	11.2	10.6	12	10
Trp	2.2	1.7	N.D.	2	2
GlcN	<0.2	<0.2	0.5		
GalN	<0.2	<0.2			
Molecular weight ...				19 383	19 253

* As cysteic acid.

† As methionine sulphone.

Fig. 5. Peptide 'maps' of interferon mixtures IFN- α A (a) and IFN- α B (b)

The peptide 'maps' of performic acid oxidized proteins were prepared as described in the text. Peptides were detected with the ninhydrin/Cd reagent, and gave pink spots except where indicated: y, yellow, y→p, yellow turning pink. Other symbols: W, tryptophan-containing peptides (fluorescent); R, arginine-containing peptides; Δ, internal marker, *N*^ε-dinitrophenyl-lysine; X, internal marker, Xylene Cyanol FF; I, origin; - ↔ +, direction of electrophoresis at pH 4.4; †, direction of chromatography.

digestions (2)–(5) by chromatography on Sephadex G-50 columns followed by thin-layer peptide 'mapping' of pools of fractions is described in Supplement SUP 50117. Amino acid analyses, where determined, are also presented in Supplement SUP 50117. A number of mainly small peptides were not subjected to quantitative analysis, either because their 'map' positions, *N*-terminal amino acids and semi-quantitative compositions determined by the dansyl method (Allen, 1981) were identical with those of known peptides, or, more often, because of insufficient yield, in which case the total amount of peptide recovered was used for sequence determinations. There were, in addition, some peptides obtained from the chymotryptic digests that were not sufficiently pure for reliable interpretation of the analytical results.

Peptide sequences that were determined are presented in Fig. 6. The sequences could readily be aligned by homology with complete amino acid sequences deduced from nucleotide sequences of cloned complementary DNA; the sequence of IFN- α 1 (Mantei *et al.*, 1980) is given at the top of the Figure. Many peptide sequences were only partially determined, owing to limitations in the amounts of purified peptides, and it is likely that additional

sequences would be found with further study on larger amounts of protein. However, some clear comparisons with the amino acid sequences of interferon species deduced from sequences of complementary DNA can be made, and these are discussed below.

Discussion

Eight molecular species of IFN- α from Namalwa cells could be resolved by chromatography on Sephadex G-75 followed by SDS/polyacrylamide-gel electrophoresis. Eight major species were also reported for human leucocyte interferon subjected to high-pressure liquid chromatography (Rubinstein *et al.*, 1981). The low recovery of antiviral activity after gel electrophoresis could be accounted for partly through losses in sample preparation and extraction from gel slices, but mainly through inactivation by reduction of IFN- α (Mogensen & Cantell, 1974; Stewart *et al.*, 1974).

Non-reduced IFN species were poorly resolved by SDS/polyacrylamide-gel electrophoresis, but the distribution of protein stain and antiviral activity in gel slices is consistent with previous studies on crude leucocyte-type human interferon (Stewart &

Fig. 6. Amino acid sequences of peptides from IFN- α A and IFN- α B. Peptide sequences were determined by either the dansyl-Edman method (→) or the dimethylaminoazobenzene isothiocyanate method (→) or both (→); results not considered definitive are shown thus: →, →, →. The standard three-letter code is used, with the addition of Ile for either Leu or Ile (not readily distinguished by the dimethylaminoazobenzene isothiocyanate method). Residues identified in peptides by analysis, but not placed in sequence, are enclosed in parentheses and separated by commas. Peptides are aligned by homology with the complete sequence of IFN- α 1 deduced from the sequence of the complementary DNA (Mantel *et al.*, 1980). The numbering system is for this sequence: IFN- α A sequences have a deletion at residue 44. An unidentified dansyl derivative, in addition to dansyl-serine, was observed in some peptides at position 25 (marked S α). Peptides are labelled according to their origin (IFN- α A or IFN- α B), the digestion method (T, trypsin; Ch, chymotrypsin; Sp, staphylococcal proteinase) and their separation on Sephadex G-50 and on thin-layer peptide 'maps'.

Desmyter, 1975; Desmyter & Stewart, 1976; Bridgen *et al.*, 1977; Havell *et al.*, 1977; Chadha *et al.* 1978; Grob & Chadha, 1979).

Although both gel filtration and SDS/polyacrylamide-gel electrophoresis separate proteins primarily on the basis of molecular weight, differences in amino acid compositions and sequences between different proteins of the same molecular weight can lead to pronounced differences in their behaviour in these systems. A single amino acid change is sufficient for this (de Jong *et al.*, 1978; Noel *et al.*, 1979). From the amino acid sequence results, particularly the high yields of both *N*- and *C*-terminal peptides, and comparison with several complete sequences of IFN- α species deduced from sequences of complementary DNA (Mantei *et al.*, 1980; Streuli *et al.*, 1980; Goeddel *et al.*, 1981), it is likely that all species studied in the present work have 165 or 166 residues and M_r 19 300 $\pm$ 200, unless so-far-undetected post-translational modifications are present. The 27 500- M_r species was not studied by sequencing and may be glycosylated (see below). Forms of IFN- α lacking the ten *C*-terminal amino acids, as described by other workers (Levy *et al.*, 1981), were not detected in the present work.

The gel-filtration results show that human IFN- α is a typical globular protein, comparable with soya-bean trypsin inhibitor in hydrodynamic properties. IFN- α A may be slightly more elongated, and is capable of reversible dimerization, with a very approximate $K_{diss.}$ value of 1 μ M. An accurate dissociation constant was not determined, owing to the complexity of the mixture (e.g. the 18 600- M_r species has a smaller tendency to associate) and lack of information on rates of association. At a concentration of about 20 μ g/ml in the eluate from the Sephadex G-75 column, IFN- α A is about half dimerized, as seen in several chromatographic runs. The 27 500- M_r component (almost absent from the sample shown in Fig. 3) does not associate under these conditions.

The amino acid compositions of IFN- α A and IFN- α B were similar to that reported by Rubinstein *et al.* (1979) for a single component of leucocyte interferon; only the tryptophan content was significantly different, and the spectrophotometric method used in the present work is generally more reliable than methods employing hydrolysis. The analysis reported by Zoon *et al.* (1979) differed for several amino acids, again probably because of technical problems. The compositions found in the present work are close to those expected for mixtures of IFN- α species of known sequence (see below) and are similar to those reported for purified components of leucocyte interferon (Rubinstein *et al.*, 1981).

No glucosamine or galactosamine, typically found

in glycoproteins, were detected in IFN- α A or IFN- α B mixtures lacking the 27 500- M_r component. However, glucosamine was present in mixtures containing the 27 500- M_r component as a major species, in amounts consistent with two residues per molecule. Although these results do not demonstrate conclusively that the 18 400–20 900- M_r species are not glycosylated whereas the 27 500- M_r species is, they are strongly suggestive of this interpretation. Attempts to recover sufficient pure 27 500- M_r component from polyacrylamide-gel slices after electrophoresis for determination of amino sugars were unsuccessful: contaminants eluted from the gel interfered with the analysis. No periodate/Schiff-reagent-positive bands were seen on polyacrylamide gels; however, this reagent has low sensitivity for some glycoproteins.

Additional evidence that most IFN- α species are not glycoproteins, at least of the asparagine-linked type, is that no Asn-Xaa-(Ser or Thr) sequences, which form potential sites for glycosylation, were found in the present work. All but one of the eight sequences of IFN- α species deduced from the sequence of complementary DNA lack such sequences (Goeddel *et al.*, 1981). In addition, no glycopeptides, which typically have very low R_F values on t.l.c., were detected in the present work on the 18 400–20 900- M_r species. This situation is in contrast with that for INF- β , which has an Asn-Glu-Thr sequence (Taniguchi *et al.*, 1980a) and was shown to be a glycoprotein (Knight, 1976; Tan *et al.*, 1979).

Although IFN- α was reported to be glycosylated (Bose *et al.*, 1976; Bridgen *et al.*, 1977; Stewart *et al.*, 1979; Salit & Ogburn, 1980), some studies with the use of tunicamycin (Chadha *et al.*, 1980) or concanavalin A affinity chromatography (Grob & Chadha, 1979), suggest that only a small proportion is glycosylated, in agreement with my conclusion. Rubinstein *et al.* (1981) have also concluded that most, and perhaps all, human IFN- α species are devoid of carbohydrate.

The amino acid sequences determined in the present work agree well with those deduced from complementary DNA species. IFN- α A contains at least two very closely related sequences corresponding to IFN- α 2 (Streuli *et al.*, 1980) or Le IFN- α A (Goeddel *et al.*, 1981), which differ only at residue 23 (Lys or Arg). There is microheterogeneity in IFN- α A at positions 34 (Arg + His), 53 (Ala + Thr), 55 (Ser + Pro) and possibly 74 (weakly detected Glx as well as Ser). All IFN- α A species have a deletion at residue 44 in comparison with most IFN- α species. It is likely that Asp-84 and Pro-88 (Allen & Fantes, 1980) were incorrectly identified; the former was probably an overlap from Asp-83, and the latter an artifact spot occasionally seen on chromatograms with the dimethylamino-

Fig. 7. Homologous sequences in cholera toxin B-subunit, *E. coli* heat-labile toxin and human IFN species

The sequence data were taken from the following sources: cholera toxin B, Lai (1977) and Kurosky *et al.* (1977); *E. coli* heat-labile toxin (LT), Dallas & Falkow (1980); Hu IFN- β , Taniguchi *et al.* (1980a); Hu IFN- α_1 , Mantei *et al.* (1980); Le IFN- α A, Goeddel *et al.* (1981). Allowance has been made for a single deletion in each sequence to maximize homology. Residues in either toxin that are identical with one or more interferon sequences are enclosed in boxes. The homology between interferon and cholera toxin was pointed out by Anfinsen (1981).

azobenzene isothiocyanate method. An unidentified weak spot was seen on t.l.c. plates in addition to dansyl-serine when sequencing through residue 25 in some peptides (see Fig. 6); this might be indicative of a modified residue.

IFN- α B is a more complex mixture than IFN- α A. There is little, if any (<10%), of Le IFN- α B or Le IFN- α H (Goeddel *et al.*, 1981) present, since peptides characteristic of these species were not detected. However, significant amounts (>10%) of species very similar to, if not identical with, Le IFN- α C, Le IFN- α D, Le IFN- α F and possibly Le IFN- α G (Goeddel *et al.*, 1981) are probably present. Since there are 10–15 different IFN- α gene sequences (Nagata *et al.*, 1980, 1981; Goeddel *et al.*, 1981), some of which are highly homologous, it is not surprising that some differences were found, e.g. no Gly-19 (present in Le IFN- α C) and no Leu-60 or Val-114 (present in Le IFN- α D) were seen. Ala-114 was present in IFN- α B, as in IFN- α 1 (Mantei *et al.*, 1980), which differs only at this residue from Le IFN- α D.

No evidence for blocked *N*-terminal peptides was obtained, in contrast with the observations made by Rubinstein *et al.* (1981) and Levy *et al.* (1981). Nor was any *N*-terminal serine detected, as reported for one component of IFN- α by Zoon *et al.* (1980), and for other components by Levy *et al.* (1981).

Since the peptide sequences determined in the present work are not complete, it has not been possible to identify with certainty individual species of known complete sequence. It may be concluded, however, that partial structures of Namalwa IFN- α are consistent with structures of individual species of leucocyte or lymphoblastoid interferon deduced in other laboratories from sequences of complementary DNA, and that the mature proteins extend up to the termination codon determined in the

sequences of the complementary DNA.

The amino acid sequences of IFN- α species are unremarkable, lacking long stretches of hydrophobic, acidic or basic residues. Analysis of the sequences of Le IFN- α A, Le IFN- α C, Le IFN- α D and Le IFN- α F (Goeddel *et al.*, 1981) shows that the *N*-terminal and *C*-terminal regions are relatively basic: the sequence 1–34 contains 2–4 acidic but 6–8 basic residues, and the sequence 121–166 contains 4 or 5 acidic but 10–12 basic residues; where the central portion of the molecule is acidic, containing 15–17 acidic but only 4–5 basic residues in the sequence 35–120. Comparison of published sequences shows (Goeddel *et al.*, 1981), as observed in the present work, that residues 115–151 form a highly conserved region, which is also similar in human IFN- β (Taniguchi *et al.*, 1980b). Part of this sequence is homologous with sequences in cholera toxin (Anfinsen, 1981) and a heat-labile toxin from *Escherichia coli*, as shown in Fig. 7. This homology, which does not extend to the rest of the molecules, suggests a related function, possibly binding to cell surfaces (Anfinsen, 1981), although cholera toxin does not appear to act directly on the IFN- α -receptor site (Fuse & Kuwata, 1979).

I thank Dr. K. H. Fantes for supplying interferon and for constant advice and help during this work, and Ms. P. Shockley for performing interferon assays.

References

- Allen, G. (1981) *Laboratory Methods in Biochemistry and Molecular Biology* (Work, T. S. & Burdon, R. H., eds.), Vol. 9: *Sequencing of Proteins and Peptides*, Elsevier/North-Holland Biomedical Press, Amsterdam.
- Allen, G. & Fantes, K. H. (1980) *Nature (London)* **287**, 408–411.
- Anfinsen, C. B. (1981) *Interdiscip. Sci. Rev.* **6**, 110–118.
- Berg, K. & Heron, L. (1980) *J. Gen. Virol.* **50**, 441–446.

- Bose, S., Gurari-Rotman, D., Ruegg, V. T., Corley, L. & Anfinsen, C. B. (1976) *J. Biol. Chem.* **251**, 1659–1662
- Bridgen, P. J., Anfinsen, C. B., Corley, L., Bose, S., Zoon, K. C., Ruegg, V. T. & Buckler, C. E. (1977) *J. Biol. Chem.* **252**, 6585–6587
- Bruton, C. J. & Hartley, B. S. (1970) *J. Mol. Biol.* **52**, 165–178
- Chadha, K. C., Sclair, M., Sulkowski, E. & Carter, W. A. (1978) *Biochemistry* **17**, 196–200
- Chadha, K. C., Grob, P. M., Hamill, R. L. & Sulkowski, E. (1980) *Arch. Virol.* **64**, 109–117
- Chang, J. Y., Brauer, D. & Wittmann-Liebold, B. (1978) *FEBS Lett.* **93**, 205–214
- Chen, J. K., Jankowski, W. J., O'Malley, J. A., Sulkowski, E. & Carter, W. A. (1976) *J. Virol.* **19**, 425–434
- Christofinis, G. J. (1970) *J. Med. Microbiol.* **3**, 251–258
- Chua, N. H. & Bennoun, P. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 2175–2179
- Dallas, W. S. & Falkow, S. (1980) *Nature (London)* **288**, 499–501
- de Jong, W. W., Zweers, A. & Cohen, L. H. (1978) *Biochem. Biophys. Res. Commun.* **82**, 532–539
- Desmyter, J. & Stewart, W. E., II (1976) *Virology* **70**, 451–458
- Edelhoc, H. (1967) *Biochemistry* **6**, 1948–1954
- Fantes, K. H. (1970) *Ann. N.Y. Acad. Sci.* **173**, 118–121
- Fantes, K. H. & Allen, G. (1981) *J. Interferon Res.* **1**, 465–472
- Fantes, K. H., Burman, C. J., Ball, G. D., Johnston, M. D. & Finter, N. B. (1980) *Biochemical Characterization of Lymphokines* (de Weck, A. L., Kristensen, F. & Landy, M., eds.), pp. 343–351, Academic Press, London
- Finter, N. B. (1969) *J. Gen. Virol.* **5**, 419–427
- Finter, N. B. (1982) *Tex. Rep. Biol. Med.* **41**, in the press
- Fuse, A. & Kuwata, T. (1979) *Infect. Immun.* **26**, 235–239
- 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
- Grob, P. M. & Chadha, K. C. (1979) *Biochemistry* **18**, 5782–5786
- Havell, E. A., Yip, Y. K. & Vilček, J. (1977) *Arch. Virol.* **55**, 121–129
- Heiland, I., Brauer, D. & Wittmann-Liebold, B. (1976) *Hoppe-Seyler's Z. Physiol. Chem.* **357**, 1751–1770
- Hirs, C. H. W. (1967) *Methods Enzymol.* **11**, 197–199
- Knight, E., Jr. (1975) *J. Biol. Chem.* **250**, 4139–4143
- Knight, E., Jr. (1976) *Proc. Natl. Acad. Sci. U.S.A.* **73**, 520–523
- Kurosky, A., Markel, D. E. & Peterson, J. W. (1977) *J. Biol. Chem.* **252**, 7257–7264
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Lai, C.-Y. (1977) *J. Biol. Chem.* **252**, 7249–7256
- Levy, W. P., Rubinstein, M., Shively, J., Del Valle, 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
- Mantei, N., Schwarzstein, M., Streuli, M., Panem, S., Nagata, S. & Weissmann, C. (1980) *Gene* **10**, 1–10
- Mogensen, K. E. & Cantell, K. (1974) *J. Gen. Virol.* **22**, 95–103
- Mogensen, K. E., Pyhälä, L., Törmä, E. & Cantell, K. (1974) *Acta Pathol. Microbiol. Scand. Sect. B* **82**, 305–310
- Moore, G. E., Gerner, R. E. & Franklin, H. A. (1967) *J. Am. Med. Assoc.* **199**, 519–524
- Nagata, S., Mantei, N. & Weissmann, C. (1980) *Nature (London)* **287**, 401–408
- Nagata, S., Brack, C., Hencø, K., Schamböck, A. & Weissmann, C. (1981) *J. Interferon Res.* **1**, 333–336
- Noel, D., Nikaido, K. & Ames, G. F. L. (1979) *Biochemistry* **18**, 4159–4165
- Paucker, K., Dalton, B. J., Törmä, E. T. & Ogburn, C. A. (1977) *J. Gen. Virol.* **35**, 341–351
- Roth, M. & Hampai, A. (1973) *J. Chromatogr.* **83**, 353–356
- Rubinstein, M., Rubinstein, S., Familletti, P. C., Miller, R. S., Waldman, A. A. & Pestka, S. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 640–644
- Rubinstein, M., Levy, W. P., Moschera, J. A., Lai, C.-Y., Hershberg, R. D., Bartlett, R. T. & Pestka, S. (1981) *Arch. Biochem. Biophys.* **216**, 307–318
- Salit, M. G. & Ogburn, C. A. (1980) *Arch. Virol.* **63**, 133–142
- Stewart, W. E., II, & Desmyter, J. (1975) *Virology* **67**, 68–73
- Stewart, W. E., II, DeClercq, E. & De Somer, P. (1974) *Nature (London)* **249**, 460–461
- Stewart, W. E., II, Lin, L. S., Wiranowska-Stewart, M. & Cantell, K. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4200–4204
- Stewart, W. E., II, Wiranowska-Stewart, M., Koistinen, V. & Cantell, K. (1979) *Virology* **97**, 473–476
- Stewart, W. E., II, Blalock, J. E., Burke, D. C., Chany, C., Dunnick, J. K., Falcoff, E., Friedman, R. M., Galasso, G. J., Joklik, W. K., Vilček, J. T., Youngner, J. S. & Zoon, K. C. (1980) *Nature (London)* **286**, 110
- Strander, H., Mogensen, K. E. & Cantell, K. (1975) *J. Clin. Microbiol.* **1**, 116–117
- Streuli, M., Nagata, S. & Weissmann, C. (1980) *Science* **209**, 1343–1347
- Tan, Y. H., Barakat, F., Berthold, W., Smith-Johannsen, H. & Tan, C. (1979) *J. Biol. Chem.* **254**, 8067–8073
- Taniguchi, T., Ohno, S., Fujii-Kurimaya, Y. & Muramatsu, M. (1980a) *Gene* **10**, 11–15
- Taniguchi, T., Mantei, N., Schwarzstein, M., Nagata, S., Muramatsu, M. & Weissmann, C. (1980b) *Nature (London)* **285**, 547–549
- Thang, M. N., Thang, D. C., Chelbi-Alix, M. K., Robert-Galliot, B., Commoey-Chevalier, M. J. & Chany, C. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 3717–3721
- Törmä, E. T. & Paucker, K. (1976) *J. Biol. Chem.* **251**, 4810–4816
- Wetzel, R. (1981) *Nature (London)*, **289**, 606–607
- Zoon, K. C., Smith, M. E., Bridgen, P. J., zur Nedden, D. & Anfinsen, C. B. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 5601–5605
- Zoon, K. C., Smith, M. E., Bridgen, P. J., Anfinsen, C. B., Hunkapiller, M. W. & Hood, L. E. (1980) *Science* **207**, 527–528