Characterization of human interleukin 2 derived from *Escherichia coli*

Shu-Mei LIANG,* Bernard ALLET,* Keith ROSE,† Marlis HIRSCHI,* Chi-Ming LIANG*
and David R. THATCHER*†

* Biogen S.A., 3 route de Troinex, Carouge, Geneva, Switzerland, and † Département de Biochimie Médicale, Centre Médical Universitaire, 9 avenue de Champel, 1211 Geneva 4, Switzerland

(Received 27 December 1984/15 February 1985; accepted 22 March 1985)

Interleukin 2 isolated from *Escherichia coli* cells expressing the human interleukin gene has been characterized. The observed properties of the protein have been compared with those properties which can be deduced from the DNA sequence alone and the published properties of natural human interleukin 2. The purified *E. coli*-derived interleukin 2 is a monomeric protein of M_r 15000 with a sedimentation velocity of 1.86S. The amino acid composition of the protein and isoelectric point (7.7) are consistent with that part of the translated DNA sequence of the gene corresponding to the mature protein. A single disulphide bridge was identified between Cys-58 and Cys-105. C.d. suggested that interleukin 2 is predominantly α -helical in secondary structure. The *E. coli*-derived protein differed from natural interleukin 2 in the presence of *N*-terminal methionine and also in the absence of a carbohydrate moiety. Removal of the coding region for the first three amino acids of the natural interleukin 2 protein sequence (Ala-Pro-Thr) by site-specific mutagenesis resulted in a protein with *N*-terminal serine. The possibility that the specificity of the *E. coli* ribosomal methionine aminopeptidase may not recognize the sequence NH_2 -Met-Xaa-Pro is discussed (where Xaa is any amino acid residue).

Interleukin 2 is an immunomodulatory factor produced by certain subsets of T-lymphocytes. This factor or lymphokine was first identified by its ability to promote the long term proliferation of activated T-cells (Morgan *et al.*, 1976) and was originally called T-cell growth factor. Recently, interleukin 2 has been shown to have other immunomodulatory effects on T-cells and related cell types: the promotion of NK cell proliferation and their activation (Ortaldo *et al.*, 1984), the induction of γ -interferon secretion (Farrar *et al.*, 1982), induction of B-cell growth factor secretion (Inaba *et al.*, 1983), and the differentiation of lymphokine-activated killer (LAK) cell activity (Grimm *et al.*, 1982) from large granular leukocytes.

Human interleukin 2 has been partially purified from lectin-stimulated peripheral blood lymphocytes (Welte *et al.*, 1982; Godard *et al.*, 1984), purified to apparent homogeneity from stimulated

buffy coats (Knip *et al.*, 1984) and isolated from a constitutive T-cell-leukaemia-derived cell line JURKAT (Robb *et al.*, 1983a).

The cloning of the interleukin 2 gene (Taniguchi *et al.*, 1983; Devos *et al.*, 1983), and its expression at high levels in *Escherichia coli*, has created the possibility of producing large quantities of purified protein that is not contaminated by other natural lymphokines, growth factors or interferons. In the present paper we describe characterization of the protein produced from *E. coli* harbouring an expression plasmid containing the interleukin 2 gene and compare these properties with existing information on the properties of natural interleukin 2.

Materials and methods

Materials

Interleukin 2 was purified from an *E. coli* strain (C600) harbouring a pBR322 derivative plasmid (Devos *et al.*, 1983). At high expression levels the protein was found to be aggregated into inclusion bodies in the cells and this property was exploited in the purification strategy. The protein aggregates

† To whom correspondence and reprint requests should be addressed.

Abbreviation used: SDS, sodium dodecyl sulphate; Ilx, leucine or isoleucine (Rose *et al.*, 1983b).

were purified by differential centrifugation from cells ruptured by passage through a French press at 55.2 MPa (8000 lbf/in²). After solubilization, interleukin 2 was purified by reversed-phase h.p.l.c. and gel filtration.

Iodo[2-³H]acetic acid was obtained from Amersham International and had a specific radioactivity of 50 mCi/mmol. All other laboratory reagents were purchased from Fluka AG, Buchs, Switzerland, were of analytical grade and were used without further purification.

Gel electrophoresis and isoelectric focusing

SDS/polyacrylamide-gel electrophoresis was performed as described by Laemmli (1970) and gels were stained with silver as described by Wray *et al.* (1981). Gel isoelectric focusing was performed with pre-prepared polyacrylamide gels, ampholine pH range 3.5–9.5 (LKB, Bromma, Sweden) on a Pharmacia FBE 3000 electrophoresis apparatus (Pharmacia, Uppsala, Sweden) using the exact methods described by the manufacturers. The isoelectric point was predicted from the amino acid composition by using a simple computer program written in BASIC for a Sharp PC1500 pocket computer. This program calculates the fractional charge of each particular charged amino acid species by using the Henderson–Hasselbach equation and pK_a values taken from Cantor & Schimmel (1980). All the fractional charges are summed at each pH value (every 0.1 pH unit between 1 and 14) and the pH value where the net summed charge is 0 is assumed to be the isoelectric point.

Gel permeation chromatography

An LKB Blue column (TSK gel G30005w) (7.5 mm × 600 mm) was equilibrated with 0.1 M-NaH₂PO₄ buffer, pH 4.0. Interleukin 2 (100 µl of a 1 mg/ml solution) was injected by using a Pharmacia V7 valve. The column was run at a flow rate of 0.5 ml/min with a Pharmacia P500 pump and the elution stream was monitored with a Pharmacia UV-1 detector at a wavelength of 280 nm. A solution (1 mg/ml) of each of α -globulin, bovine serum albumin, ovalbumin, chymotrypsinogen and ribonuclease was used to calibrate the column.

Analytical ultracentrifugation

Analytical ultracentrifugation was performed using a Beckman Prep UV Scanner and an AN-F two-place rotor fitted with double sector centrifuge cells. A stock solution of protein was prepared by lyophilizing 3.5 mg of interleukin 2 which had been dialysed against water and then dissolving the residue in 2 ml of 50 mM-acetic acid. Of this stock, 1 ml was taken and the other 1 ml was used to form a dilution series down to 0.13 mg/ml. The rotor

speed was adjusted to 60000 rev./min and the scanner was set to scan the sample compartment at 10 min intervals. Plots of the logarithm of the distance of the boundary mid-point versus time were drawn and the slope of the line was calculated by linear regression analysis. The s value was calculated from the slope of the regression line.

The theoretical value for the sedimentation coefficient was calculated from the Svedberg equation by using a partial specific volume calculated from the amino acid composition by the method of Cohn & Edsall (1943), an M_r predicted from the DNA sequence and a diffusion coefficient calculated from the Einstein–Sutherland equation. The intrinsic viscosity was assumed to be 1 and the anhydrous molecular radius for the Einstein equation was approximated from an assumed hydrated M_r of 1.6 times the sequence M_r , by using the equation described by Ackers (1970).

Chemical analysis

Amino acid analysis was performed on samples hydrolysed in 6 M-HCl, 10% (v/v) with respect to ethanethiol at 110°C, with a Beckman 6300 high performance analyser.

S-Carboxymethylation was performed with iodo[2-³H]acetic acid. The alkylating reagent contained 5 mCi of iodo[2-³H]acetic acid diluted with 826 µl of 0.5 M-iodoacetic acid and 4 ml of buffer containing 8 M-guanidinium chloride, 0.5 M-Tris and 0.05 M-EDTA (final pH adjusted to 8.5 with conc. HCl). Interleukin 2 (1 mg samples) was freeze-dried into 2 ml glass vials at a shelf temperature of –10°C. The vials were then capped, 200 µl of alkylating reagent was added and the freeze-dried residue dissolved. The tubes were then flushed with N₂ and incubated for 30 min at 37°C. The reaction was quenched by desalting the protein by gel filtration into 0.1 M-NH₄OH. The protein was then freeze-dried and the disulphide bridge was reduced by dissolving the residue in 0.5 ml of guanidinium chloride buffer and adding 20 µl of 0.1 M-dithiothreitol. After incubation for 30 min at 37°C under N₂, 20 µl of 0.5 M-iodoacetic acid was added. The reaction was quenched by gel filtration after further incubation at 37°C for 30 min. The desalted protein was digested by the addition of 20 µl of a 1 mg/ml solution of trypsin (Worthington Diagnostic Systems, batch 32N798) and incubation for 4 h at 37°C. The peptides were then freeze-dried.

Peptides were separated by reversed-phase h.p.l.c. A Spherisorb (0.4 cm × 30 cm) octadecyl stationary phase (Spectra Physics) was chosen and elution was performed using a 0–75% acetonitrile organic modifier gradient in a 0.1% trifluoroacetic acid mobile phase at a flow rate of 1.4 ml/min. Peptides were detected at 214 nm and 0.7 ml

fractions were collected. Portions (100 μ l) of each fraction were mixed with 2 ml of Aquasol-2 (New England Nuclear) and counted for ^3H in a Beckman LS 7000 scintillation counter.

Radioactive peptides were analysed after acid hydrolysis as described for the protein analysis above.

Sequence studies on the intact protein were performed by the g.l.c.-m.s. method. The method of identifying the sequence of the C-terminal peptide has been described in detail by Rose *et al.* (1983b) and of N-terminal peptide analysis by Rose *et al.* (1983a,b).

Carbohydrate content was estimated by the anthrone method of Roe (1955) using glucose as a standard and ovalbumin as a positive control.

Secondary structure analysis

C.d. spectra were obtained with a Jasco J500 machine. A sample of interleukin 2 (0.182 mg/ml) was dialysed against 50 mM-acetic acid and the spectra were recorded using a 2 mm cell, a time constant of 16 s and a scan speed of 4 nm/min. Scale expansion was set at 5 millidegrees/cm. Secondary structure was predicted from the sequence by the approach of Chou & Fasman (1974). The sequence was scanned at 6 residues/scan to establish centres of high secondary structural potential.

Bioassay

Interleukin 2 activity was measured by the stimulation of [^3H]thymidine incorporation into a murine interleukin 2 dependent T-cell line CTLL-2 (Gillis *et al.*, 1978). The cells were routinely cultured on 2 ng of recombinant interleukin 2/ml in the presence of Dulbecco's minimal essential medium supplemented with 10% (v/v) foetal calf serum.

Radioimmunoassay

Antibodies against *E. coli* proteins were raised by injecting a rabbit with 200 μ g of *E. coli* proteins in Freund's complete adjuvant. The rabbit was boosted with the same amount of antigen twice at 21 day intervals. The antibodies were then affinity purified by protein A-Sepharose. Some of the purified antibodies were labelled with ^{125}I using Enzymobead (Bio-Rad) to a specific radioactivity of 10–15 Ci/g.

Solid phase radioimmunoassay was performed by coating a 96-well polystyrene plate with 50 μ l of the purified antibody (5 μ g/ml) at room temperature overnight and then saturating it with human serum albumin (10 mg/ml, 200 μ l/well).

The plate was subsequently incubated with standard and unknown amounts of *E. coli* proteins at room temperature for 5 h. After washing with phosphate-buffered saline, 100 000 c.p.m. of ^{125}I -labelled rabbit antibodies were added (50 μ l/well)

and incubated at 4°C for 2 h. After washing four times thoroughly with phosphate-buffered saline containing 0.5% Tween 20, 100 μ l of 1 M-NaOH was added to each well. The solution was then counted in a Beckman Gamma 4000 counter. The sensitivity of the assay was 10–20 ng of *E. coli* proteins/ml.

Endotoxin analysis

The rabbit pyrogen test and the *Limulus* amoebocyte lysate test were performed as described in the U.S. Pharmacopeia.

Construction of truncated interleukin 2

The genetic coding sequence for the first three amino acid residues of the mature natural protein sequence (Ala-Pro-Thr) was removed by site-specific mutagenesis and the truncated gene was expressed in *E. coli* using a pBR322-derived expression vector (B. A., unpublished work). The level of expression, as observed by SDS/polyacrylamide-gel electrophoresis of crude *E. coli* lysates (5–10% of total protein) was the same both for the original clone and for the deletion mutant and no differences in behaviour were observed during the purification of this mutant. In both the mutant and the original clone, the proteins were laid down as insoluble cellular aggregates which were purified by differential centrifugation. The specific activity of the partially purified protein was 2.3×10^6 units/mg, similar to the specific activity of the material produced by the original clone (2×10^6 units/mg of protein) at the same stage of purification.

Results

Homogeneity

Although the presence of other proteinaceous growth factors in preparations of interleukin 2 produced by recombinant DNA technology is unlikely, *E. coli* cellular constituents can have an effect on the immune system. *E. coli* outer membrane proteins are also extremely potent antigenic stimuli and lipopolysaccharide, the *E. coli* endotoxin, has specific effects on the immune system, e.g. the induction of interleukin 1 synthesis and the activation of B-cells. If recombinant lymphokine products are going to be useful reagents in immunology, it is important that the levels of these constituents be determined.

The interleukin 2 showed no extraneous protein bands on silver-stained SDS/polyacrylamide-gel electrophoresis (Fig. 1). The weak band, visible at high protein loading and running at an apparent M_r of 35 000, cross-reacts with a monoclonal antibody raised against natural interleukin 2 (Gearing *et al.*, 1985) and is therefore probably

Fig. 1. M_r of interleukin 2 by SDS/polyacrylamide-gel electrophoresis

A linear 12.5% acrylamide gel was used in conjunction with the Tris/glycine discontinuous buffer system of Laemmli (1970). The markers (left-hand) were lysozyme, soya bean trypsin inhibitor, carbonic anhydrase, ovalbumin, bovine serum albumin and phosphorylase. The other tracks show 2-fold dilutions in interleukin 2 concentration from 90 µg/slot in the first track. The gel was stained with silver.

covalent interleukin 2 dimer. The absence of protein contaminants comigrating with recombinant interleukin 2 was established by using a solid phase radioimmunoassay for *E. coli* protein (Table 1). Lipopolysaccharide was not detected either by elevation in body temperature (the rabbit pyrogen test) or by the formation of a clot in the *Limulus* amoebocyte lysate test.

M_r value

Discontinuous SDS/polyacrylamide-gel electrophoresis with samples reduced in 2-mercaptoethanol (Fig. 1) suggested a covalent M_r of approx. 14400 (comigration with lysozyme) in good agreement with the theoretical value of 15513, deduced from the amino acid sequence. Using a similar technique, both JURKAT and peripheral blood lymphocyte interleukin 2 (Welte *et al.*, 1982; Robb *et al.*, 1983a; Godard *et al.*, 1984; Knip *et al.*, 1984) have been shown to migrate with subunit M_r values of approx. 15000.

The monomeric nature of *E. coli*-derived interleukin 2 in aqueous solution can be deduced from an observed M_r of 15000 on calibrated gel-permeation chromatography (see the Materials and methods section). Also, sedimentation velocity experiments showed that in solution interleukin-2 exists as a slowly sedimenting species having an $s_{20,w}$ of 1.86S (see the Materials and methods section). This value is close to a theoretical value of 1.76S for an ideal hard sphere of M_r 15513.

Table 1. Solid phase radioimmunoassay for *E. coli* proteins

The recombinant interleukin 2 preparation and solutions containing known amounts of *E. coli* proteins were examined under standard solid phase radioimmunoassay conditions, as described in the Materials and methods section. The amounts of radioactively labelled rabbit anti-(*E. coli* proteins) IgG bound to the plate are listed (values are means $\pm$ S.E.M.).

Sample	Concentration (ng/ml)	IgG bound (c.p.m.)
Buffer alone	—	233.7 $\pm$ 16.1
<i>E. coli</i> protein	30	313.3 $\pm$ 9.6
	100	362.0 $\pm$ 23.9
	1000	1340.7 $\pm$ 38.7
	10000	5131.7 $\pm$ 162.4
Interleukin 2	500000	243.0 $\pm$ 37.3

Reported values for the M_r of the natural interleukin 2 preparations are 20000 for JURKAT (Robb, 1982) and 25000 for peripheral blood lymphocyte derived material.

Isoelectric point

Gel isoelectric focusing on the preparation suggested a value of 7.7 for the single major constituent species. A minor component (<10%) of pI 8.2 has also been observed in some batches. The theoretical amino acid composition of the mature interleukin 2 protein was used to calculate a theoretical isoelectric point of 7.9 using a simplistic additive model and perturbed pK_a values for the charged amino acid species derived from the literature (see the Materials and methods section). Isoelectric point determinations on natural interleukin 2 vary from 6.8 to 8.2 depending on the method of preparation and the source material. There is now good circumstantial evidence that this heterogeneity is due to differing amounts of terminal sialic acid on the carbohydrate moiety of the molecule. Robb (1982) working with tonsil lymphocyte derived interleukin 2 found three major biologically active species (pI values 6.8, 7.6 and 7.9) which were converted to a single species of pI 7.9 upon chemical removal of sialic acid (neuramidase treatment or mild acid hydrolysis).

Chemical structure

The observed amino acid composition (Table 2a) of acid-hydrolysed interleukin 2 was consistent with that predicted from the amino acid sequence (Fig. 4).

The N-terminal sequence of the *E. coli*-derived interleukin 2 was obtained by combined g.l.c.-m.s. Analysis was performed by the method of Rose *et*

Table 2. Amino acid composition of recombinant interleukin 2

Samples of interleukin 2 or interleukin 2-derived peptides were hydrolysed in 6M-HCl for 24, 48 and 72h. The quantitative recovery of amino acids was determined. Values for 48 and 72h were normalized on aspartic acid, glutamic acid, glycine, alanine, leucine and phenylalanine, and the 'best value' given is an average of all three values except for serine and threonine, which were extrapolated to zero hydrolysis time, and valine and isoleucine, which were extrapolated to infinite hydrolysis time.

(a) Analysis of whole protein

Amino acid	'Best value' of recovery (nmol)	Residues/mol	
		('Best value'/6.18)	Theoretical
Asx	74.78	12.1	12
Thr	77.70	12.6	13
Ser	45.0	7.3	8
Glx	111.89	18.1	18
Pro	27.27	4.4	5
Gly	14.0	2.2	2
Ala	30.7	4.9	5
Cys*	—	2.5	3
Val	26.0	4.2	4
Met	18.52	3.6	5
Ile	53.6	6.9	9
Leu	136.96	22.16	22
Tyr	17.07	2.8	3
Phe	36.25	5.9	6
His	19.43	3.14	3
Lys	68.92	11.1	11
Arg	24.48	4.0	4
Trp†	—	0.73	1

(b) Analysis of fraction 90 from the experiment described in Fig. 3(a)

Amino acid	Recovery from labelled peptide (nmol)	Residues/mol	Theoretical composition		
			Peptide 124–33	Peptide 98–120	Peptide 50–76
Cm-Cys	12.60	1.0	1	1	1
Asx	0.12	0.09	0	2	1
Thr	26.5	2.09	2	4	1
Ser	25.5	2.02	2	1	1
Glx	14.6	1.16	1	4	8
Pro	0.0		0	0	1
Gly	0.0156		0	0	1
Ala	0.0		0	2	2
Val	0.0		0	1	1
Met	0.0		0	1	0
Ile	20.35	1.61	2	1	0
Leu	13.54	1.2	1	1	7
Tyr	0.0		0	1	0
Phe	0.0		0	2	0
His	0.0		0	0	1
Lys	0.026		0	0	2
Arg	0.0		0	0	0

* Quantification by iodo[2-³H]acetic acid modification.
† Spectrophotometric determinations.

al. (1983a) (Fig. 2a) indicating an *N*-terminal sequence of:

The finding that two sets of sequence ions were in a 1:1 ratio proves that the *N*-terminal methion-

ine residue was not blocked. In *E. coli* the most commonly expected group blocking *N*-terminal methionine is formyl, translation usually being initiated by *N*-formylmethionine. Experiments with model compounds (results not shown) show that formylmethionine is not cleaved during the

Fig. 2. Mass spectra of peptides from interleukin 2

(a) Mass spectrum of the *N*-terminal peptide of recombinant interleukin-2. This spectrum is that of the peptide Met-Ala-Pro as its *N*^α-acetyl-*N,O*-permethyl derivative. In the Figure and legend, the conventional three-letter code is used to denote *N*^α-acyl-*N,O*-permethyl amino acid residues. Interpretation of the spectrum is straightforward and follows well-established rules (see Rose *et al.*, 1983b). Iminium composition ions, of general structure $H^+NCH_3:CHR$, where R represents the *N,O*-permethylated side chain of an amino acid, appearing at *m/z* 58, 70, 118 inform us of the presence of Ala, Pro and Met, respectively. The thioether side chain of Met is responsible for the strong signal at *m/z* 61. A series of sequence ions develops, commencing with *N*^α-acetyl-Met at *m/z* 188. This series is accompanied by a second series, 3 mass units heavier than the first, because a mixed acetyl/perdeuteroacetyl reagent was used. The presence of this second series permits us to deduce that the *N*-terminal amino group was free (and not naturally acetylated) prior to sample preparation. The molecular ion was too weak to be recorded. The strong pair of signals at *m/z* 327, 330 is due to a loss of 74 from the molecular ion of the tripeptide derivative, a typical fragmentation of methionyl peptides, and confirms the sequence assigned. (b) Mass spectrum of the *N*-terminal peptide of truncated interleukin-2. This spectrum is rather weak but clearly shows the presence of *N*-terminal serine (see legend to a). The iminium composition ion appears here at *m/z* 88 (no signal at *m/z* 91 since the structure of this ion does not contain the acetyl group). (c) Mass spectrum of the *C*-terminal peptide of recombinant interleukin-2. This spectrum is that of the peptide Thr-Ile-Thr as its *N*^α-trifluoroacetyl-*N,O*-permethyl derivative (see legend to a). The peptide was amongst those produced upon subtilisin digestion of recombinant human interleukin 2 in a medium containing approx. 50 atom% $H_2^{18}O$. It was the only peptide not found to have incorporated the stable isotope into its *C*-terminal carboxyl group and so must be the *C*-terminal peptide of the protein (Rose *et al.*, 1983b). This is the raw mass spectrum obtained as scan 179 of the g.l.c.-m.s. run. No background subtraction or signal averaging has been performed. Signals at *m/z* 59 and 100 arise from the side chain of Thr and the iminium composition ion of Ile, respectively. A series of sequence ions develops, commencing at *m/z* 226 (*N*^α-trifluoroacetyl-Thr), which permits us to deduce the sequence Thr-Ile-Thr. The amino acid sequence predicted from the DNA terminates Thr-Ile-Thr, in agreement with the mass spectrometric result.

formation of acetylated derivatives and is permethylated in the normal way. If one assumes that the initiator methionine is deformylated during synthesis but not cleaved from the chain, then

these results are in accord with the DNA sequence of the plasmid vector (Devos *et al.*, 1983).

Natural JURKAT interleukin 2 (Robb *et al.*, 1983a; Stern *et al.*, 1984) has been shown to have

the following sequence by automated gas phase Edman degradation:

NH₂-Ala-Pro-Xaa-Ser-Ser. . .

The DNA sequence of human genomic clones of interleukin 2 (Degraeve *et al.*, 1983) indicates that *in vivo* the protein is translated as a precursor 20 residues longer at the *N*-terminus than the secreted extracellular species. This hydrophobic signal peptide is specifically cleaved between Ser-19 and Ala-20 of the preprotein. As *E. coli* cannot process mammalian signal peptides efficiently, the original gene sequence was engineered (Devos *et al.*, 1983) so that an initiator methionine (ATG) codon lay immediately before the first alanine codon of the mature natural interleukin 2 coding sequence. *E. coli* does possess an enzyme which removes formyl groups from initiator methionyl *N*-termini (Adams, 1968) and a little studied enzyme of wide specificity for removing the *N*-terminal methionine from newly synthesized proteins. However, mammalian proteins produced in *E. coli* by recombinant means need not necessarily be recognized by these processing enzymes, particularly in such cases as interleukin 2, where the natural *N*-terminal sequence (signal peptide) has already been removed by genetic engineering.

In order to investigate further the specificity of this processing enzyme system, a deletion mutant of recombinant interleukin 2 lacking the *N*-terminal Ala-Pro-Thr coding sequence was purified and analysed by the g.l.c.-m.s. method (Fig. 2b). In this case we found that the *N*-terminal amino acid was serine, the initiator methionine having been removed by the *E. coli* ribosomal methionine aminopeptidase.

In natural interleukin 2, Thr-3 is linked via an *O*-glycosidic bond to *N*-acetyl-D-galactosamine (Robb *et al.*, 1983b) and hence has never been identified as phenylthiohydantoin-threonine after gas phase sequencer analysis. Recombinant interleukin 2 is not a glycoprotein. Samples of purified protein (60 and 300 nmol) were analysed for total carbohydrate by the anthrone method with negative results.

The C-terminal peptide of recombinant interleukin 2 was identified and sequenced by the g.l.c.-m.s. method (Fig. 2c). The sequence was:

. . .-Thr-Ilx-Thr-CO₂H

in agreement with that predicted from the DNA sequence of the gene and as observed by carboxypeptidase A digestion of the natural mature protein (Clark *et al.*, 1984).

S-Carboxymethylation with iodo[2-³H]acetic acid in guanidinium chloride indicated that interleukin 2 has a single free thiol residue in the sequence. Alkylation of recombinant interleukin 2

reduced in guanidinium chloride and dithiothreitol led to the incorporation of 2.5 mol of iodo[2-³H]acetic acid/mol of protein (Table 2a). Unreduced recombinant interleukin 2 incorporated 1.0 mol of iodo[2-³H]acetic acid/mol of protein. This reactive thiol residue was identified as Cys-125 in the sequence by tryptic peptide mapping (Fig. 3a) and amino acid analysis of the labelled peptide. As shown in Table 2(b), the observed amino acid composition of the labelled peptide is clearly related to the theoretical composition of pseudotryptic cleavage peptide 124-133 and cannot be reconciled with the theoretical compositions of the other possible cysteine-containing peptide sequences. The two other cysteine residues in the protein sequence, Cys-58 and Cys-105, were shown to be involved in a disulphide bridge by reduction prior to alkylation with labelled iodoacetic acid and mapping the resultant tryptic peptides (Fig. 3b).

The presence of a disulphide bridge has been suggested indirectly in JURKAT interleukin 2 (Robb *et al.*, 1983a). This natural protein preparation was seen to change mobility on SDS/polyacrylamide-gel electrophoresis after reduction and also to incorporate more radioactivity on alkylation with labelled iodoacetic acid. Wang *et al.* (1984), however, report that their human recombinant interleukin 2 preparation, although fully active, consisted of a mixture of all three possible disulphide bridges.

Physical structure

Chou & Fasman's (1974) method of secondary structure prediction on the interleukin 2 mature protein sequence (results not shown) shows that the protein should be predominantly α -helical in nature. This prediction is confirmed by the c.d. spectrum which is dominated by an α -helical component. An approximate calculation of the percentage helix, using a value of -3.6×10^4 degree⁻¹·cm⁻¹·dmol (Gratzer, 1978) for the molar residue ellipticity of α -helix gave a value of 31% α -helix in interleukin 2.

Biological activity

The ability of *E. coli*-derived interleukin 2 to promote the proliferation of a murine cytotoxic T-cell line (CTLL-2) was used as the basis of a biological assay (Gillis *et al.*, 1978). A unit of biological activity is defined as the reciprocal of the dilution needed to bring about a 50% stimulation of [³H]thymidine incorporation. The observed specific activity of *E. coli*-derived interleukin 2 was 1×10^7 units/mg. Published specific activities for purified natural interleukin 2 using the thymidine incorporation methodology range from 6.4×10^3

Fig. 3. Identification of cysteine-containing tryptic peptides

See the text for experimental methods. Radiolabelled recombinant interleukin 2 was digested with trypsin and the resulting tryptic peptides were loaded onto a C-18 reversed-phase analytical h.p.l.c. column. The column was eluted with a linear gradient of acetonitrile (---) and eluate fractions of 0.7 ml were collected. Each fraction was monitored for ${}^3\text{H}$ radioactivity (—) and A_{214} (.....). (a) Pattern of labelling after modification of the protein with iodo[2- ${}^3\text{H}$]acetic acid in 8M-guanidinium chloride. (b) As in (a) except the protein was reduced with 1 mM-dithiothreitol prior to the addition of iodo[2- ${}^3\text{H}$]acetic acid.

```

0      10      20
M A P T S S S T K K T Q L Q L E H L L L D

      30      40
L Q M I L N G I N N Y K N P K L T R M L

      50      60
T F K F Y M P K K A T E L K H L Q C L E

      70      80
E E L K P L E E V L N L A Q S K N F H L

      90     100
R P R D L I S N I N V I V L E L K G S E

      110     120
T I F M C E Y A D E T A T I V E F L N R

      130     133
W I T F C Q S I I S T L T

```

Fig. 4. Amino acid sequence of mature human interleukin 2. The sequence is taken from Degrove *et al.* (1983) and is written in the one-letter code.

(Godard *et al.*, 1984) to 1×10^9 units/mg (Stern *et al.*, 1984). This variation is due to differences in (a) the cell type used in the assay and (b) the correction factors and reference standards used in the calculation. The same assay using the CTLL-2 cell line gives a specific activity for natural interleukin 2 of a similar order of magnitude (e.g. Robb *et al.*, 1983a). As the potency of a natural interleukin-2 reference standard preparation [National Cancer Institute's Biological Response Modifiers Program (specific activity claimed 1.3×10^7 units/mg)] is of a similar level to that of the *E. coli*-derived material, the synthesis of interleukin-2 in an unnatural microbial host does not affect the biological activity of the protein.

Discussion

The chemical structure and biological potency of *E. coli*-derived interleukin 2 are similar to those of natural interleukin 2. Both are low- M_r monomeric proteins whose isoelectric point and secondary structural properties are consistent with the amino acid sequence. Differences in chemical structure are due to differing mechanisms of post-translational modification, i.e. glycosylation at Thr-3 in the human-derived protein and the presence of initiator methionine in the *E. coli*-derived molecule.

The finding that the initiator methionine was not cleaved during synthesis is remarkable. The processing of the sequence *N*-formyl-Met-Ala-... is known to occur during the synthesis of *E. coli* proteins, over 30 having *N*-terminal alanine (Dayhof, 1983). The apparent absence of this activity during the synthesis of *E. coli* interleukin 2 could be due to (a) the sequestration of the substrate in an inclusion body, (b) a high rate of synthesis,

saturating the processing enzyme or (c) steric hinderance of this processing enzyme by either an unfavourable global conformation of the interleukin 2 molecule or the local conformation of the *N*-terminal region. As the *E. coli*-derived interleukin 2 mutant lacking three *N*-terminal amino acid residues is a good substrate for the methionine-removing enzyme, arguments based on rates of synthesis, compartmentalization in inclusion bodies and overall protein conformation can be eliminated. What element of the *N*-terminal sequence of interleukin 2 makes it a poor substrate for the processing enzyme? The National Biomedical Research Foundation data base (Dayhof, 1983) lists 30 *E. coli* protein sequences containing *N*-terminal alanine, only one protein with the sequence NH₂-Xaa-Pro-Xaa... and five with the sequence NH₂-Xaa-Xaa-Thr... (where Xaa is any amino acid except methionine). It is possible therefore that the low frequency of proline in this sequence position reflects the creation of a local conformation preventing productive binding to the methionine amino peptidase. This hypothesis is strengthened by the finding that the human growth hormone gene is expressed in *E. coli* as a protein with a sequence NH₂-Met-Phe-Pro-Thr... and is therefore not processed (Olsen *et al.*, 1981).

No evidence has emerged that either the presence of an extra methionine or the absence of glycosylation modify either the biological activity *in vitro* or the physicochemical properties of *E. coli*-derived interleukin 2. The physiological significance of interleukin 2 glycosylation is still unclear. Experiments with semisynthetic peptides suggest that the half-life of a protein or peptide *in vivo* may be extended by glycosylation (Mencke & World, 1982).

In the case of interleukin-2, this postulated role for the carbohydrate on some glycoproteins will be tested during the analysis of pharmacokinetic data as both natural and recombinant interleukin 2 are developed for therapeutic uses. The National Cancer Institute Biological Response Modifiers Program have compared the biological properties of this and other recombinant interleukin 2 preparations (Thurman *et al.*, 1985).

We would like to thank Dr. J.-P. Piguet, University of Geneva, for the gift of the CTLL-2 cell line, J. Gil, R. Gaffner, D. Williams, E. Magnenat and M. Barnes for skilful technical assistance, Heather Carr for performing the carbohydrate analysis and Marco Simona for preparing the derivatives for sequence analysis. We also acknowledge J. Rosa and J. Smart for help with the peptide mapping experiments and the help of Bernard Schwendimann and the University of Geneva for the circular dichroism measurements. We also like to thank Nadine Huber and Marie-Christine Vuargnier for typing the manuscript.

References

- Ackers, G. (1970) *Adv. Protein Chem.* **24**, 343–446
- Adams, J. M. (1968) *J. Mol. Biol.* **33**, 571–589
- Cantor, C. R. & Schimmel, P. R. (1980) in *Biophysical Chemistry*, p. 49, W. M. Freeman and Co, San Francisco
- Cohn, E. J. & Edsall, J. T. (1943) *Proteins, Amino Acids and Peptides*, 1st edn., pp. 374–378, Hafner Publishing Corp., New York
- Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 222–245
- Clark, S. C., Arya, S. K., Wong-Staal, F., Matsumoto-Kobayashi, M., Kay, R. M., Kaufman, R. J., Brown, E. L., Shoemaker, C., Copeland, T., Oroszlan, S., Smith, K., Sarnagadharan, M. G., Linder, S. G. & Gallo, R. C. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2543–2547
- Dayhof, M. O., *Protein Sequence Data Base: Computer System NBR Rep.* 08710-81001, National Biomedical Research Foundation, Washington DC
- Degrave, W., Tavernier, J., Duerinck, F., Plaetinck, G., Devos, R. & Fiers, W. (1983) *EMBO J.* **2**, 2349–2353
- Devos, R., Plaetinck, G., Cheroute, M., Simons, G., Degrave, W., Tavernier, J., Remaut, E. & Fiers, W. (1983) *Nucleic Acids Res.* **11**, 4307–4323
- Farrar, J. J., Benjamin, W. R., Hilfiker, M. L., Howard, M., Farrar, W. L. & Fuller-Farrar, J. (1982) *Immunol. Res.* **63**, 129–165
- Gearing, A., Thorpe, R., Bird, C. & Spitz, M. (1985) *Immunol. Lett.*, in the press
- Gillis, S., Fern, M. M., Ou, W. & Smith, K. A. (1978) *J. Immunol.* **120**, 2027–2032
- Godard, A., Naulet, J., Peyrat, M. A., Vie, H., Moreau, J. F., Bignon, J. D. & Soullilou, J. P. (1984) *J. Immunol. Methods* **70**, 233–244
- Gratzer, W. (1978) in *Techniques in the Life Sciences* (Kornberg, H. L., Metcalfe, J. C., Northcote, D. H., Pogson, C. I. & Tipton, K. F., eds.), vol. B1, Elsevier/North-Holland
- Grimm, E. A., Mazumder, A. Zhang, H. Z. & Rosenberg, S. A. (1982) *J. Exp. Med.* **155**, 1823–1841
- Inaba, K., Ganelli-Piperno, A. & Steinman, R. M. (1983) *J. Exp. Med.* **158**, 2040–2057
- Kniep, E. M., Kniep, B., Grote, W., Conradt, H. S., Monner, D. A. & Mühlradt, P. F. (1984) *Eur. J. Biochem.* **143**, 199–203
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Mencke, A. J. & World, F. (1982) *J. Biol. Chem.* **257**, 14799–14805
- Morgan, D. A., Ruscetti, F. W. & Gallo, R. (1976) *Science* **193**, 1007–1008
- Olsen, K. C., Fenno, J., Lin, N., Harkins, R. N., Snider, C., Kohr, W. H., Ross, M. J., Fodge, D., Prender, G. & Stebbing, N. (1981) *Nature (London)* **293**, 408–411
- Ortaldo, J. R., Mason, A. T., Gerard, J. P., Henderson, L. E., Farrar, W., Hopkino, R. F., Herkerman, R. B. & Robin, H. (1984) *J. Immunol.* **133**, 779–783
- Robb, R. J. (1982) *Immunobiology* **161**, 21–50
- Robb, R. J., Kutny, R. M. & Chowdry, V. (1983a) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5990–5994
- Robb, R. J., Kutny, R. M., Panico, M., Morris, H., De Grado, W. F. & Chowdny, V. (1983b) *Biochem. Biophys. Res. Commun.* **116**, 1049–1055
- Roe, J. M. (1955) *J. Biol. Chem.* **212**, 335–339
- Rose, K., Simona, M. G. & Offord, R. E. (1983a) *Biochem. J.* **215**, 261–272
- Rose, K., Simona, M. G., Offord, R. E., Prior, C., Otto, B. & Thatcher, D. R. (1983b) *Biochem. J.* **215**, 273–277
- Stern, A., Pan, Y. C., Urdal, D. L., Mochizuki, D. Y., De Chira, S., Blacker, R., Wideman, J. & Gillis, S. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 871–875
- Taniguchi, T., Matsui, M., Fujita, T., Takasha, C., Kashima, N., Yoshimoto, R. & Hanubo, J. (1983) *Nature (London)* **302**, 305–309
- Thurman, G. B., Maluish, A. E., Rossio, J. L., Schlick, E., Onozaki, K., Talmadge, J. E., Procopio, A. D., Ortaldo, J. R., Ruscetti, F. W., Stevenson, H. C., Cannon, G. B., Iyer, S. & Herberman, R. B. (1985) *J. Clin. Immunol.*, in the press
- Wang, A., Lu, S. D. & Mark, D. F. (1984) *Science* **224**, 1431–1433
- Welte, K., Wang C. Y., Mertelomann, R., Salvatore, V., Feldman, S. P. & Moore, M. A. S. (1982) *J. Exp. Med.* **156**, 454–464
- Wray, W., Bouliskas, T., Wray, V. P. & Hancock, R. (1981) *Anal. Biochem.* **118**, 197–203