Biochem. J. (1992) 286, 825–828 (Printed in Great Britain)

Human interleukin-5 expressed in *Escherichia coli* has *N*-terminal modifications

Keith ROSE,*§ Pierre-Olivier REGAMEY,* Robert ANDEREGG,† Timothy N. C. WELLS‡ and Amanda E. I. PROUDFOOT‡

*Département de Biochimie Médicale, Centre Médical Universitaire, 1 rue Michel Servet, 1211 Geneva 4, Switzerland, †Glaxo Inc., 5 Moore Drive, Research Triangle Park, NC 27709, U.S.A., and ‡Glaxo Institute for Molecular Biology, 14 Chemin des Aulx, 1228 Plan-les-Ouates, Geneva, Switzerland

Recombinant human interleukin-5 exists as four major isoforms all possessing *N*-terminal methionine. Peptide mapping and subsequent analysis by fast-atom-bombardment mass spectrometry (f.a.b.-m.s.) have shown that *N*-terminal modifications are the cause of the charge heterogeneity. In order of decreasing abundance, these are unmodified methionine, retention of *N*-terminal formyl group, oxidation of *N*-terminal methionine to sulphoxide and carbamoylation of the *N*-terminus. These results were confirmed by analysis of the reduced and alkylated intact protein by electrospray-ionization mass spectrometry. The implications of these findings for the production and characterization of recombinant proteins are briefly discussed.

INTRODUCTION

Interleukin-5 (IL-5) is the haemopoietic factor which causes eosinophil differentiation. Only this activity has been described for the human protein, whereas the mouse form also has B-cell growth-factor activity [1]. Its low abundance has meant that the natural form has only been purified from mice. The mouse protein is a glycosylated disulphide-linked homodimer of molecular mass 46–60 kDa [2,3]. The human gene was cloned by using mouse IL-5 cDNA as a probe [4]. The human protein has been obtained by expression in *Escherichia coli* [5] and in eukaryotic expression systems [6,7]. In *E. coli*, the protein is produced in inclusion bodies, and can be re-folded, giving a fully biologically active protein with a molecular mass of 25 kDa [5,8]. Recombinant human IL5 (rhIL-5) in all systems results in disulphide-linked homodimeric topology, as is observed for the natural mouse protein.

Previous work has shown that IL-5 produced in both *E. coli* and mammalian expression systems exists in several isoforms, and shows charge heterogeneity in both glycosylated and non-glycosylated forms [9]. In the present paper, we show that the observed heterogeneity of the *E. coli* produced protein can be explained by *N*-terminal modifications.

MATERIALS AND METHODS

Unless otherwise stated, all solvents and reagents were of analytical grade or better, were obtained from commercial sources and were used without further purification. All reactions were carried out at room temperature (22 °C) unless otherwise stated.

Reduction and carboxymethylation

rhIL-5 was isolated, renatured, purified and analysed by SDS/PAGE and isoelectric focusing (IEF) as previously described [5]. Approx. 3 mg of protein was reduced in 4.6 ml of a solution containing 4 M-guanidine hydrochloride, 5 mM-EDTA, 5 mM-dithiothreitol and 0.1 M-Tris/HCl, pH 7.4. After 1 h at room temperature, iodoacetate was added from a 1 M stock at pH 4.8 to a final concentration of 30 mM. The solution was incubated for 30 min, and dithiothreitol was added to a final concentration of 25 mM. The sample was dialysed extensively at 4 °C against 125 mM-NH_4HCO_3.

Enzymic digestion

Reduced and carboxymethylated rhIL-5 was diluted to 30 ml with 125 mM-NH_4HCO_3, and 30 μl of a freshly prepared solution of bovine trypsin (treated with tosylphenylalanylchloromethane; Worthington; 1 mg/ml in water) was added. After 2 h at 37 °C, a further 30 μl of trypsin was added and the incubation was continued for a further 2 h. The sample was freeze-dried and then redissolved in 3 ml of 0.1% trifluoroacetic acid (TFA)/5% (v/v) acetic acid. The small quantity of insoluble material was removed by centrifugation.

H.p.l.c.

Peptides were separated on a Waters h.p.l.c. system, monitored at 214 nm and 280 nm. Solvent A was prepared by adding 1 g of TFA (h.p.l.c. grade; Applied Biosystems) to 1000 ml of h.p.l.c.-grade water (Milli-Q system), followed by vacuum filtration. Solvent B was prepared by adding 1 g of TFA to 100 ml of h.p.l.c.-grade water (previously degassed by vacuum filtration) and then adding 900 ml of acetonitrile (gradient grade, Merck; previously degassed by vacuum filtration). A column (250 mm × 4 mm internal diam.) of Nucleosil 300-A 5 μm C_8 (Machery Nagel) was used at a flow rate of 0.6 ml/min. Samples were injected into the column equilibrated with solvent A. After 5 min, a linear gradient (1%/min) of solvent B was applied for 60 min. The column was then washed with 100% B for 5 min before re-equilibration with solvent A. Peptides were collected manually at the detector outlet and dried without heat in a vacuum centrifuge (Speedvac, Savant).

Abbreviations used: IL-5, interleukin-5; rhIL-5, recombinant human IL-5; f.a.b.-m.s., fast-atom-bombardment mass spectrometry; IEF, isoelectric focusing; TFA, trifluoroacetic acid.

§ To whom correspondence should be addressed.

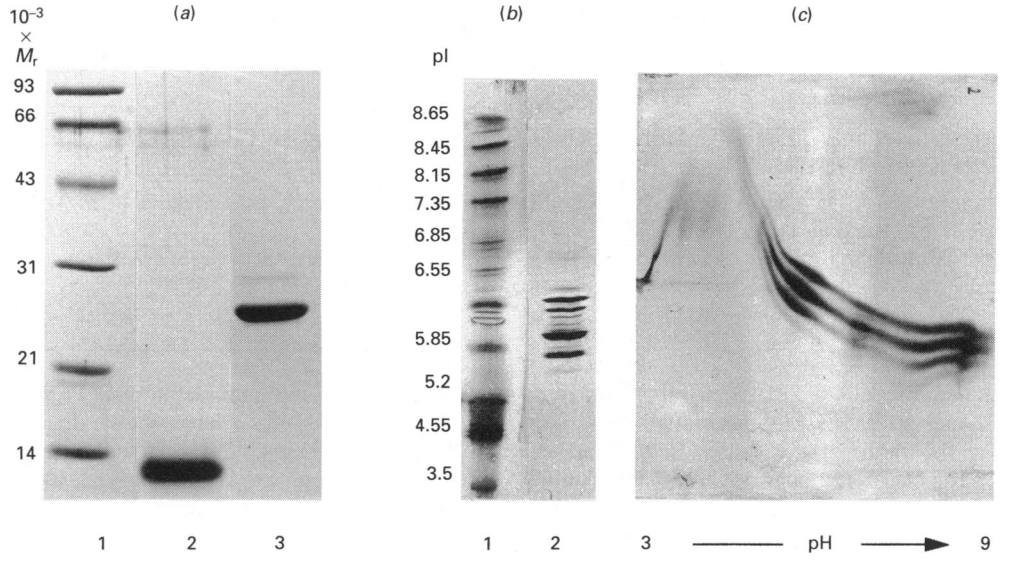

Fig. 1. (*a*) SDS/PAGE of rhIL-5; (*b*) polyacrylamide IEF gel; (*c*) pH-titration curve of rhIL-5

(*a*) Lane 1, standards (M_r indicated); lane 2, rhIL-5 in presence of reductant; lane 3, rhIL-5 in absence of reductant. (*b*) Lane 1, pI standards (pH values indicated); lane 2, rhIL-5.

Carbamoylation with potassium cyanate

Approx. 2 μg of the *N*-terminal peptide was dissolved in 50 μl of 8 M-urea containing 1 μl of acetic acid. The peptide was carbamoylated by addition of 5 μl of 1 M-potassium cyanate and incubation for 30 min at room temperature. The sample was immediately diluted with 100 μl of 0.1% TFA in water and analysed by h.p.l.c.

Deformylation in acid

Approx. 2 μg of peptide was dissolved in 100 μl of 0.1% TFA in water and incubated at 37 °C for 24 h before injection into the h.p.l.c. apparatus.

Fast-atom bombardment mass spectrometry (f.a.b.-m.s.)

F.a.b.-m.s. was performed on an MS 50 S mass spectrometer (Kratos Analytical, Manchester, U.K.) with a full f.a.b. retro-fit (MScan Ltd.) and operated under control of a DS 55 M data system (Kratos). The mass-spectrometric resolution was about 3000, with an accelerating voltage of 4 kV and a scan rate of 30 s/decade. The f.a.b. gun was operated with xenon at 9 kV and 7 μA. External calibration of the mass scale was achieved with a mixture of the iodides of sodium and caesium (3:1 molar ratio). For analyses in positive-ion mode, the stainless-steel target was loaded with 1 μl of a mixture of glycerol and 1-thioglycerol (2:1, v/v), followed by 0.5 μl of acetic acid. Samples were taken up in about 20 μl of 50% acetonitrile containing 0.1% TFA, and about 1 μl was applied to the loaded target.

Electrospray m.s.

Electrospray (ion-spray) mass spectra were collected with a Sciex API-III mass spectrometer. Protein was reduced and pyridylethylated at cysteine [10], and was dissolved in aq. 50% (v/v) methanol containing 2% acetic acid. The solution was infused at a rate of 2 μl/min into the ion-spray needle, which was maintained at 5000 V. The orifice voltage was 120 V. Spectra were co-added by repetitive scanning until an acceptable signal/noise level was obtained (usually 3–4 min).

RESULTS AND DISCUSSION

The rhIL-5 used in this study gives a single band on SDS/PAGE. However, when analysed by IEF and pH-titration curves, it is clear that there are several isoforms (Fig. 1). Edman degradation shows there is a single *N*-terminal sequence, and *C*-terminal analysis by m.s. shows a single *C*-terminal sequence [5]. The amino acid analysis was in good agreement with the composition predicted from the DNA sequence and confirms that the initiating methionine is fully retained.

The convergence of the pH-titration curves (Fig. 1*c*) towards pH 4 suggested that the charge heterogeneity may be caused by deamidation. In order to investigate this possibility, we attempted to isolate peptides of a convenient mass for f.a.b.-m.s., covering all the glutamine and asparagine residues in rhIL-5.

The reduced carboxymethylated protein was digested with trypsin, and the peptides were separated by h.p.l.c. and analysed by f.a.b.s.-m.s. Fig. 2 shows an analytical chromatogram of the tryptic digest. All the glutamine and asparagine residues were mapped, and none was found to be extensively deamidated.

Three of the tryptic peptides, T7, T5 and T10, were found to correspond to the *N*-terminus. Fig. 3(*a*) shows the partial mass spectrum of peak T7. m/z 1189 corresponds to the expected protonated molecular ion for residues 1–11 (MTEIPTSALVK) in an unmodified form, and this assignment was confirmed by the presence of sequence ions at m/z 1058, 957, 828 and 715, corresponding to successive losses from the *N*-terminus of methionine, threonine, glutamate and isoleucine respectively. Fig. 3(*b*) shows the partial mass spectrum of peak T5, showing m/z of 1205. This addition of 16 relative to peptide T7 suggests an oxidized form of the *N*-terminal residues 1–11, an interpretation confirmed by the first sequence ion at m/z 1058, which corresponds to loss of methionine sulphoxide. Although sulphoxide formation occurs readily, it is unlikely to have occurred during peptide isolation, since peak T7 showed no readiness to oxidize spontaneously. Fig. 3(*c*) shows the partial mass spectrum of peak T10. Once again, the sequence ions at m/z 1058, 957, 828 and 715 confirmed that T11 is derived from the *N*-terminal residues 1–11. However, protonated molecular ions appear at m/z 1217 and 1232. T10 therefore contains two peptides which

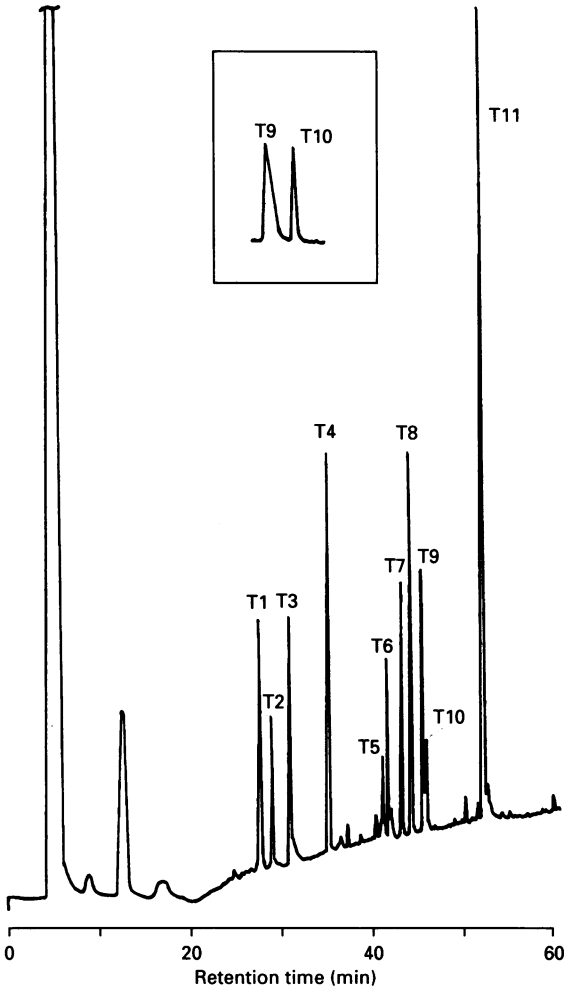

Fig. 2. Analysis of the tryptic digest of rhIL-5 by reversed-phase h.p.l.c.

Conditions were as described in the Materials and methods section. The A_{214} (0.2 AUFS) is shown. The insert shows the baseline separation obtained between peaks T9 and T10 during preparative chromatography with a shallower gradient. The broad peaks eluted before peak T1 did not give an interpretable mass spectrum.

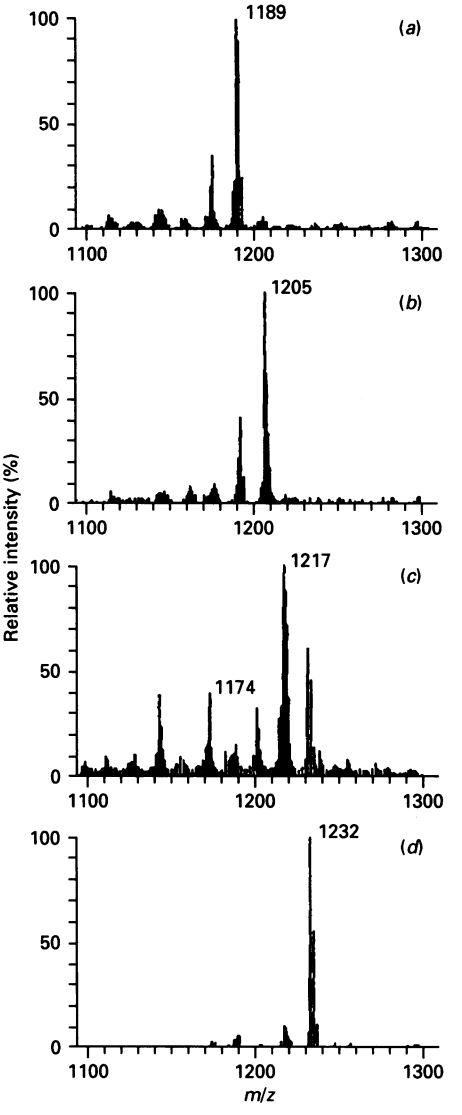

Fig. 3. Mass spectra of forms of the N-terminal tryptic peptide

Spectra were obtained by f.a.b.-m.s. in positive-ion mode as described in the Materials and methods section. (*a*) Peptide T7; (*b*) peptide T5; (*c*) peptide T10; (*d*) peptide T7 after modification with potassium cyanate.

are 28 and 43 mass units heavier than the unmodified *N*-terminal peptide. The groups responsible for these mass shifts of 28 and 43 units are most likely to be formyl and carbamoyl respectively. To test these hypotheses, 2 μg of peak T10 was treated with acid to deformylate a portion of the material. In addition, 2 μg of peak 7 was treated with potassium cyanate to transform a portion to the carbamoyl form. During mild acid treatment of T10, a small amount of material was formed which was eluted with the same retention time as T7 and which had a similar mass spectrum (results not shown). This is consistent with the expected deacylation of the formylated peptide. During treatment of T7 with potassium cyanate, a reagent known to carbamoylate amino groups [11], a species is formed which is eluted at the same retention time as T10, and which has the mass spectrum of the carbamoylated peptide (Fig. 3*d*). It is not particularly surprising that formyl and carbamoyl peptides should co-elute, under the conditions used, at a retention time later than that of the unmodified peptide: both have lost the charge of the *N*-terminal amino group, and otherwise differ only slightly. Carbamoylation of the protein probably occurred during treatment with concentrated solutions of urea [5].

To confirm our findings, we analysed the intact protein, after reduction and alkylation, by electrospray-ionization m.s. The electrospray mass spectrum is shown in Fig. 4, from which we may determine the mass of the alkylated protein. The experimentally determined value of 13282.5 $\pm$ 2 is very close to that calculated (13280.4) for monomeric IL-5, assuming that the only modification to the protein is the alkylation. The spectrum shows evidence of heterogeneity; the signals have several discrete shoulders on the high-mass/charge side (see insert to Fig. 4). Assuming that these shoulders are the result of modified IL-5 species, it is possible to measure approximate masses for each species. They are 13280.8 for the main signal, and 13292.4, 13309.4 and 13322.4 for the first, second and third shoulders respectively. The predicted masses for the unmodified *N*-terminus, sulphoxide *N*-terminus, formyl *N*-terminus and carbamoyl *N*-terminus are 13280.4, 13296.4, 13308.4 and 13324.4 respectively. It is not surprising that the main signal gives the closest correspondence between observed and expected values. On their own, the electrospray data are not sufficient to characterize rhIL-5 completely: mass agreement is not exact, and it is not possible from the electrospray data alone to attribute the signal at m/z 1025.8 (13324.4 with 13 protons) to the carbamoyl form (13323.4) rather than a form possessing both a formyl group and a sulphoxide (13324.4). However, taken together with

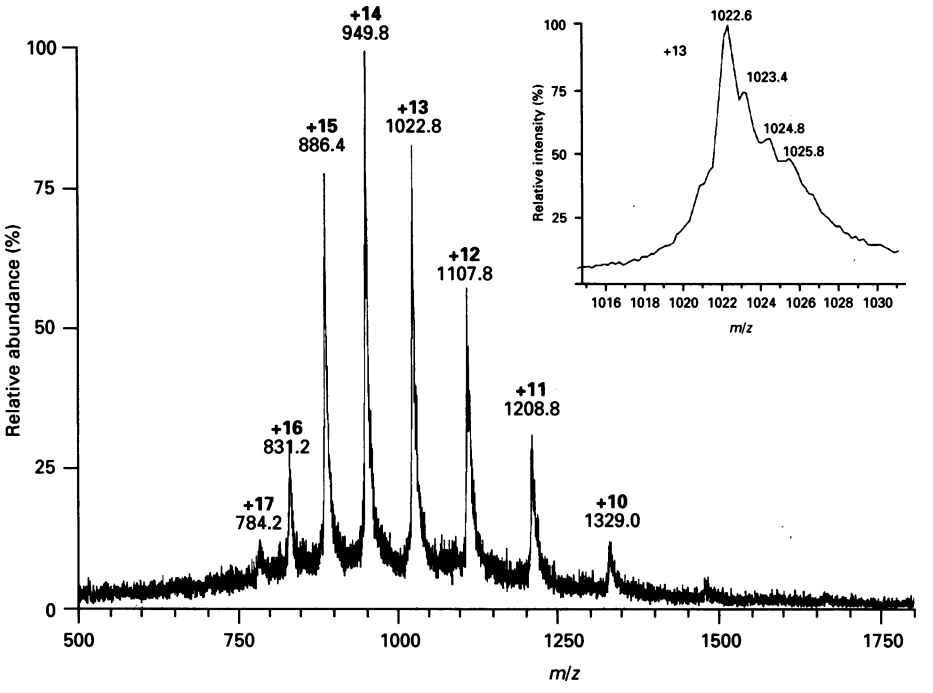

Fig. 4. Electrospray ionization mass spectrum of reduced and alkylated rhIL-5

The insert shows, with expanded scale, the signal with 13 charges. The charge state of each ion group is indicated in **bold** type above the corresponding *m/z* values.

the peptide mapping, acid and potassium cyanate-treatment results, and f.a.b.-m.s. data, it is clear that the charge isoforms are due to acylation of the *N*-terminal methionine. Partial acylation of the *N*-terminal residue by a small group such as formyl or carbamoyl would lead to three bands upon IEF of the dimeric protein. There would be a species (of lowest pI) in which both chains were acylated, a species (intermediate pI) in which only one of the chains was acylated and a species (highest pI) having both *N*-terminal amino groups free. It is possible, but not proven, that this species of highest pI suffers band splitting on IEF owing to sulphoxide formation, which also affects the pI.

The small mass increment (28 for formyl, 43 for carbamoyl) is much too small to permit resolution by SDS/PAGE, but is perfectly accessible to mass-spectrometric methods. The initiating formylmethionine residue is usually removed when exogenous proteins are expressed in *E. coli*. It is relatively rare, but not unknown [12], to find a significant amount of formylmethionyl-protein.

The production of rare proteins by recombinant technology makes feasible the determination of their structure in solution by n.m.r. It is important to identify small modifications such as those described here in order to facilitate the interpretation of the data obtained.

We thank the Fonds National Suisse de la Recherche Scientifique and the Schmidheiny Foundation for grants to purchase and maintain some of the mass-spectrometric and h.p.l.c. equipment used.

Received 5 February 1992/16 March 1992; accepted 30 March 1992

REFERENCES

1. Clutterbuck, E., Shields, J. G., Gordon, J., Smith, S. H., Boyd, A., Callard, R. E., Campbell, H. D., Young, I. G. & Sanderson, C. J. (1987) Eur. J. Immunol. **17**, 1743–1750
2. Takatsu, K., Harada, N., Hara, Y., Takahama, Y., Yamada, G., Dobashi, K. & Hamoaka, T. (1985) J. Immunol. **134**, 382–389
3. Mackenzie, D. T., Filutowitz, H. I., Swain, S. L. & Dutton, R. W. (1987) J. Immunol. **139**, 2661–2668
4. Azuma, C., Tanabe, T., Konishi, M., Kinashi, T., Noma, T., Matsuda, F., Yaoita, Y., Takatsu, K., Hammarström, L., Smith, C. I. E., Severinson, E. & Honjo, T. (1986) Nucleic Acids Res. **14**, 9149–9158
5. Proudfoot, A. E. I., Fattah, D., Kawashima, E., Bernard, A. & Wingfield, P. T. (1990) Biochem. J. **270**, 357–361
6. Tavernier, J., Devos, R., Van der Heyden, J., Hauquier, G., Bauden, R., Fache, I., Kawashima, E., Vandekerckhove, J., Contreras, R. & Fiers, W. (1989) DNA **8**, 491–501
7. Tsujimoto, M., Adachi, H., Kodama, S., Tsuruoka, N., Yamada, Y., Tanaka, S., Mita, S. & Takatsu, K. (1989) J. Biochem. (Tokyo) **106**, 23–28
8. Fattah, D., Quint, D. J., Proudfoot, A., O'Malley, R., Zanders, E. D. & Champion, B. R. (1990) Cytokine **2**, 112–121
9. Tominaga, A., Takahachi, T., Kiruchi, Y., Mita, S., Naomi, S., Harada, N., Yamaguchi, N. & Takatsu, K. (1990) J. Immunol. **144**, 1345–1352
10. Friedman, M., Krull, L. H. & Cavins, J. F. (1970) J. Biol. Chem. **245**, 3868–3871
11. Stark, G. R. (1967) Methods. Enzymol. **11**, 590–594
12. Tsunawawa, S., Yutari, K., Ogasarahara, K., Taketari, M., Yasouka, N., Kakudo, M. & Sugino, Y. (1983) Agric. Biol. Chem. **47**, 1393–1395