Biochem. J. (1983) **215**, 273–277
Printed in Great Britain

A new mass-spectrometric *C*-terminal sequencing technique finds a similarity between γ-interferon and α_2-interferon and identifies a proteolytically clipped γ-interferon that retains full antiviral activity

Keith ROSE,* Marco G. SIMONA,* Robin E. OFFORD,* Christopher P. PRIOR,† Berndt OTTO† and David R. THATCHER†
**Département de Biochimie Médicale, Centre Médical Universitaire, 9 Avenue de Champel, 1211 Geneva 4, Switzerland, and †Biogen S.A., 3 Route de Troinex, 1227 Carouge, Geneva, Switzerland*

(*Received 13 June 1983/Accepted 16 August 1983*)

A novel mass-spectrometric technique is described that permits the identification of the *C*-terminal peptide of a protein. The technique involves the incorporation of ^{18}O into all α-carboxy groups liberated during enzyme-catalysed partial hydrolysis of the protein, followed by mass spectrometry to identify as the *C*-terminal peptide the only peptide that did not incorporate any ^{18}O. The technique has been used to identify the true *C*-terminal tryptic peptide of a bacterially produced γ-interferon and to distinguish it from a peptide produced by anomalous tryptic cleavage. It was found that a closely similar sequence segment of bacterially produced α_2-interferon undergoes an analogous cleavage. The technique was also used to identify the *C*-terminus of a clipped γ-interferon that retains full antiviral activity.

Mass spectrometry offers several advantages over conventional methods for determining the amino acid sequence of polypeptides and proteins (Morris, 1980; Rose *et al.*, 1983). When mass spectrometry is combined with g.l.c. as a means of separating peptides as chemical derivatives, it becomes possible to perform peptide mapping experiments very rapidly (Herlihy *et al.*, 1980; Rose *et al.*, 1980, 1983). We have employed peptide mapping by combined g.l.c.–mass spectroscopy in studies of γ-interferon and of α_2-interferon. Although it is in principle a simple task to obtain sequence information from the *N*-terminus and generally throughout a protein (Rose *et al.*, 1983), until now there has been no simple and reliable method for identifying the *C*-terminal peptide during such experiments.

Materials and methods

General sample preparation, derivative formation, reagent preparation, g.l.c. and mass-spectrometric procedures were as previously described (Rose *et al.*, 1983). Digestions (0.1–0.2 mg of sample) were carried out in 1% (w/v) NH_4HCO_3 solution at an enzyme/substrate ratio of 1:100 (w/w) at 37°C for 4 h. 1-Chloro-4-phenyl-3-tosylamidobutan-2-one ('TPCK')-treated trypsin was obtained from Worthington–Millipore, and bovine α-chymotrypsin (type II) from Sigma Chemical Co. Water enriched with ^{18}O (99 atom%) was obtained from ICN and was diluted before use. Peptides were analysed as their $N^{\alpha\epsilon}$-trifluoroacetyl-*NO*-permethyl derivatives. The synthetic hexapeptide was obtained from Bachem. Interferon samples were prepared in *Escherichia coli* by using recombinant-DNA techniques and assayed for antiviral activity in the laboratories of Biogen S.A., Geneva.

Results and discussion

In the course of peptide-mapping experiments on intact γ-interferon we examined a tryptic digest of the native protein. We expected to find a series of peptides terminating in lysine, in addition to the *C*-terminal tryptic peptide Ala-Ser-Gln [those peptides terminating in arginine were to be invisible to us in this experiment, since we employ the permethylation reaction (Morris, 1980; Rose *et al.*, 1983) and had chosen not to form a prior derivative of arginine]. We found, in addition to various peptides terminating in lysine and the expected *C*-terminal peptide Ala-Ser-Gln, similar amounts of the tetrapeptide Leu-Thr-Asn-Tyr. This peptide, corresponding to residues 98–101 of the human γ-interferon sequence (Gray *et al.*, 1982), would be released on its *N*-terminal side by the action of trypsin, since residue 97 is lysine. On the other hand, the cleavage at tyrosine (estimated by g.l.c. to have been almost quantitative) is, though not being

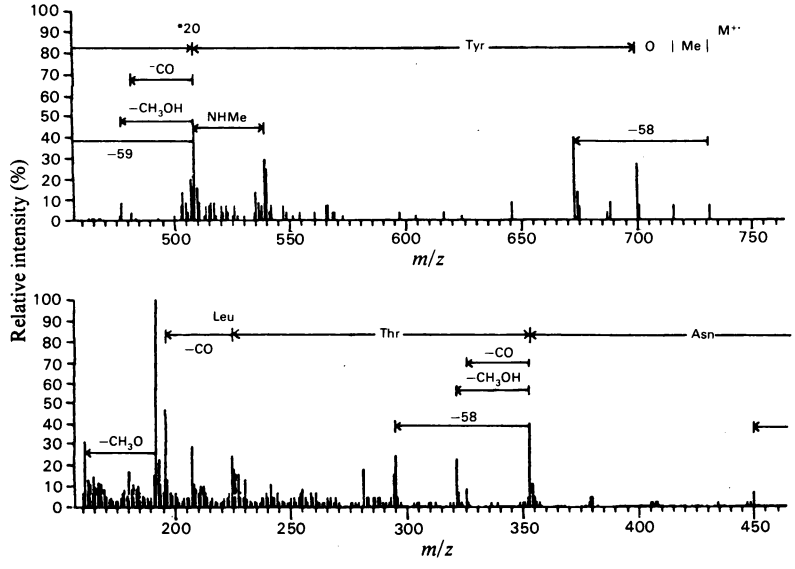

Fig. 1. *Mass spectrum of Leu-Thr-Asn-Tyr*

The peptide, excised from γ-interferon as a result of an untypical cleavage between Tyr-101 and Ser-102, was identified, along with others, after trifluoroacetylation, permethylation and combined g.l.c.–mass spectrometry. In addition to the normal sequence-determining ions, the strong signal at *m/z* 192 informs us that tyrosine is *C*-terminal (structure $H_3C–O–C_6H_4–CH=CO_2CH_3$). Background signals, in particular those at *m/z* 207 and 281, due to bleeding of the stationary phase of the g.l.c. column, have not been subtracted from this spectrum. Assignments corresponding to losses of the side chain of threonine with and without hydrogen transfer (58 and 59 mass units respectively) and other common fragmentations have been made. Intensities of signals above *m/z* 500 have been multiplied by 20.

without precedent, anomalous. We present the mass spectrum of the tetrapeptide in Fig. 1.

The simplest explanations for the presence of this peptide were (*a*) a proportion of the interferon molecules terminated at residue 101 instead of 146, or (*b*) that chymotryptic activity of the trypsin preparation was responsible. Neither one of these explanations could be adopted without confirmation. The interferon sample was known to run as a single band of correct molecular mass on polyacrylamide-gel electrophoresis in the presence of sodium dodecyl sulphate, which tends to exclude explanation (*a*). Also, the trypsin preparation was a solution of purified enzyme that had been treated with the specific inhibitor of chymotrypsin, 1-chloro-4-phenyl-3-tosylamidobutan-2-one, and that had not before shown any signs of untypical activity: indeed, in the digest of γ-interferon examined by mass spectrometry, no other anomalous peptides were present. Had chymotryptic activity been generally expressed, we should have expected to find peptides such as Asn-Trp, Asn-Phe and many others.

To resolve the issue as to whether or not the Tyr–Ser bond was being cleaved during the tryptic digestion or was already cleaved in the sample of protein taken for study, we performed the digestion with trypsin in water enriched with ^{18}O (50–55 atom%). In this way, ^{18}O would be incorporated about equally along with ^{16}O into the tyrosine carboxy group on hydrolysis of the Tyr–Ser bond if this hydrolysis were indeed taking place. Incorporation of ^{18}O into the carboxy groups of the lysine peptides produced by the normal action of the enzyme was to serve as a check that label was not being lost by subsequent sample handling, and the *C*-terminal peptide Ala-Ser-Gln was to serve as a control to determine whether or not label was being incorporated indiscriminately into pre-existing carboxy groups (i.e. without hydrolysis having to take place).

As an additional control for this new method of *C*-terminal-peptide identification by mass spectrometry, we performed a parallel digestion of the synthetic hexapeptide Lys-Glu-Thr-Tyr-Ser-Lys in the same ^{18}O-enriched water, using a mixture of trypsin and chymotrypsin. Analytical high-voltage electrophoresis on paper showed that the Tyr–Ser bond of the hexapeptide was cleaved almost quantitatively. Interpretation of the mass-spectrometric results was very simple after separation of the peptide derivatives by g.l.c. The dipeptide Ser-Lys

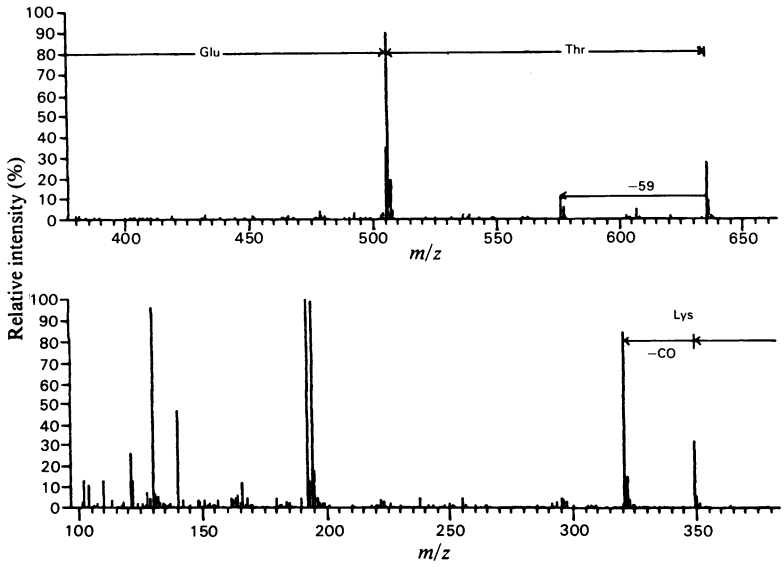

Fig. 2. *Mass spectrum of Lys-Glu-Thr-Tyr*

This tetrapeptide was produced by the action of chymotrypsin on the hexapeptide Lys-Glu-Thr-Tyr-Ser-Lys and the digestion was performed in ^{18}O-enriched water (50–55 atom%). Through signals at *m/z* 192 and 194 (see the legend to Fig. 1) it is clear that tyrosine is in the *C*-terminal position and that the tyrosine residue has incorporated the ^{18}O excess of the medium in which hydrolysis of the Tyr–Ser bond occurred. No label has been incorporated elsewhere. (The intense signal at *m/z* 130 informs us that glutamic acid is present in the peptide and is not in the *N*-terminal position.) Background signals have been subtracted from this spectrum.

was found to have no incorporation of ^{18}O, whereas the tetrapeptide Lys-Glu-Thr-Tyr contained approximately the full enrichment of ^{18}O, uniquely in the carboxy group of tyrosine. The mass spectrum demonstrating this is presented in Fig. 2. This is exactly as predicted by supposing incorporation can take place only on hydrolysis, and that no complications occur due to exchange or loss of label during derivative formation. Had significant enzyme-catalysed exchange been occurring as well, then the Lys-Glu-Thr-Tyr (tyrosine residues being specificity sites for trypsin) and the Ser-Lys (lysine residues being specificity sites for trypsin) would have both contained up to twice the ^{18}O excess of the medium, there being two potentially exchangeable oxygen atoms in a carboxy group.

The digest of γ-interferon in $H_2{}^{18}O$ gave the same sequences as found when normal water had been used. The mass spectrum of the *C*-terminal tryptic peptide Ala-Ser-Gln showed no ^{18}O incorporation, whereas the peptides having lysine at the *C*-terminus (e.g. Leu-Phe-Lys, Thr-Gly-Lys) were found to have incorporated ^{18}O into the carboxy group of the lysine to a proportion equivalent to the isotopic abundance in the medium. The tetrapeptide Leu-Thr-Asn-Tyr had also been incorporated ^{18}O to the isotopic abundance of the medium, and its mass spectrum is presented in Fig. 3. We may therefore conclude that the Tyr–Ser bond in the interferon was completely intact before digestion and that an anomalous cleavage occurred during digestion. Such cleavages have been noted in studies of other proteins, e.g., to take a single example, alcohol dehydrogenase from *Drosophila* (Thatcher, 1980).

It is noteworthy that, despite the apparent lack of any general homology between the sequences of α- and γ-interferons, there is a segment of sequence Arg-Ile-Thr-Leu-Tyr in α-interferon, corresponding to residues 126–130 (Streuli *et al.*, 1980), that resembles the Lys-Leu-Thr-Asn-Tyr sequence of γ-interferon. In addition, digestion of native α_2-interferon with trypsin under similar conditions was found to give rise, in addition to expected peptides, to the peptide Ile-Thr-Leu-Tyr, which had been excised in high yield. Peptides were identified as before by combined g.l.c.–mass spectrometry. We had previously shown that the α_2-interferon sample used possessed its $Tyr_{(130)}$–$Leu_{(131)}$ bond intact: the peptide Leu-Tyr-Leu was found by g.l.c.–mass spectrometry in a subtilisin digest of *S*-carboxymethylated material and could have come from nowhere else in the sequence. It therefore appears that both α_2- and γ-interferons contain a closely similar sequence segment and that both proteins in

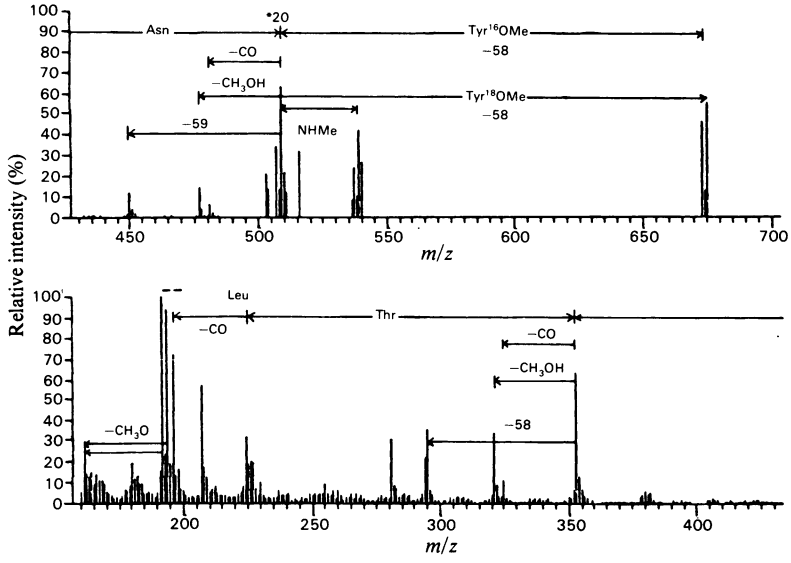

Fig. 3. *Mass spectrum of Leu-Thr-Asn-Tyr produced by digestion of γ-interferon in water enriched with* ^{18}O
The significant difference between this spectrum and that presented in Fig. 1 concerns the signals at *m/z* 194 and 675 that are visible in the present Figure. Although the molecular-ion region was too weak to be recorded, the presence of the signals at *m/z* 194 and 675 with an intensity comparable with those of the signals at *m/z* 192 and 673 respectively establishes beyond doubt that the *C*-terminal tyrosine residue had incorporated the ^{18}O excess of the digestion medium. Compare also Fig. 2. There is no evidence for incorporation of ^{18}O elsewhere in the peptide. As in Fig. 1, no background subtraction has been performed.

the native state are susceptible to proteolytic cleavage at the tyrosine residue that terminates the segment. Levy *et al.* (1981) observed the partial excision of the tetrapeptide Ile-Thr-Leu-Tyr on treatment of α_1-, α_2- and β_1-interferons with trypsin.

We have also examined a sample of what should have been the same γ-interferon, but isolated by a different procedure. This new sample, which has antiviral activity indistinguishable from that of intact γ-interferon, was found by sodium dodecyl sulphate/polyacrylamide-gel electrophoresis to have an M_r value lower than expected. By using a modification (Rose *et al.*, 1983) of a published procedure (Gray & del Valle, 1970), the *N*-terminus was shown by g.l.c.–mass spectrometry to have the correct sequence Cys-Tyr-Cys-, so a clip in the *C*-terminal region was suspected. Digestion of the sample with a mixture of trypsin and chymotrypsin in ^{18}O-enriched water produced a number of peptides, among them Leu-Thr-Asn-Tyr. No Ala-Ser-Gln was found, in agreement with the hypothesis of a *C*-terminal clip. The Leu-Thr-Asn-Tyr and all other peptides except one had incorporated ^{18}O into the carboxy group to the expected extent. The peptide Thr-Gly-Lys, which had not incorporated any ^{18}O, defines the *C*-terminus of the γ-interferon isolated in clipped form by the modified procedure: examination of the known sequence (Gray *et al.*, 1982) indicates that the clip takes place between residues 131 and 132 of this 146-residue protein, which is in agreement with the lower M_r value found for this sample on gel electrophoresis. This clip occurs at the first of a series of four consecutive basic residues: it is possibly produced by the action of a membrane-associated proteinase of *E. coli*. These findings exclude the presence of any significant amount of intact protein in the sample studied. The possibility cannot be excluded of there being additional components in the sample with clips at points nearby to residue 131: all sequence methods are to a greater or lesser degree open to objections of this type. However, if present at all, such molecules would constitute only a minority of those present.

Truncated versions of interferons other than γ-interferon that retain activity are known: forms of α_1-, α_2- and β_1-interferons lacking the *C*-terminal ten amino acid residues predicted by DNA data (and found by protein sequence analysis of α-interferon isolated from Namalwa lymphoblastoid cells; Allen, 1982) have been isolated from chronic myelogous leukaemia cells and shown to retain activity (Levy *et al.*, 1981).

Conclusion

The ^{18}O-labelling technique was developed to identify the *C*-termini of proteins in a peptide, and

not as an amino acid. The method seems to be a useful complement to the mass-spectrometric and conventional procedures that are available for characterizing the *N*-termini of proteins. In cases where trypsin or chymotrypsin are inconvenient, it may be possible to develop different procedures for the partial hydrolysis in $H_2{}^{18}O$. It may also be possible to use different mass-spectrometric techniques, such as fast atom bombardment or field desorption, for the peptide identification.

We thank the Fonds National Suisse de la Recherche Scientifique for a grant to purchase the equipment. We thank Professor F. Goldberger for his suggestion that a membrane-associated proteinase may be responsible for the clip observed in γ-interferon.

References

Allen, G. (1982) *Biochem. J.* **207**, 397–408

Gray, P., Leung, D. W., Pennica, D., Yellrerton, E., Najarian, R., Simonsen, C. C., Rik, D., Sherwood, P. J., Wallace, D. M., Berger, S. L., Lerinson, A. D. & Goeddel, D. V. (1982) *Nature* (*London*) **295**, 503–508

Gray, W. R. & del Valle, U. E. (1970) *Biochemistry* **9**, 2134–2137

Herlihy, W. C., Royal, N. J., Biemann, K., Putney, S. D. & Schimmel, P. R. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6531–6535

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

Morris, H. R. (1980) *Nature* (*London*) **286**, 447–452

Rose, K., Priddle, J. D., Offord, R. E. & Esnouf, M. P. (1980) *Biochem. J.* **187**, 239–243

Rose, K., Simona, M. G. & Offord, R. E. (1983) *Biochem. J.* **215**, 261–272

Streuli, M., Nagata, S. & Weissman, C. (1980) *Science* **209**, 1343–1347

Thatcher, D. R. (1980) *Biochem. J.* **187**, 875–886