Amino acid sequence determination by g.l.c.–mass spectrometry of permethylated peptides

Optimization of the formation of chemical derivatives at the 2–10 nmol level

Keith ROSE,* Marco G. SIMONA and Robin E. OFFORD

Département de Biochimie Médicale, Centre Médical Universitaire, 9 Avenue de Champel, 1211 Genève 4, Switzerland

(Received 28 March 1983/Accepted 14 July 1983)

A new technique is described that permits the permethylation of acylated peptides at the 2–10 nmol level. The presence of up to 400 µg of sodium dodecyl sulphate per sample does not affect the reaction yields. The technique, which is a miniaturization of the widely used methyl iodide/dimethylsulphanyl carbanion procedure, employs a layer of hexane to exclude moisture and oxygen from the reaction mixture. Analysis of the peptide derivatives by combined g.l.c.–mass spectrometry permits amino acid sequence information to be obtained. In addition to studies of digests of a model substrate (glucagon), the new permethylation technique has been used to identify a peptide of interest from a digest of a cytochrome and to define the *N*-termini of two proteins at the 5 nmol level.

In spite of, and to some extent because of, the recent advances in DNA sequencing methodologies, there is a growing need for amino acid sequence determination of proteins and polypeptides. In many cases a full sequence determination is not immediately required, but only a rapid survey of sequence segments to check a DNA reading frame, to define an *N*-terminus or to generate enough sequence information to permit the construction of a complementary DNA probe.

G.l.c.–m.s. analysis of volatile peptide derivatives lends itself very well to this kind of sequence survey (Herlihy *et al.*, 1980; Rose *et al.*, 1980). In this approach, a protein, or a polypeptide fragment, is digested by one or a series of proteolytic agents. The resulting mixture, ideally of di- to hexa-peptides, is converted into volatile derivatives and separated by g.l.c. As each component is eluted it is analysed by a mass spectrometer directly coupled to the g.l.c. column.

For an *N*-terminal sequence assignment, which proceeds slightly differently (Gray & del Valle,

1970) and produces only a single peptide per polypeptide chain for analysis, this may also be introduced to advantage via the gas chromatograph if desired.

Another technique that is applicable to peptide sequencing is fast-atom-bombardment mass spectrometry. This promising method can be used to determine the sequence of up to about 15 amino acid residues in peptides and is frequently successful with approx. 20 nmol of sample, although 2 nmol is adequate in favourable cases (Williams *et al.*, 1982; see also Barber *et al.*, 1981; Morris *et al.*, 1981).

Although g.l.c.–m.s. analysis of derivatives of peptides produces sequence data rapidly and reliably even if a blocked *N*-terminus or post-translational modification is present, the quantity of sample required has been so high as to prohibit certain applications. The difficulty lies in the yields during the formation of the derivatives rather than in the sensitivity of the analysis that follows. The preferred procedures for the formation of derivatives (if peptides larger than dipeptides are to be encountered) are either (a) esterification, acylation and reduction to poly(amino alcohols), which are then trimethylsilylated, or (b) simply acylation followed by permethylation. Using poly(amino alcohols), Biemann's group, which has been responsible for most of the activity in the field, has been able to decrease the sample requirement to 50–100 nmol per g.l.c.–m.s. experiment (Carr *et al.*, 1981). Pre-

Abbreviations used: g.l.c.–m.s., combined gas–liquid chromatography–mass spectrometry; f.i.d., flame ionization detector; SDS, sodium dodecyl sulphate; Dmp, dimethylpyrimidylornithine, the product of the condensation of pentane-2,4-dione with arginine (Gilbert & O'Leary, 1975).

* To whom correspondence should be addressed.

vously 5–10 times this amount was required, but it remains the case that a separate experiment is needed in order to assign side-chain amides. The quantity required then increases to 100–200 nmol.

Other workers have preferred to use the permethylated derivatives, which are less volatile but possess simpler mass-spectral characteristics. With permethylated derivatives, assignment of side-chain amides does not require a separate experiment. By improvements in sample handling, Morris and co-workers have been able to decrease the sample requirement for permethylation from 100–1000 nmol (Morris *et al.*, 1971) down to 30–50 nmol (Dell & Morris, 1981). Up to now, we have required 100 nmol per g.l.c.–m.s. experiment, although, as is the case with the poly(amino alcohols), only a fraction of the product is used for the analysis.

We now report the introduction of a new technique that permits the scaling down of the permethylation reaction sufficiently to produce at least an order of magnitude decrease in sample requirement: 2–10 nmol is all that is required, whether for an *N*-terminal assignment, or for a sequence survey experiment as described above.

Experimental

Glassware

All glassware, except micropipettes, was steeped in chromic acid (Merck, Darmstadt, Germany) for at least 24 h, rinsed thoroughly with ten changes of distilled water and then baked overnight in an oven at 120°C. Afterwards, aluminium foil was used to exclude dust.

Reagents

Unless otherwise specified, solvents and reagents were analytical grade and used without further purification.

Distillation of dimethyl sulphoxide

The procedure followed is a modification of the one described by Charransol & Desfosses (1967). The details are important. Glassware was baked at 120°C immediately before use. A two-necked 1-litre distillation flask was charged via a funnel with oil-free NaH (approx. 3 g, 99% NaH; EGA-Chemie, Steinheim, W. Germany), triphenylmethane (approx. 0.3 g, purum; Fluka, Buchs, Switzerland) and dimethyl sulphoxide (500 ml, puriss., Fluka) in a glove-bag (Instruments for Research and Industry, Cheltenham, PA, U.S.A.) under an atmosphere of dry O₂-free N₂ (<3 p.p.m. H₂O; <2 p.p.m. O₂). A fine glass bleed was introduced, and the flask was then attached to a fractionating column, thermometer, Liebig condenser, adaptor and a pair of

collecting flasks. The apparatus, outside the glove-bag, was immediately put under rotary-pump vacuum. Dry O₂-free N₂ entered via the bleed. Cones were lubricated with silicone high-vacuum grease along only the top two-thirds of their length. The flask contents were warmed to about 40°C and a brisk effervescence ensued. As the effervescence subsided, the contents took on a deep-red hue and the distillation commenced. The flask contents must remain red throughout the procedure. The first 50 ml was collected in one flask, and the major fraction, about 300 ml, in the second at a temperature of about 38°C. The apparatus, still under vacuum, was allowed to cool before being flooded with N₂ via a three-way tap on the vacuum line. The flask containing the redistilled dimethyl sulphoxide was immediately stoppered and transferred to the glove-bag, where the contents were pipetted into Sovirel bottles, which were closed with a polytetrafluoroethylene-lined screw-cap. The bottles were stored over NaOH at room temperature in the dark in a desiccator evacuated by means of a rotary pump. From time to time, a bottle was withdrawn and the contents were dispensed into N₂-flushed 2 ml ampoules, which were sealed in the flame, care being taken to avoid any pyrolysis of the dimethyl sulphoxide. Each 2 ml ampoule contained approx. 1 ml of dimethyl sulphoxide. Residual NaH in the still was safely disposed of by cautious addition of propan-2-ol.

Preparation of dimethylsulphinyl sodium

The method is similar to that of Rose (1979), but the modifications (type of NaH, use of glove-bag, improved apparatus) are important. Glassware was baked at 120°C immediately before use. A 250 ml flask was charged via a funnel with oil-free NaH (approx. 3.6 g, accurately weighed) in a glove-bag as above. Redistilled dimethyl sulphoxide was then added (27.7 ml/g of NaH), the operation still being performed in the bag. An inverted glass U-tube was attached and clipped to the flask via a cone at one end greased along the top two-thirds of its length. The other end of the U-tube was closed with a short length of silicone tubing, which was clamped off. The reaction flask was withdrawn from the bag and swirled to generate a slight positive pressure of H₂. At this point, the silicone tube was attached to a bubble-trap containing NaH/dimethyl sulphoxide/triphenylmethane, which had been warmed briefly to 40°C to generate dimethylsulphinyl sodium and was therefore deep red. The clamp was released and the H₂ was then bubbled steadily through the trap. The flask was maintained at about 40°C (no higher) and swirled from time to time until the evolution of gas subsided: approx. 3 h. The reaction does not go to completion, since a small proportion of the NaH is very slow to react. The flask contents must be

almost colourless, a slight grey tone being imparted by fine bubbles and a suspension of the unaltered NaH. If the liquid is yellow, or darker still, this is due to the entry of air, or to overheating, or both, and such preparations should be discarded. The flask was allowed to cool, and was clamped off when liquid from the trap, which must remain red throughout the procedure, began to rise. The flask, still clamped off, was then detached from the trap and transferred to the glove-bag, where the flask contents were pipetted into a series of 15 ml conical centrifuge tubes, which were then fitted with polythene caps and wrapped with sealing film. Care was taken, when pipetting, not to touch the silicone grease on the neck of the flask. The tubes were centrifuged at maximum speed in a swing-out bench-top centrifuge for 5 min. They were then tapped gently to dislodge any NaH from the surface of the liquid, and re-centrifuged for 5 min. They were then carefully removed, and traces of sediment were prevented from rising by freezing the tips of the tubes in powdered solid CO₂. The tube contents should be water-clear, with very little sediment. Each tube was opened in turn, and the contents were immediately taken up into a large drawn-out pipette that had been purged with N₂. The liquid was then injected into N₂-flushed ampoules made from drawn-out 12 mm × 75 mm Pyrex test tubes (commercial vials have a tendency to split when the contents are frozen and then thawed), which were sealed in the flame. The drawn-out pipette was wiped each time. If any liquid was deposited by accident near the neck of the ampoule, pyrolysis occurred during flame-sealing: such ampoules were rejected. The ampoules were stored at -18°C. One ampoule was not frozen, but was centrifuged in a bench-top centrifuge for 5 min at maximum speed (no deposit was normally visible) and then opened. Known volumes of the contents (e.g. 3 × 200 μl) were each diluted into 5 ml of water containing 1 drop of Methyl Orange as indicator and were titrated against standard 0.1 M-HCl or -H₂SO₄. If the preparation had proceeded normally, the dimethylsulphinyl sodium was found to be about 1.1 M. If the reaction is stopped too soon, the concentration may be as low as 0.6 M, but this is still adequate: the volumes of the reagent solution used must simply be adjusted in consequence.

Digestion

Samples of polypeptide or protein (0.5–10 nmol) were incubated at 37°C for 4 h with subtilisin (type VIII; Sigma Chemical Co., St. Louis, MO, U.S.A.) (enzyme/substrate ratio 1:100–1:50, w/w) in small tubes (30 mm × 4.5 mm internal diam.) in 50 μl of 1% (w/v) NH₄HCO₃ containing 0–400 μg of SDS (minimum 99%; Serva, Heidelberg, W. Germany). The tubes were closed with aluminium foil held on with sealing film. After the incubation the foil was

pierced and the contents were freeze-dried over fresh NaOH.

Acetylation

Freeze-dried digests (see above) were taken up in 50 μl of pyridine/water (1:1, v/v) and, during vortex-mixing, 5 μl of acetic anhydride or of a 1:1-molar acetic anhydride/[²H₆]acetic anhydride mixture was added. (The labelled reagent, >99 atom% ²H, was obtained from EGA-Chemie.) Mixing was continued for 1 min, and the tube was then closed with aluminium foil held on with sealing film and set aside. After a further 10 min, 5 μl of reagent was again added during mixing, and the tube was closed and set aside. After 10 min, the foil was pierced, and solvent and reagents were removed in a desiccator over NaOH under rotary-pump vacuum. In the case of an undigested sample for *N*-terminal assignment (2–5 nmol), the above operations were performed as usual, except that, instead of the initial addition of 50 μl of pyridine/water, the sample was dissolved in 25 μl of water containing up to 400 μg of SDS, and 25 μl of pyridine was then added. The purpose of the SDS is to improve protein solubility, where necessary.

Trifluoroacetylation

To the freeze-dried digest (see above) was added 50 μl of pyridine. Then 5 μl of phenyl trifluoroacetate (pure; Koch–Light Laboratories, Colnbrook, Bucks., U.K.) was added during vortex-mixing, and mixing was continued for 1 min. The tube was then closed with aluminium foil held on with sealing film, and was set aside at room temperature, with vortex-mixing occasionally over a period of 1 h. (Note: SDS is not very soluble in pyridine.) After this time the foil was pierced, and solvent and reagent were removed in a desiccator over NaOH, with pumping overnight with a rotary pump. The residue was dissolved in 50 μl of 1% (w/v) NH₄HCO₃ and freeze-dried to remove traces of phenol.

Permethylation

An ampoule of dimethylsulphinyl sodium was taken from the freezer and thawed rapidly by rolling between the palms of the hands (if left to thaw naturally, the ampoule may split as the contents expand). It may be preferable, on safety grounds, to agitate the ampoule in a beaker of tepid water, taking care not to break the seal. After reaching room temperature, the ampoule was centrifuged in a bench-top centrifuge at maximum speed for 5 min, although there was seldom any visible deposit. The vial of dimethylsulphinyl sodium was then opened and immediately filled to the neck with hexane (pro analysi; Merck, Darmstadt, W. Germany). An ampoule of dimethyl sulphoxide was opened and at

once was filled to the neck with hexane. To a tube containing a sample to be permethylated (either an acylated digest or, in the case of an *N*-terminal analysis, a digest of an acetylated protein or polypeptide) was added slowly 200 μl of hexane. A 50 μl micropipette attached to a syringe was then used to take up an approx. 0.5 cm column of hexane followed by 50 μl of dimethyl sulphoxide. A further 0.5 cm column of hexane was drawn in after the dimethyl sulphoxide so that the dimethyl sulphoxide was sandwiched without air bubbles between the columns of hexane. The contents of the micropipette, but no air, were then injected into the sample tube. (The same vial of dimethyl sulphoxide was used for a series of up to six samples, in which each sample was carried through to the extraction with chloroform before addition of dimethyl sulphoxide to the next). A 10 μl portion of dimethylsulphinyl sodium (1.1 M) was withdrawn into a micropipette into which a 0.5 cm column of hexane had already been drawn. If the concentration of dimethylsulphinyl sodium was not 1.1 M, the volume taken was adjusted accordingly. The outside of the micropipette was then wiped, and the contents were injected into the sample tube, the point of injection being beneath the surface of the dimethyl sulphoxide (and thus under the hexane also). Care must be taken not to expel any air from the micropipette. The micropipette, still containing a few microlitres of hexane, was used to stir the dimethyl sulphoxide layer regularly for 10 s. Then 3 μl of the dimethyl sulphoxide layer was withdrawn, sandwiched between 5 mm lengths of hexane, and injected either directly on to a layer of fresh triphenylmethane or into a test tube containing a saturated solution of triphenylmethane in hexane. A transient, but definitely red, coloration indicated the presence of sufficient base, otherwise a further quantity was added to the sample and tested in the same way. The coloration lasts longer if a solution of triphenylmethane is used. (For digests of 5 nmol of protein up to M_r 20 000, extra dimethylsulphinyl sodium should not be necessary, and a negative test with triphenylmethane would rather indicate imperfect removal of acylation reagents or the presence of moisture in the apparatus, in the hexane or in the dimethyl sulphoxide.) Approx. 18 μl of methyl iodide (puriss.; Fluka) from a micropipette containing 25 μl of methyl iodide was then rapidly injected beneath the surface of the dimethyl sulphoxide, followed by stirring of the reaction mixture for 5 s. (Methyl iodide is not sandwiched between hexane columns as, although it is not extracted to a prohibitive extent from the reaction mixture by hexane, it is nonetheless miscible with it in the pure state.) At 90 s after the addition of the methyl iodide most of the hexane layer was removed (Pasteur pipette) and the reaction was quenched by the addition of 50 μl of

water. This was followed by 100 μl of chloroform (pro analysi for residue analysis; Merck), and the tube was carefully vortex-mixed for 30 s. The aqueous (upper) phase was removed, after centrifugation to assist separation of the phases. The walls of the tube were rinsed with water, which was then discarded. The organic phase was then washed three times by vortex-mixing with 100 μl portions of water. After the final wash, the organic phase was carefully evaporated in a desiccator over NaOH at the water pump, the desiccator tap being only partially opened to prevent bumping of the solvent. When the sample appeared to be completely dry, the desiccator was transferred to a rotary pump and pumped on overnight.

G.l.c. analysis with a flame ionization detector (f.i.d.)

The dried permethylated sample was taken up in 10 μl of chloroform and warmed very briefly. A volume of 0.2 μl was taken up into a 5 μl syringe primed with hexane (the syringe should register 0.5 μl of hexane). The liquid was then withdrawn into the barrel of the syringe and injected directly via the on-column injector on to a fused silica column (typically 8 m $\times$ 0.32 mm internal diam.; 0.14 μm coating of Sil 5; Chrompack BV, Middelburg, The Netherlands). After 1 min, the oven of the Carlo Erba 4160 gas chromatograph was programmed from an initial 80°C to 320°C at 6°C/min. The detector was maintained at 360°C. Helium was used as carrier gas at an inlet pressure of 0.3 kg/cm². Output was to a 10 mV pen recorder.

G.l.c.-m.s. analysis

The instrument consisted of a Carlo Erba 4161 gas chromatograph coupled to a Kratos MS 50S mass spectrometer (Kratos, Manchester, U.K.). A fused silica capillary column similar to that used for the f.i.d. analysis was inserted directly into the ion source of the mass spectrometer through a modified g.l.c.-m.s. interface (Rose, 1983). The instrument was operated in low-resolution electron-impact mode under field control of the magnet current. The source temperature was 200°C, the accelerating voltage 8 kV, and the electron-beam energy 70 eV. The source housing was kept warm by leaving the low-source bake permanently selected (except when changing columns or the source). The instrument was scanned at 1 s/decade over a calibrated mass range of 28–832 atomic mass units with a cycle time of about 2.5 or 3.5 s depending on the interscan delay selection. Data were acquired and processed by a DS 55 M data system (Kratos). Samples (0.2–1.0 μl) in solution in chloroform were slowly injected via the on-column injector from a needle primed with and registering 0.2 μl of hexane. The oven temperature was 120°C and the He carrier-gas

inlet pressure was 0.1 kg/cm² (above ambient). While an injection was being made, the filament, ion gauges and accelerating voltage were turned off and the valve isolating source and analyser regions was closed to finger tightness. The oven temperature was maintained at 120°C for 3–5 min, during which time the carrier-gas flow rate was adjusted to a known value between 0.7 and 1.5 ml/min, as measured at the outlet of the source rotary pump or taken from the reading of the source ion gauge (which was periodically calibrated against the measured flow at the pump outlet), thus causing the column inlet pressure to fall below ambient (Rose, 1983). After this time, the oven temperature was programmed at 4°C/min to 280–320°C, depending on the result obtained with the f.i.d. The interface temperature was maintained at 250°C. A set of alkane standards, n-C₁₈H₃₈ to n-C₄₄H₉₀, run on the f.i.d. and g.l.c.–m.s. instruments from time to time, served to correlate approximately the retention times under the different conditions.

Results and discussion

Permethylation technique

In the past, particularly when direct-insertion techniques have been used, mass spectrometry of unknown peptides has not been at all quantitative. Either an interpretable spectrum was obtained, or not (see, e.g., Morris, 1972; Morris *et al.*, 1973). The new permethylation procedure of acylated digests (whether for *N*-terminal assignment or sequence survey) is semi-quantitative. This is to say, we are able to estimate from the f.i.d. response, with a precision of about 100%, the number of picomoles injected into the gas chromatograph. For an acetylated digest of a medium-sized polypeptide (30 residues), the amount of derivative corresponding to 100 pmol of peptide should give a response of about 2–3 pA under the typical chromatographic conditions in the Experimental section. Under these conditions, 10 ng of n-C₃₄H₇₀ gives a response of about 6 pA. Of course, a digest with a non-specific proteinase will produce peptides in various yields, and this will be reflected in the analysis that follows. Nevertheless, certain peptides will be present in fairly high yield.

We have performed acetylation and permethylation of digests of glucagon (2–10 nmol) and of soya-bean trypsin inhibitor (0.5–10 nmol) in the presence of various amounts of SDS (0–400 µg). (The effect of the addition of SDS was studied since SDS often improves sample solubility, and also because we hope eventually to work with samples eluted from gels containing SDS.) Analysis by g.l.c. (f.i.d.) produced a response for the respective peptides that varied approximately linearly with the amount permethylated in the range 2–10 nmol

Table 1. *Yields of permethylated peptide derivatives*

This Table gives the yields of acetylated permethylated peptides determined by g.l.c. with f.i.d. detection. The sample which was digested with subtilisin, was either pig glucagon or soya-bean trypsin inhibitor. The formation of derivatives was performed with various amounts of sample (nmol, *n*) in the presence of SDS (0–400 µg per sample). The global yield, Σ , in arbitrary units (mm on the recorder trace), is the sum of the peak heights of nine prominent components with glucagon and of 14 prominent components with soya-bean trypsin inhibitor. The ratio Σ/n expresses the global yield per nmol of sample and should ideally be constant for a given sample. It is seen to be more nearly constant for soya-bean trypsin inhibitor than for glucagon, and not to be greatly affected by the presence of SDS up to 400 µg per sample. The presence of SDS was found to give rise to small amounts of additional components, which were shown by g.l.c.–m.s. analysis to be peptide derivatives carrying too many or too few methyl groups. The peak heights of such components have not been included in the value for the principal components.

Sample	Amount permethylated (nmol) (<i>n</i>)	SDS present (µg)	Global yield (Σ)	Σ/n
Glucagon	10	0	444	44
	5	0	221	44
	2	0	126	63
	5	50	279	56
	5	100	419	84
	5	200	264	53
	5	400	352	70
Soya-bean trypsin inhibitor	5	0	992	198
	9	400	1753	195
	5	400	1023	205
	2	400	459	230

(Table 1). No evidence was found to suggest that the presence of SDS (up to 400 µg, at least) decreased the yields of peptide derivative. Selected samples were then examined by g.l.c.–m.s. and gave amino acid sequence information (see below). Below 2 nmol, reliable results were not obtained; to work below this level either a carrier substance would have to be added or further miniaturization of the permethylation reaction would be required.

In order to obtain reproducible results it was found necessary to increase the ratio of methyl iodide to dimethylsulphonyl sodium (20 µl of methyl iodide/10 µl of 1.1 M-dimethylsulphonyl sodium) over that used previously (15 µl of methyl iodide/50 µl of 2 M-dimethylsulphonyl sodium; Rose, 1979). Presumably this excess is required to compensate for losses by extraction into the hexane.

The semi-quantitative nature of the new permethylation is especially important for *N*-terminal assignments, where the finding of the expected f.i.d.

Fig. 1. Mass spectrum of an *N*-terminal peptide

Ions appearing at m/z 88 [$\text{HN}(\text{CH}_3)=\text{CH}-\text{CH}_2-\text{O}-\text{CH}_3$] and 100 [$\text{HN}(\text{CH}_3)=\text{CH}-\text{C}_4\text{H}_9$] inform us of the presence of serine and leucine/isoleucine respectively, but do not indicate their position in the sequence. The signal at m/z 158 indicates the presence of serine as *N*-terminal residue, with characteristic losses of 28 (CO) and 32 (CH_3OH) atomic mass units. The interval $285-158=127$ atomic mass units corresponds to the residue leucine/isoleucine, so the sequence Ser-Ilx- is assigned to the signal at m/z 285. Again, losses of CO and CH_3OH are observed, confirming the assignment. The presence of an ion series 3 atomic mass units higher than those discussed is due to the $^2\text{H}_3\text{C}-\text{CO}-$ group used for the acetylation (1:1 with $\text{CH}_3-\text{CO}-$) and serves to demonstrate that the *N*-terminal amino group was free before chemical treatment.

response is a good indication of homogeneity with respect to the *N*-terminal region.

N-Terminal assignment

As was reported by Gray & del Valle (1970), if a protein is acetylated with an acetic anhydride/ $[\text{H}_6]$ acetic anhydride mixture, then digested with a proteinase and finally permethylated, only the *N*-terminal peptide should form a volatile derivative for subsequent mass-spectral analysis. All other peptides, having a free α -amino group, become quaternized by the methyl iodide and are thereby rendered involatile. Use of the isotopically labelled acetyl group not only facilitates mass-spectral interpretation, but also provides a distinction between an *N*-terminal peptide that was naturally acetylated and one that was acetylated only during the chemical treatment (Gray & del Valle, 1970).

For the original work described by Gray & del Valle (1970), 100–200 nmol of protein was used. By using the new permethylation procedure, we have been able to perform similar analyses at the 5 nmol level. For example, 5 nmol of a newly isolated protein of approximate M_r (by SDS/polyacryl-

amide-gel electrophoresis) 21000 (kindly given by Dr. Paul Wingfield, Biogen S.A., Geneva, Switzerland) was acetylated in the presence of 100 μg of SDS. After digestion with subtilisin followed by permethylation, g.l.c. (f.i.d.) of one-fiftieth of the product showed one prominent peak (in addition to the phthalate plasticizer that is a contaminant of the methyl iodide). Analysis by g.l.c.-m.s. of one fifth of the remaining derivative, corresponding to approx. 1 nmol of protein, produced the mass spectrum shown in Fig. 1. The protein clearly commences Ser-Ilx-, and had a free *N*-terminus before formation of the derivative. We propose here the three-letter code Ilx to mean leucine or isoleucine not determined, by analogy with Glx (glutamic acid or glutamine not determined) and Asx (aspartic acid or asparagine not determined). We prefer Ilx to the previous suggestions Jeu (Auffret *et al.*, 1981) and Lle (Allen, 1982) as being more explicit of an ambiguity. One may nevertheless use J as the one-letter code (Auffret *et al.*, 1981): compare Trp = W, for instance. The DNA sequence of the construction thought to be responsible for production of the protein translates as Met-Ser-Leu-etc.

Fig. 2. Mass spectrum of the *N*-terminal peptide of a γ -interferon

Ions appearing at m/z 104 [$\text{HN}(\text{CH}_3)=\text{CH}-\text{CH}_2-\text{S}-\text{CH}_3$]⁺ and 121 [$\text{CH}_2-\text{C}_6\text{H}_4-\text{O}-\text{CH}_3$]⁺ inform us of the presence of cysteine and tyrosine respectively, but do not indicate their positions in the sequence. The signal at m/z 192 [$(\text{C}_6\text{H}_5-\text{CH}=\text{CH}-\text{CO}_2\text{CH}_3)^+$] indicates that tyrosine is in the *C*-terminal position, and *N*-terminal cysteine is assigned to the $\text{CH}_3/^3\text{H}_3$ doublet at m/z 174 and 177 (see the legend to Fig. 1). By its retention position on the g.l.c. column, we know that we are dealing with a dipeptide, and the sequence Cys-Tyr is confirmed by the presence of a signal at m/z 323, which corresponds to *N*-C cleavage at the cysteine residue. Such a cleavage, well documented for aspartic acid, asparagine, phenylalanine, histidine, tyrosine and tryptophan residues (Dell & Morris, 1981), is consistently observed in this laboratory for residues of *S*-alkylcysteine also. The *N*-terminal sequence deduced is thus Cys-Tyr-.

The result of the g.l.c.-m.s. experiment is in agreement with this and permits one to postulate that the *N*-formylmethionine residue is removed by processing after translation.

Since such analyses require the *N*-terminus to be blocked, they are equally possible when the *N*-terminus is blocked naturally.

We should point out that the above simple procedure fails if residues of either cystine or arginine, or both, are to be found in the *N*-terminal peptide produced by digestion of the acetylated protein. This is because the side chains of these amino acids require modification before the permethylation reaction. In the case of cystine, reduction and *S*-alkylation is most suitable, in particular *S*-methylation (Rose, 1979); sulphitylation is acceptable, but oxidation with performic acid is not appropriate when one is attempting to sequence cysteine-containing peptides (Rose, 1979). Reduction and alkylation normally precede the acetylation step, serving in some cases to denature the protein and so better expose the *N*-terminus to the acetylating reagent. In the case of arginine, several pro-

cedures have been recommended for rendering the side chain of arginine compatible with the permethylation reaction. Although hydrazinolysis to ornithine is simple and in the case of small peptides often effective (Morris *et al.*, 1973), nucleophile-labile peptide bonds are broken during the reaction (Priddle *et al.*, 1978; Rose, 1979), thus creating new termini. Treatment with hydrazine would therefore have to follow and not precede acetylation of the protein (and follow digestion, for convenience). Because of the difficulty of removing hydrazine hydrate afterwards, we prefer a quite different approach: condensation with pentane-2,4-dione. This reaction may be performed under mild conditions with intact protein. Both of the arginine residues of glucagon may be modified without any detectable degree of deamidation occurring, and the reaction works well at the 6 nmol level with cytochrome *c* (K. Rose, unpublished work).

So far we have not optimized the reduction and alkylation of cystine at the 5 nmol level. When enough protein is available, this step may be performed with a larger amount (e.g. 50 nmol) and

Fig. 3. Mass spectrum of Gly-Dmp

This spectrum, which has not been subjected to any averaging or background subtraction procedures, was obtained from a g.l.c.-m.s. run of a digest of a mixture of fragments 1-38 and 39-104 of horse heart cytochrome c. Signals at m/z 150, 164 and 178 are due to the side chain of Dmp.

the product then divided for an *N*-terminal analysis and for various digests to provide general sequence information. Nevertheless, we have succeeded in adapting and scaling down the *S*-methylation procedure of Rochat *et al.* (1970), replacing 5 M-guanidinium chloride by 50% (v/v) propan-2-ol, the phosphate buffer by 1% (w/v) NH_4HCO_3 , and gel filtration by acetone precipitation. Thus 5 nmol of a γ -interferon (kindly given by Dr. Berndt Otto, Biogen S.A.), which has cystine (or cysteine) as the *N*-terminal residue, gave, after acetylation, *S*-methylation, digestion with subtilisin and permethylation, on analysis by g.l.c.-m.s., the spectrum shown in Fig. 2. The sequence Cys-Tyr- is clearly established, and is in agreement with the sequence expected on the basis of DNA data. The response obtained on the f.i.d. was, however, significantly lower than expected. At present, we do not know if this was due to low yield during *S*-methylation and acetone precipitation, or whether it reflects a characteristic of the molecule concerned; nevertheless, the sequence information contained in Fig. 2 was obtained with no more than 5 nmol of protein as starting material.

In the case of the *N*-terminal assignment method discussed in the present paper, the direct-insertion probe could be used to introduce the sample into the mass spectrometer, as it was originally (Gray & del Valle, 1970). The advantages of using g.l.c. are

fourfold. Firstly, g.l.c. with f.i.d. detection provides a semi-quantitative measure of the number of nano-moles of sample. Secondly, g.l.c. provides a reliable measure of volatility, the Retention Index (Kovats, 1965), which may be used to assist the mass-spectral interpretation (van de Graaf *et al.*, 1978; Nau *et al.*, 1975). Thirdly, g.l.c.-m.s. provides an efficient means of transferring the sample, already in the gas phase, as a vapour pulse into the mass spectrometer. Fourthly, in cases where a mixture is present, e.g. as a result of an impurity or there being more than one polypeptide chain linked by disulphide bonds, g.l.c.-m.s. provides a much better separation than may be achieved by using the direct-insertion probe, thus simplifying interpretation.

Sequence survey

By the term sequence survey we mean an attempt to generate, in a single g.l.c.-m.s. experiment, sequence information from all parts of the molecule of interest. Through a series of such surveys, a complete, fully overlapped, sequence of the sample molecule may in principle be obtained. We now illustrate the application of the new permethylation technique to two sequence surveys.

Finding a single peptide of interest

In order to assist protein semi-synthetic work being undertaken in this laboratory, it was important

Fig. 4. Results of g.l.c.-m.s. analysis of a digest of glucagon

Key to one-letter code: A, Ala; D, Asp; F, Phe; G, Gly; H, His; K, Lys; L, Leu; M, Met; N, Asn; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr. Shown aligned beneath the sequence of pig glucagon are the peptides identified by g.l.c.-m.s. analysis of a subtilisin digest of glucagon that had been modified at both arginine residues with pentane-2,4-dione. The upper presentation shows those peptides found on acetylation and permethylation of 5 nmol of digest. One-fifth of the treated mixture was used for the g.l.c.-m.s. analysis. The lower presentation shows those peptides found on trifluoroacetylation and permethylation of 10 nmol of the same digest. Again, one-fifth of the treated mixture was used for the g.l.c.-m.s. analysis.

to developed a method that would definitively and rapidly, and consuming only a small amount of material, establish that a peptide link had been formed between two fragments of horse heart cytochrome *c*. The fragments concerned were 1-38 and 39-104, and the peptide link that was sought was -Arg₍₃₈₎-Lys₍₃₉₎-. Examination of the published sequence showed that the action of a combination of trypsin and chymotrypsin on the intact protein, after condensation of arginine residues with pentane-2,4-dione, would be expected to release the tripeptide Gly-Dmp-Lys. The protein contains two arginine residues; the other would be expected to reside in a peptide too large to be eluted under our typical chromatographic conditions.

A 5 nmol portion of horse heart cytochrome *c*, which had been condensed with pentane-2,4-dione and digested with trypsin and chymotrypsin, was acetylated and permethylated, and one-fifth of the product was analysed by g.l.c.-m.s. in a single experiment. The selected ion record of *m/z* 150, diagnostic for the side chain of the dimethylpyrimidylornithine residue, showed only one maximum and in a position corresponding to that of a tripeptide of the expected composition. The corresponding mass spectrum clearly established the existence of a peptide bond between residues 38 and 39. What is more, the f.i.d. response obtained on a one-fiftieth portion of the same treated mixture was consistent with the number of nanomoles taken. The desired information was thus obtained semi-quantitatively and with very little material. A control experiment was then performed with, instead of the native cytochrome, a mixture of fragments 1-38 and 39-104. G.l.c.-m.s. analysis of the corresponding treated digest of this sample showed none of the Gly-Dmp-Lys, and instead produced a strong spectrum of the dipeptide Gly-Dmp (Fig. 3). We are thus in a position to screen rapidly samples that may contain various amounts of coupled material in the presence of the uncoupled fragments, and, more

significantly, in the presence of the fragments coupled at points other than at the peptide bond in question.

Sequencing a polypeptide

In order to obtain the maximum amount of sequence information from a protein, it is necessary to perform a preliminary cleavage and separation of the fragments. Under suitable conditions, the separation may be efficiently and rapidly performed by high-pressure liquid chromatography. As a model of a polypeptide produced in such a fashion, we have studied pig glucagon, a 29-residue polypeptide chosen as a model for other sequencing techniques (e.g. Kelley *et al.*, 1975; Shimonishi *et al.*, 1980; Hunt *et al.*, 1981).

A subtilisin digest of 5 nmol of glucagon (with arginine as dimethylpyrimidylornithine) gave, after acetylation and permethylation, the sequence information contained in Fig. 4. Only one-fifth of the chemically modified material, corresponding to 1 nmol of glucagon, was used for the g.l.c.-m.s. experiment. Examples of the mass spectra are reproduced in Fig. 5: there are no sensitivity problems with this amount of sample, and the spectra are clear and strong. A fully overlapped sequence cannot be reconstructed from the results of this one experiment, but a considerable amount of information has been acquired in a g.l.c.-m.s. experiment lasting about 45 min; 21 of the 29 residues are represented.

A similar experiment was performed with a subtilisin digest of 10 nmol of glucagon (with arginine as dimethylpyrimidylornithine) except that trifluoroacetylation was used rather than acetylation. Again only one-fifth of the chemically modified material, this time corresponding to 2 nmol of glucagon, was used for the g.l.c.-m.s. analysis, which produced the sequence information contained in Fig. 4. All 29 residues are represented, and some overlaps have been obtained. Examples of the mass

Fig. 5. Mass spectra of (a) acetylated permethylated Dmp-Ala-Gln and (b) acetylated permethylated Leu-Asp-Ser-Dmp. (a) This spectrum, interpreted along the lines indicated in the legends to Figs. 1–3, demonstrates how easily the assignment of side-chain amides may be made mass-spectrometrically. The peptide Dmp-Ala-Glu would have a molecular ion doublet at m/z 564 and 567. (b) This spectrum again demonstrates the facility with which amides may be assigned by mass spectrometry of the permethylated derivatives. The sequence Leu-Asn would give rise to a doublet of ions at m/z 326 and 329 instead of at 313 and 316 as here. In this spectrum, the ion at m/z 100 indicates the presence of leucine (see Fig. 1). We may write leucine rather than leucine/isoleucine since glucagon contains no isoleucine residue. The signal at m/z 365 is due to N–C cleavage at serine with charge retention on the C-terminal fragment. Such N–C cleavages, normally less prominent, are more pronounced in peptides containing Dmp (see also Fig. 6b).

spectra are reproduced in Fig. 6, permitting comparison with the acetyl counterparts (Fig. 5).

The yields of permethylated product after trifluoroacetylation at the 10 nmol level seem to be about half of those that follow acetylation.

Repetition of the cytochrome *c* experiment (see above) at the 5 nmol level but with the use of trifluoroacetylation rather than acetylation also produced an f.i.d. response about half of that obtained when acetylation is used. Ample de-

Fig. 6. Mass spectra of (a) trifluoroacetylated permethylated Dmp-Ala-Gln and (b) trifluoroacetylated permethylated Leu-Asp-Ser-Dmp

(a) This spectrum, from which the sequence is easily deduced, may be compared with its acetyl counterpart (Fig. 5a). The intensity distribution is more even in the case of the trifluoroacetyl derivative, so no scale expansion was required (contrast Fig. 5a). (b) Comparison of this spectrum with its acetyl counterpart (Fig. 5b) shows again the more even distribution of intensities in the trifluoroacetyl case (a smaller scale expansion used). The N-C cleavage responsible for m/z 365 (see Fig. 5b) occurs here also.

derivative is nevertheless produced to permit successful analyses at this level, in part because of the more even intensity distribution in the mass spectra (see Fig. 6), but where sample is in very short supply (2 nmol) acetylation is to be preferred at present.

Conclusion

The new permethylation procedure permits the successful formation of derivatives of acylated peptides at the 2–10 nmol level. At the 5 nmol level, the technique has been found to be without problems for

digests of polypeptides varying in size from 29 residues (glucagon) to over 100 residues (cytochrome *c*, soya-bean trypsin inhibitor and a new protein of *M_r* approx. 21 000) even in the presence of substantial quantities of SDS. The procedure is quite general in its applicability, and it would be expected to decrease the sample requirements for direct-insertion mass-spectral techniques.

We have not yet studied in detail the prior formation of derivatives of cysteine (*S*-methylation) and arginine (condensation with pentane-2,4-dione) residues at the 2–10 nmol level, although some success has already been achieved at this level (see above). It should be feasible to scale down the appropriate reactions and then to use high-pressure liquid-chromatographic techniques to remove excess of reagents. Alternatively, when sufficient sample is available, the modification of cysteine and arginine residues may be performed with larger amounts and the product divided afterwards. Of course, even if no attempt is made to modify cysteine and arginine residues it is still possible to obtain sequence information from those peptides that contain residues of neither of these two amino acids.

The sample requirement per g.l.c.–m.s. experiment has been decreased by, in the case of an acetylated digest, at least an order of magnitude over the 50 nmol needed if reduction to poly(amino alcohols) is to be performed (Carr *et al.*, 1981), and no additional experiment is required in order to assign side-chain amides as is the case for the reduced derivatives. With this increase in sensitivity, we expect g.l.c.–m.s. analysis of permethylated peptides to assist the rapid sequencing of proteins and polypeptides available in limited quantities.

We thank the Fonds National Suisse de la Recherche Scientifique for a grant, which permitted purchase of the equipment. M. G. S. thanks Biogen S.A., Geneva, for a personal stipend. We thank Professor Julian Davies for his interest and encouragement.

References

- Allen, G. (1982) *Biochem. J.* **207**, 397–408
- Auffret, A. D., Blake, T. J. & Williams, D. H. (1981) *Eur. J. Biochem.* **113**, 333–338
- Barber, M., Bordoli, R. S., Sedgwick, R. D., Tyler, A. N. & Whalley, E. T. (1981) *Biomed. Mass Spectrom.* **8**, 337–342
- Carr, S. A., Herlihy, W. C. & Biemann, K. (1981) *Biomed. Mass Spectrom.* **8**, 51–61
- Charransol, G. & Desfosses, B. (1967) *Eur. J. Steroids* **2**, 423–441
- Dell, A. & Morris, H. R. (1981) *Biomed. Mass Spectrom.* **8**, 128–136
- Gilbert, H. F. & O'Leary, M. H. (1975) *Biochemistry* **14**, 5194–5199
- 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
- Hunt, D. F., Buko, A. M., Ballard, J. M., Shabanowitz, J. & Giordani, A. B. (1981) *Biomed. Mass Spectrom.* **8**, 397–408
- Kelley, J. A., Nau, H., Förster, H. J. & Biemann, K. (1975) *Biomed. Mass Spectrom.* **2**, 313–325
- Kovats, E. (1965) *Helv. Chim. Acta* **58**, 1915–1932
- Morris, H. R. (1972) *FEBS Lett.* **22**, 257–260
- Morris, H. R., Williams, D. H. & Ambler, R. P. (1971) *Biochem. J.* **125**, 189–201
- Morris, H. R., Dickinson, R. J. & Williams, D. H. (1973) *Biochem. Biophys. Res. Commun.* **51**, 247–255
- Morris, H. R., Panico, M., Barber, M., Bordoli, R. S., Sedgwick, R. D. & Tyler, A. (1981) *Biochem. Biophys. Res. Commun.* **101**, 623–631
- Nau, H., Förster, H.-J., Kelley, J. A. & Biemann, K. (1975) *Biomed. Mass Spectrom.* **2**, 326–339
- Priddle, J. D., Rose, K. & Offord, R. E. (1978) *Adv. Mass Spectrom.* **7B**, 1502–1505
- Rochat, C., Rochat, H. & Edman, P. (1970) *Anal. Biochem.* **37**, 259–267
- Rose, K. (1979) D.Phil. Thesis, University of Oxford
- Rose, K. (1983) *J. Chromatogr.* **259**, 445–452
- Rose, K., Priddle, J. D., Offord, R. E. & Esnouf, M. P. (1980) *Biochem. J.* **187**, 239–243
- Shimonishi, Y., Hong, Y.-M., Kitagishi, T., Matsuo, T., Matsuda, H. & Katakuse, I. (1980) *Eur. J. Biochem.* **112**, 251–264
- van de Graaf, B., Schuyt, P. J. W., Beyerman, H. C. & Knol-Kalkman, A. (1978) *Recl. Trav. Chim. Pays-Bas* **97**, 175–176
- Williams, D. H., Bradley, C. V., Santikarn, S. & Bojesen, G. (1982) *Biochem. J.* **201**, 105–117