Journal of Chromatography, 116 (1976) 305–314

CHROM. 8685

APPLICATIONS OF GAS–LIQUID CHROMATOGRAPHY IN PROTEIN CHEMISTRY

DETERMINATION OF C-TERMINAL SEQUENCES ON NANOMOLAR AMOUNTS OF PROTEINS

KENNETH W. M. DAVY and COLIN J. O. R. MORRIS

Department of Experimental Biochemistry, London Hospital Medical College, Queen Mary College, London E1 4NS (Great Britain)

(Received July 31st, 1975)

SUMMARY

A method for the determination of C-terminal amino acids and C-terminal amino acid sequences in nanomolar amounts of proteins is described, based on carboxypeptidase A digestion of the protein, followed by removal of the partially digested protein and quantitative gas–liquid chromatographic determination of the amino acids released after known time intervals. Sequences deduced from the kinetics of release of specific amino acids are compared with the known C-terminal sequences of well-characterized proteins.

INTRODUCTION

The techniques for amino acid sequence determination on millimolar and larger amounts of protein have now become virtually standardized, and in some cases automated, so that only minor advances are to be expected in this range. On the other hand, corresponding methods for the micromolar and nanomolar ranges have not yet been standardized to the same extent, and often rely heavily on non-quantitative methods such as the various dansyl techniques. These have marked limitations for the kinetic studies such as those involving enzymatic degradations. Although the ion-exchange method for amino acid analysis is continually being improved and its sensitivity increased, these advances are only obtained by increased complexity, cost and with very specialised instrumentation. The recent striking improvements in amino acid analysis by gas–liquid chromatography (GLC) pioneered by Darbre and Islam[1–3] and by Gehrke and co-workers[4,5] suggested to us that these methods, which are applicable to the microgram scale, might be usefully combined with scaled-down versions of well-known methods for the determination of amino acid sequences in order to extend the latter to nanomolar amounts of proteins and peptides. This paper describes the application of GLC amino acid analysis by the method of Islam and Darbre[2], using N-trifluoroacetyl methyl esters, to C-terminal sequence determination on nanomolar

amounts of protein by carboxypeptidase A digestion. The methodology of this enzymatic method on the millimolar scale has been reviewed in detail by Ambler[6,7].

MATERIALS AND METHODS

Chemicals

Acetone, acetyl chloride, dichloromethane and ammonium bicarbonate were of analytical grade. Specially dried methanol, the silicone stationary phases QF-1 and MS-200 for GLC were supplied by BDH, Poole, Great Britain. The silicone stationary phase XE-60 and the supporting phase HP Chromosorb W, 80–100 mesh, were obtained from Field Instruments, Richmond, Surrey, Great Britain.

Proteins

Carboxypeptidase A, twice recrystallised (EC 3.4.2.1), diisopropylphosphofluoridate treated, and crystalline ribonuclease A were obtained from Worthington Biochemical, Freehold, N.J., U.S.A. Lysozyme (twice recrystallised) and bovine serum albumin (recrystallised) were supplied by Koch–Light, Colnbrook, Great Britain). α-Chymotrypsin and bovine insulin (IUPAC standard) were from BDH. Chymotrypsinogen A and ribonuclease B were from Miles–Seravac, Maidenhead, Great Britain. Bovine glucagon was from Eastman–Kodak, Rochester, N.Y., U.S.A. and α-lactalbumin was from Gallard–Schlesinger Chemical Manufacturing Co., Carle Place, N.Y., U.S.A.

All proteins for digestion were freed from low-molecular-weight contaminants by passing a concentrated solution in 0.01 *M* ammonium bicarbonate through a short column of Sephadex G-50, eluting and lyphilizing. Certain proteins required oxidation before carboxypeptidase A digestion. This was carried out with performic acid at 0°, according to the method of Hirs[8].

Preparation of 4 M HCl in methanol

Acetyl chloride (14 ml) was added in 1-ml portions to specially dried methanol cooled in an acetone–solid carbon dioxide mixture. The methanol was allowed to warm up to room temperature after each addition of acetyl chloride, and again cooled before the next addition. The solution was finally made up to 50 ml with dried methanol.

Carboxypeptidase A digestion

A suitable aliqout of the commercial suspension of the enzyme was centrifuged, and the solid suspended in the same volume of water and re-centrifuged. The solid enzyme was suspended in an appropriate volume of 2 *M* ammonium bicarbonate and left for 30 min until the enzyme was in solution. An aliquot of the enzyme solution was added to a solution or suspension of the substrate protein (10–70 nmoles) in 0.2–0.5 ml of 0.01 *M* ammonium bicarbonate containing norleucine (1 mole per calculated mole of substrate protein) as internal standard. The final enzyme:substrate ratio (w/w) was usually 0.02–0.01, but higher dilutions down to 0.001 were used when preliminary experiments indicated that the initial reaction was too rapid. Actual ratios used are shown in the legends to the figures. The final ammonium bicarbonate concentration was in the range 0.025–0.05 *M*. Digestion was carried out at 37° for up to 6 h, samples being removed for analysis usually at 5-, 15-, 30-, 60-, 120-, 240- and

360-min intervals. The reaction was terminated in each sample by the addition of 40% (v/v) acetic acid, bringing the pH to 3–4. The sample was cooled to 0°, and cooled redistilled acetone added to a final acetone concentration of 75% (v/v). The mixture was cooled to 0° and stored at that temperature for 2–3 h. It was then centrifuged and the supernatant liquid evaporated to dryness under reduced pressure.

Chromatographic separation of amino acids as N-trifluoroacetyl methyl esters

A Pye Series 104 gas chromatograph fitted with a wide range amplifier and a dual flame ionisation detector system was used throughout this work. Peak areas were measured with an Autolab Model 6300 digital integrator with a printed output expressed in counts/sec. Operating conditions were as follows:

Column temperature, initial 80°, final 210°; detector temperature, 250°; injection port temperature, 200°; rate of temperature increase, 2°C/min (programmed); nitrogen carrier flow-rate, 25 ml/min; hydrogen flow-rate, 30 ml/min; air flow-rate, 400 ml/min.

Column preparation. The mixed stationary phase (46% XE-60, 27% QF-1, 27% MS-200, w/w/w) was kept as a 1% (w/v) solution in dichloromethane. HP Chromosorb W, 80–100 mesh, (9.75 g) was heated at 200° overnight. The cold supporting phase was transferred to a large evaporation basin, and just covered with dichloromethane. 25 ml of the stationary phase solution were added, and the mixture gently swirled while the solvent was slowly removed by blowing air over the surface of the fluid. The still damp supporting phase was transferred to a flask and the remainder of the solvent removed under reduced pressure at 100°. The coated supporting phase was packed into two glass columns (3 m × 2.4 mm I.D.), leaving the first 7 cm of the columns empty. The columns were plugged with silanised glass thread and conditioned in the chromatograph at 210° for 24 h.

Derivatisation and chromatography. The sample residue, containing up to 10 μg of total amino acids, was dissolved in 4 *M* anhydrous methanolic hydrochloric acid (0.2 ml) and heated at 70° for 90 min in a PTFE-stoppered tube. The solution was then evaporated to dryness under reduced pressure at 80°. Trifluoroacetic anhydride (0.2 ml) was then added to the dry residue of amino acid methyl ester hydrochlorides, and allowed to react for 30 min at room temperature. The reaction mixture was then transferred to a small pre-column (7 cm × 1.5 mm I.D.) packed with HP Chromosorb W. The excess solvent was removed under reduced pressure at room temperature during 10 min. The pre-column was then introduced into the main separation column as described by Darbre and Islam[3], and the subsequent GLC separation of the N-trifluoroacetyl amino acid methyl esters carried out as described by these authors[2].

Relative molar responses (RMR) of the protein amino acids were determined relative to norleucine by individual derivatization and transfer by the pre-column methods. Responses were measured as integrated counts·sec^{-1}·μmole^{-1} for each amino acid.

$$\text{RMR of amino acid} = \frac{\text{counts}\cdot\text{sec}^{-1}\cdot\mu\text{mole of amino acid}^{-1}}{\text{counts}\cdot\text{sec}^{-1}\cdot\mu\text{mole of norleucine}^{-1}}$$

Mean values of the RMR for the protein amino acids based on five replicate determinations are given in Table I together with the retention temperatures under the specified conditions. The retention temperatures were also checked with the

TABLE I

RELATIVE MOLAR RESPONSES OF THE PROTEIN AMINO ACIDS WITH REFERENCE TO NORLEUCINE

All amino acids were taken through the complete derivatization and chromatographic procedure. Values are means of five determinations.

Amino acid	*Retention temperature (°C)*	*RMR*
Ala	93	0.46
Val	96	0.80
Gly	98	0.25
Ile	103	0.85
Thr	105	0.75
Leu	106	1.00
Nle	110	1.00
Ser	114	0.55
Pro	117	0.86
Asp	124	0.64
Met	140	0.71
Glu	141	0.81
Phe	143	1.65
Tyr	167	1.17
Lys	191	0.80
Try	193	1.40
Arg	201	—
Met-SO_2	206	—

reference set supplied by BDH. A number of non-protein amino acids as well as the hexosamine derivatives N-acetylglucosamine and N-acetylgalactosamine were also resolved by the mixed silicone stationary phase GLC column. Retention temperatures for these are given in Table II. Asparagine and glutamine are converted into aspartic and glutamic acid trifluoroacetyl methyl esters during the derivatization process, so that no distinction between the amides and free acids is possible by our method.

TABLE II

RETENTION TEMPERATURES OF CERTAIN NON-PROTEIN AMINO ACIDS AND AMINO SUGAR DERIVATIVES RESOLVED BY THE MIXED SILICONE STATIONARY PHASE GLC COLUMN

Amino acid/sugar	*Retention temperature (°C)*
Abu	95
Nva	103
β-Ala	106
γ-Abu	126
Cys–Cys	129
Hyp	137
Homo-Cys	142
Aspr	163
N-acetylglucosamine	180
Orn	184
N-acetylgalactosamine	187

RESULTS AND DISCUSSION

C-Terminal amino acid residue determinations

In order to test the method for the presence of contaminants (which need not necessarily be amino acids) introduced prior to GLC, it was applied to two proteins, chicken lysozyme (C-terminal sequence Arg–Leu[9]) and α-lactalbumin (C-terminal sequence Lys–Leu[10]) in which the action of carboxypeptidase A is virtually limited to the release of the C-terminal amino acid by the presence of a basic amino acid residue in the penultimate position.

As shown in Fig. 1, leucine only was released in both cases, 1.0 mole per mole of protein in the case of α-lactalbumin, and 0.74 mole per mole in the case of performic acid-oxidised lysozyme. No leucine was released from unoxidised lysozyme. The lower yield in the case of lysozyme probably reflects the greater three-dimensional complexity of this protein, even after oxidation. No contaminant peaks were observed in either case, and the absence of contaminants was also confirmed by control experiments carried out in the absence of substrate proteins in the digestion mixture.

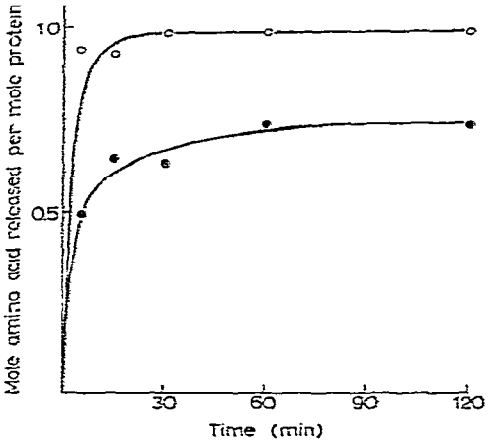

Fig. 1. Release of C-terminal leucine from (○) α-lactalbumin and (●) oxidised chicken lysozyme. In both cases enzyme: substrate ratio 1/100, substrate concentration 0.5% (w/v) buffer 0.2 *M* ammonium bicarbonate.

C-Terminal amino acid sequence determinations

Ribonuclease A and ribonuclease B. The C-terminal sequences of these two proteins which differ only in the presence of a carbohydrate group in ribonuclease B[11,12] were compared in order to ascertain whether our general method was applicable to glycoproteins. As shown in Figs. 2a and 2b, no differences were detected between the two proteins, and the correct C-terminal sequence –Phe–Asp–Ala–Ser–Val was confirmed in both cases. No interference from the carbohydrate was observed in the case of ribonuclease B, although the yields of C-terminal amino acids were lower in this case, possibly reflecting the somewhat lower purity of the commerical preparation used.

Bovine serum albumin. The correct C-terminal sequence of this protein is still

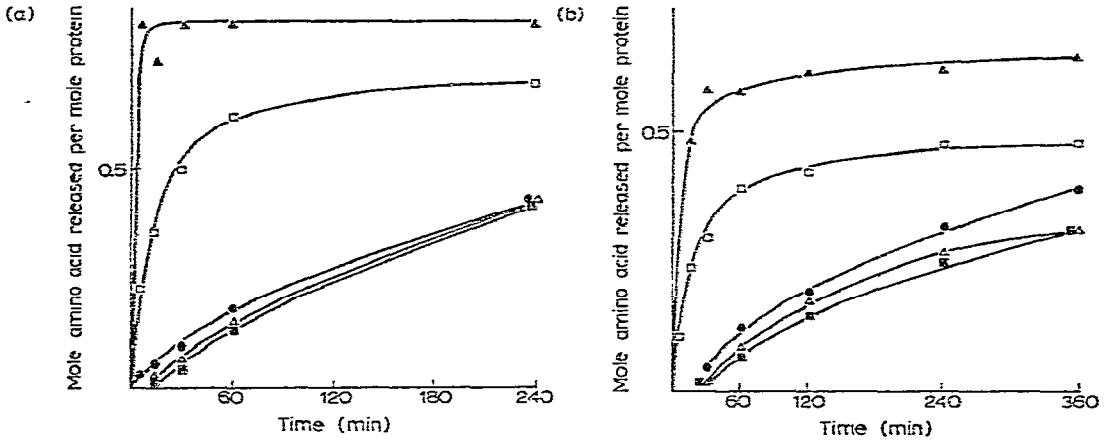

Fig. 2. Release of C-terminal amino acids from (a) ribonuclease A, (b) ribonuclease B. Enzyme: substrate ratio 1/130, substrate concentration 0.34% (w/v, ribonuclease A) and 0.37% (w/v, ribonuclease B), buffer 0.025 *M* ammonium bicarbonate. ▲, Val; □, Ser; ●, Ala; △, Asx; ■, Phe.

open to discussion. King and Spencer[13] determined the complete amino acid sequence of a 36 residue C-terminal peptide obtained by them by cyanogen bromide fission of the intact protein. They employed chemical, mainly Edman degradation, methods. They also noted that the preparation of bovine serum albumin used by them was not homogeneous, containing two components, respectively 70% and 30% of the whole, by isoelectric focusing. The relevant portion of the C-terminal sequence was –Lys–Leu–Val–Val–Ser–Thr–Gln–Ala–Leu–Ala.

Peters and Hawn[14] used carboxypeptidase A to determine the C-terminal sequence of a C-terminal fragment of the intact protein obtained by them by peptic digestion. Their results are consistent with the C-terminal sequence (Gly, Ser, Val)–Thr–Ala–Leu–Ala.

Our own investigations (Fig. 3) support the C-terminal sequence –Ser–Val–

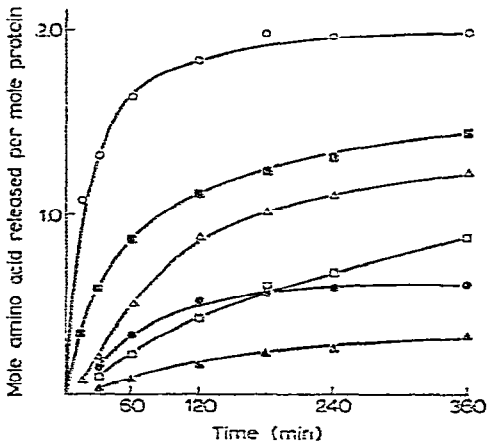

Fig. 3. Release of C-terminal amino acids from bovine serum albumin. Enzyme: substrate ratio 1/100, substrate concentration 0.16% (w/v), buffer 0.04 *M* ammonium bicarbonate. ○, Ala; ■, Leu; △, Thr; ●, Glx; □, Val; ▲, Ser.

(Val)–Glx–Thr–Ala–Leu–Ala in complete agreement with Peters and Hawn[14] for the last four residues, but only in partial agreement with King and Spencer[13]. However, only we and the latter investigators have detected glutamic acid (or glutamine) in the C-terminal portion of this protein, although there have been several earlier studies using various carboxypeptidase preparations.

Two commercial preparations (Koch–Light and Armour) gave identical results by our method, but most likely the observed disagreement is essentially due to differences (and heterogeneity) in the preparations of this poorly characterised protein. This is confirmed by reported differences in N-terminal residues, and by various electrophoretic homogeneity studies, although the latter are complicated by the capacity of serum albumin to bind fatty acids.

Bovine glucagon. The complete amino acid sequence of this polypeptide hormone has been determined by Bromer *et al.*[15]. The relevant C-terminal sequence is –Ala–Gln–Asp–Phe–Val–Gln–Trp–Leu–Met–Asn–Thr.

Our carboxypeptidase A kinetic study is consistent with the sequence (Fig. 4) –Ala–Asx–Phe–Glx–(Val,Trp,Leu,Met)–Asx–Thr which is in reasonable agreement with the correct sequence, but has the undetermined section (Val,Trp,Leu,Met). Our subsequent investigations have shown that our preparation of carboxypeptidase A releases methionine residues very slowly, so that the sequencing of earlier amino acid residues may become ambiguous, and depend markedly on individual amino acid release kinetics. Phenylalanine is also released slowly by our enzyme preparation. It is however interesting to observe that the alanine residue, ten in the C-terminal residue can be detected by the carboxypeptidase A–GLC method.

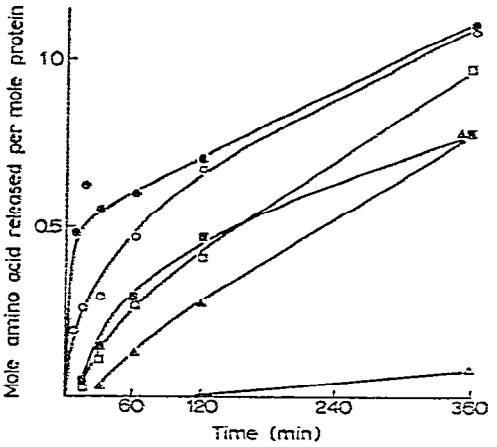

Fig. 4. Release of C-terminal amino acids from bovine glucagon. Enzyme: substrate ratio 1/100, substrate concentration 0.04% (w/v), buffer 0.025 *M* ammonium bicarbonate. ●, Thr; ○, Asx; ■, Trp, Val, Leu, Met; □, Glx; ▲, Phe; △, Ala.

Bovine insulin. This dual-chain protein has the two C-terminal sequences –Lys–Ala and –Cys–Asn (ref. 16), in which both penultimate residues are resistant to carboxypeptidase A hydrolysis, so that only alanine and aspartic acid should be detected by our method. This expectation is confirmed in 0.02 *M* ammonium bicar-

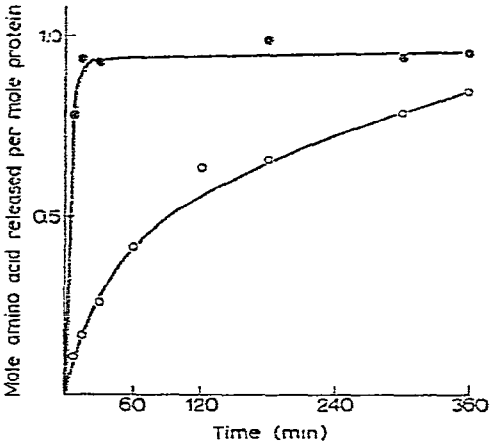

Fig. 5. Release of C-terminal amino acids from bovine insulin. Enzyme: substrate ratio 1/50, substrate concentration 0.16% (w/v), buffer 0.02 *M* ammonium bicarbonate. ●, Ala; ○, Asx.

bonate buffer (Fig. 5), while in 0.2 *M* ammonium bicarbonate buffer, alanine only is released.

α-Chymotrypsin and chymotrypsinogen A. This pair of proteins (for structures see Hartley[17], and Hartley and Kauffmann[18]) were studied by the carboxypeptidase A–GLC method in order to determine whether the new C-terminal residues formed during the activation reaction could readily be detected in a complex system on the nanomolar scale.

The kinetics of release of amino acids during the carboxypeptidase A digestion of chymotrypsinogen A were consistent with the sequence (Fig. 6) –Val–($Glx_{1.5}$,Thr)–Leu–(Ala_2,Asx) which although in reasonable agreement with the known sequence –Val–Gln–Gln–Thr–Leu–Ala–Ala–Asn has two ambiguous sections, and illustrates

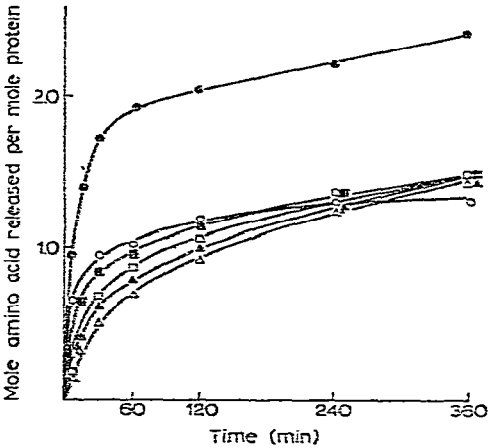

Fig. 6. Release of C-terminal amino acids from oxidised chymotrypsinogen A. Enzyme: substrate ratio 1/300, substrate concentration 0.63% (w/v), buffer 0.025 *M* ammonium bicarbonate. ●, Ala; ○, Asx; ■, Leu; □, Glx; ▲, Thr; △, Val.

the difficulties which may arise in attempts at carboxypeptidase sequencing when two identical amino acid residues (Gln_2 and Ala_2) follow a single different residue, especially as asparagine was somewhat slowly released by our preparation of carboxypeptidase A. Even an enzyme:substrate ratio of 0.001 did not enable the correct sequence to be determined unambiguously, although the kinetics of release of threonine did vary with this ratio.

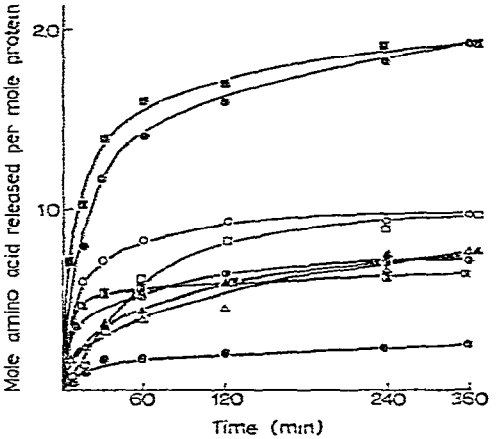

Fig. 7. Release of C-terminal amino acids from oxidised α-chymotrypsin. Enzyme: substrate ratio 1/1000, substrate concentration 0.63% (w/v), buffer 0.025 *M* ammonium bicarbonate. ●, Ala; ○, Asx; ■, Leu; □, Glx; ▲, Thr; △, Val; ⊖, Ser; ◑, Gly; ◧, Tyr.

The C-terminal sequence analysis of α-chymotrypsin (Fig. 7) shows that two new amino acid residues, a second leucine and a tyrosine have been superimposed on the pattern of Fig. 6, as well as the delayed appearance of serine and glycine, suggesting a C-terminal pattern such as

$$\text{–Ser–Gly}\begin{cases}\text{Leu}\\ \\ \text{Tyr}\end{cases}$$

–Val–(Glx,Thr)–Leu–(Ala_2,Asx)

in comparison with the actual C-terminal pattern of this three-chain protein:

$$\begin{array}{l}\text{–Ser–Gly–Leu}\\ \text{–Arg–Tyr}\end{array}$$

–Val–Gln–Gln–Thr–Leu–Ala–Ala–Asn

A difference analysis of this kind requires less than 500 μg of protein in all for the complete kinetic analysis, and it may be of special value for the quantitative identification of new C-terminal amino acid residues uncovered by activation or dissociation phenomena.

The essential problem to be solved in this investigation was the quantitative recovery of at most 10 μg of amino acids from 10–20 times that amount of protein, as preliminary studies had shown that introduction of the complete hydrolysis mixture on to the GLC column resulted in unacceptable blanks due to pyrolysis of the protein. Several alternative methods of deproteinization were examined, including

microgram-scale ion-exchange chromatography and hollow-fibre dialysis. The simple acetone precipitation method described however gave the best recovery of amino acids provided the final acetone concentration did not exceed 75% (v/v). Table III shows that appreciable losses of the acidic amino acid amides occurred at higher acetone concentrations. Fortunately this acetone concentration was sufficiently high to eliminate interference with the subsequent GLC amino acid analysis.

TABLE III

RECOVERY OF ASPARAGINE AND GLUTAMINE FROM ACETONE PRECIPITATION AT VARIOUS ACETONE CONCENTRATIONS

Acetone concentration (v/v)	*Recovery (%)*	
	Asn	*Gln*
88.5	29	57
80	54	77
75	93	95

Subsequent to this work similar experiments were initiated with leucine aminopeptidase (EC 3.4.1.1) to investigate whether similar methods could be employed for N-terminal sequence analysis. Unfortunately although the hydrolysis, working-up, derivatization and GLC analysis presented no new problems, the presence of endopeptidase contaminants in both commercial leucine aminopeptidase preparations examined rendered the results untrustworthy. We are currently investigating methods of elimination or inactivation of such endopeptidase contaminants.

Although more sensitive detector systems such as the electron capture or conductivity detectors might be employed with this enzymatic method to give a 10–100 fold increase in sensitivity, the reliability of the flame ionisation detector recommends it for multiple kinetic analysis of this type.

REFERENCES

1 A. Darbre and A. Islam, *Biochem. J.*, 106 (1968) 923.
2 A. Islam and A. Darbre, *J. Chromatogr.*, 43 (1969) 11.
3 A. Darbre and A. Islam, *J. Chromatogr.*, 49 (1970) 293.
4 C. W. Gehrke, K. Kuo and R. W. Zumwalt, *J. Chromatogr.*, 57 (1971) 209.
5 C. W. Gehrke and K. Leimer, *J. Chromatogr.*, 57 (1971) 219.
6 R. P. Ambler, *Methods Enzymol.*, 11 (1967) 155, 436.
7 R. P. Ambler, *Methods Enzymol.*, 25 (1972) 143, 263.
8 C. H. W. Hirs, *Methods Enzymol.*, 11 (1967) 197.
9 R. E. Canfield, *J. Biol. Chem.*, 238 (1962) 2698.
10 K. Brew, T. C. Vanaman and R. L. Hill, *J. Biol. Chem.*, 242 (1967) 3747.
11 D. G. Smyth, W. H. Stein and S. Moore, *J. Biol. Chem.*, 238 (1963) 229.
12 T. H. Plummer and C. H. W. Hirs, *J. Biol. Chem.*, 239 (1964) 2530.
13 T. P. King and E. M. Spencer, *Arch. Biochem. Biophys.*, 153 (1972) 627.
14 T. Peters and C. Hawn, *J. Biol. Chem.*, 242 (1967) 1566.
15 W. W. Bromer, A. Staub, E. R. Diller, H. L. Bird, L. G. Sinn and O. K. Behrens, *J. Amer. Chem. Soc.*, 79 (1957) 2794.
16 A. P. Ryle, F. Sanger, L. F. Smith and R. Kitai, *Biochem. J.*, 60 (1955) 541.
17 B. S. Hartley, *Nature (London)*, 201 (1964) 1284.
18 B. S. Hartley and D. L. Kauffmann, *Biochem. J.*, 101 (1966) 229.