Biochem. J. (1974) **141**, 701–713
Printed in Great Britain

A Mass-Spectrometric Sequence Study of the Enzyme Ribitol Dehydrogenase from *Klebsiella aerogenes*

By HOWARD R. MORRIS and DUDLEY H. WILLIAMS
University Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.
and
GRAEME G. MIDWINTER* and BRIAN S. HARTLEY†
M.R.C. Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, U.K.

(*Received* 5 *March* 1974)

The first detailed results of the application of a low-resolution mixture analysis approach to the sequence analysis of an enzyme, ribitol dehydrogenase, are given. Examples of the interpretation of the spectra of peptide mixtures derived from this protein are described. Evidence for new fragmentation patterns observed is reported, together with an explanation of the generation of ambiguous sequences by use of a low-specificity enzyme, thermolysin. The overall sequencing strategy evolved is assessed.

In recent years, there has been a continued interest in protein sequencing by mass spectrometry, despite the lack of general application of the technique to genuine structural problems. Critics have been able, justifiably, to point to the relatively poor sensitivity compared with the highly developed classical approaches, and to the inability to handle certain naturally occurring amino acids effectively. Despite these problems, the mass-spectrometric method has obvious potential, and in our view the particular strategy with which the method is used is of prime importance, since this governs the time and effort spent, and the reliability of the sequence information produced. For example, little can be gained by following the conventional isolation and purification procedures and then applying mass-spectrometric methods to the single pure peptides obtained. For this reason, we have been developing the concept of 'mixture analysis', the potential of which was first indicated in the work on the protein silk fibroin (Geddes *et al.*, 1969; Morris, 1970). Deliberate use of this strategy of mixture analysis in sequence studies (Morris *et al.*, 1971) optimizes peptide yields by minimizing isolation and purification steps, and lessens the effort required.

Another variable in the mass-spectrometric method is the choice of instrument resolving power. Some groups have chosen to operate the mass spectrometer in the high-resolution mode, as this gives valuable information on the elemental composition of ions in the spectrum, thus facilitating interpretation. More recently, and after preliminary statements of the results of our work (Morris *et al.*, 1971; Morris, 1972; Hartley *et al.*, 1972) it has been argued that high-resolution mass spectrometry is essential to the analysis of peptide mixtures, and that many ambiguities are presented if only low-resolution analysis is undertaken (Wipf *et al.*, 1973). We now produce detailed evidence and results which refute this criticism.

In addition to our own studies on peptide mixtures derived from proteins, several hypothetical mixtures of known synthetic peptides have been studied (Roepstorff *et al.*, 1971; Wipf *et al.*, 1973). Although the results of these studies have been presumed to be directly relevant to protein-sequence analysis, the composition of such mixtures has been necessarily artificial, and the sequences of the component peptides were known beforehand.

The study presented here has been a logical development from the previous work on individual unknown peptides and peptide mixtures; it was designed not as an attack on the total sequence analysis of a protein by mass spectrometry, but rather to assess the problems and optimize the speed and accuracy with which our strategy could be used in protein sequence analysis. The protein chosen for this study was the enzyme ribitol dehydrogenase from *Klebsiella aerogenes*, which is a tetramer of four identical subunits each of mol.wt. 27000 (Taylor *et al.*, 1974). This enzyme has been the subject of an 'enzyme-evolution' study (Hartley *et al.*, 1972; Hartley, 1974) and its complete sequence has recently been determined by classical techniques (C.H. Moore, S. S. Taylor, M. J. Smith & B. S. Hartley, unpublished work). The mass-spectrometric study reported here was carried out simultaneously with the latter part of the classical study, and without knowledge of the partial sequence previously obtained.

* Present address: Department of Chemistry and Biochemistry, Massey University, Palmerston North, New Zealand.

† Present address: Department of Biochemistry, Imperial College, London SW7 2AZ, U.K.

Experimental

Ribitol dehydrogenase was purified from super-producing strains of *Klebsiella aerogenes* by the method of Taylor *et al.* (1974). Treatment with iodoacetic acid and other methods of protein chemistry were as described by Shotton & Hartley (1973).

Protein digestion and peptide fractionation

Peptides were fractionated from several different digests of native or [^{14}C]carboxymethyl-ribitol dehydrogenase as follows.

Tryptic digest of carboxymethyl-ribitol dehydrogenase (digest T). Trypsin (5.0mg) was incubated with diphenylcarbamoyl chloride (2μg) in 0.1ml of 0.05M-Tris–HCl, pH7.5, and diluted to 5ml with 0.2M-NH_4HCO_3. After centrifugation, 1ml of this (1mg) was added to 112mg of carboxymethyl-ribitol dehydrogenase in 8ml of 0.1M-NH_4HCO_3. After 4h at 37°C the digest was freeze-dried, suspended in 2.5ml of pyridine–acetate buffer, pH3.1, (16.1ml of pyridine and 278ml of acetic acid/litre; 0.2M in pyridinium ion) and centrifuged. The supernatant was applied to a column (50cm×0.9cm diam.) of polystyrene sulphonic acid (Locarte L2 amino acid analyser resin) and chromatographed at 20ml/h with a linear gradient of pyridinium ion between 0.2M, pH3.1, and 1.6M, pH5.0 (161ml of pyridine, 143ml of acetic acid per litre) in a mixing device containing 356ml of pH5.0 buffer running into 178ml of pH3.1 buffer, with cross-sectional areas of the vessels in the ratio 2:1 (Shotton & Hartley, 1973). The effluent was stream split, 2% being analysed by automatic alkali hydrolysis and ninhydrin development, and the rest collected in a fraction collector. Samples from relevant tubes were screened by high-voltage paper electrophoresis at pH6.5 (Shotton & Hartley, 1973) and appropriate fractions were pooled and freeze-dried.

Cyanogen bromide cleavage of insoluble tryptic peptides (digest TM). The insoluble residue from digest T was dissolved in 10ml of 70% (v/v) formic acid and treated with CNBr (40mg) for 24h at room temperature, 22°C. The digest was evaporated *in vacuo*, diluted with 10ml of water and evaporated again. After being dissolved in 6ml of 8M-urea containing 0.1M-acetic acid it was chromatographed on a Sephadex G-50 (fine grade) column (83cm×2.5cm diam.) in the same solvent, yielding six crude fractions of decreasing molecular weight. Fractions IV, V and VI were desalted on a Sephadex G-25 column (8.3cm×2.5cm diam.) which was eluted with 0.1M-acetic acid. The desalting procedure was repeated, after which fraction IV was partially resolved into fractions IVa, IVb, and IVc.

Peptic digest of fractions TMIVa, TMIVb, TMIVc and TMVI (digest TMP). Samples of each of these fractions, after being freeze-dried, were dissolved in 1ml of 5% (v/v) formic acid containing 4μg of pepsin and incubated for 2h at 37°C. Peptides from fractions TMIVc and TMVI were partially purified by high-voltage electrophoresis at pH6.5.

Thermolysin digest of carboxymethyl-ribitol dedrogenase (digest L). Carboxymethyl-ribitol dehydrogenase (140mg) was incubated with thermolysin (3mg) and $CaCl_2$ (20mg) in 10ml of 0.2M-NH_4HCO_3 for 7h at 37°C. The digest was centrifuged to remove traces of insoluble material and fractionated on a Sephadex G-25 column (75cm×2.5cm diam.) in 0.2M-NH_4HCO_3. Samples of each fraction were monitored by high-voltage paper electrophoresis at pH6.5 and fractions were pooled into four groups in order of increasing molecular weight. Group I and group III fractions were each chromatographed in the 'peptide analyser' as described for digest T. By screening samples of the effluents by high-voltage paper electrophoresis at pH6.5 and 3.5 (Shotton & Hartley, 1973), tubes could be pooled to contain simple mixtures suitable for mass spectroscopy.

Chymotryptic digest of maleylated carboxymethyl-ribitol dehydrogenase (digest C). Carboxymethylated-ribitol dehydrogenase (370mg) in 30ml of 8M-urea was treated at 2°C with 1.7g of maleic anhydride added in small batches over 30min, the pH being kept at 9.0±0.5 by additions of 2M-NaOH. The maleylated protein was dialysed for 24h against three changes of 3 litres of water, freeze-dried, and redissolved in 10ml of 0.2M-NH_4HCO_3. Chymotrypsin (10mg) and soya-bean trypsin inhibitor (1mg) were added, and after 3h at 37°C the digest was separated on a Sephadex G-50 column (83cm×2.5cm diam.) eluted with 0.2M-NH_4HCO_3. Five fractions, in order of increasing size, were pooled and freeze-dried. Fractions IV and V (low molecular weight) were each chromatographed in the 'peptide analyser' as described for digest T. The effluents from these columns were screened by high-voltage paper electrophoresis at pH6.5 as described above and peptides staining for histidine or arginine were pooled and purified further by high-voltage paper electrophoresis. The remaining peptides were pooled into simple mixtures for mass-spectrometric analysis.

Exhaustive chymotryptic digestion of insoluble tryptic peptides (digest TC). The residue, which was insoluble at pH3.1 from a tryptic digest of 300mg of carboxymethyl-ribitol dehydrogenase (see digest T), was treated with chymotrypsin (3mg) for 24h at 37°C in 0.2M-NH_4HCO_3. After being freeze-dried, the sample was extracted with the pyridine–acetate buffer, pH3.1, described above, and half of the supernatant was chromatographed in the peptide analyser as described for digest T. Fractions were examined analytically by high-voltage paper electrophoresis at pH6.5, and those staining for histidine or arginine were further purified at pH6.5 or 3.5.

Preparation of derivatives

Most of the work described in this paper was carried out by using the procedures described previously (Morris *et al.*, 1971). Although this was the best method known at the time of examination, we were aware of its inapplicability to peptides containing histidine, methionine, cysteine and/or arginine. As a consequence, much of the sequence information available was unavoidably forfeited in the early part of this study. However, new derivative-formation procedures were applied to the later work (Morris, 1972; Morris *et al.*, 1973) and after this, a number of the peptides containing histidine, methionine, carboxymethylcysteine and arginine were sequenced.

Sample handling, temperature gradient and partial-fractionation techniques have been described previously (Morris *et al.*, 1971).

Mass spectrometry

Mass spectra were recorded on either AEI–GEC MS902 or MS12 instruments operating at a resolving power of 1200. An electron beam energy of 70eV and an accelerator voltage of 8kV were used. High-molecular-weight fragments were brought into focus at the collector slit by decreasing the accelerator voltage.

Nomenclature

For describing the amino acid sequences deduced from the mass spectra the normal three-letter code for an amino acid is used.

For describing the analysis of mass spectra or the fragmentation pattern of a peptide derivative the following symbols for amino acid derivatives are used: (i) Ac(Me)$_2$Ser means *N*-acetyl-*NO*-dimethylserine; (ii) (Me)$_2$Ser means *NO*-dimethylserine; (iii)... (Me)Tyr corresponds to a specific N–C cleavage on the *N*-terminal side of tyrosine, with hydrogen transfer yielding a fragment not containing the nitrogen atom of tyrosine.

Results

The large amount of sequence information obtained during this study precludes a description of the mass spectrum of each peptide sequenced. For brevity, the sequences deduced are presented in a tabulated form under the digest headings. Some of the digests (e.g. the chymotryptic digest) were examined fairly exhaustively, whereas others (e.g. the thermolysin digest) were only briefly examined (for reasons given in the Discussion section). Peptide spectra showing a noteworthy fragmentation pattern are either described or presented in the relevant sections, together with examples of the interpretation of the spectra of peptide mixtures.

Table 1. *Sequences derived from peptides in digest T*

See the text for further details.

Phe-Ala-Val-Gln-Ala
Phe-Ala
Ser-Gly-Leu-Gly-Leu-Glu
Leu-Asp-Leu-Phe
Asn-Thr-Ser-Leu
Glu-Gly-Glu-Lys
Thr-Leu-Leu-Gly-Ala-Gly-Ala-Lys
Asn-Val-Thr-Val

Digest T

Pooled samples from the ion-exchange column were acetylated and permethylated as described above, and examined directly by mass spectrometry. Sequences obtained are given in Table 1. Several of the samples were peptide mixtures, and indeed all the samples were impure as might be expected after only one chromatographic separation. Peptide yields, as judged by the ion-exchange data and the signal-to-noise ratio of mass spectra, suggested that the protein had only been partially digested, giving too few small fragments for a favourable mass-spectrometric examination. Of these, only those lacking the problem amino acids were amenable to full sequencing at this stage.

Digest TM

Samples of fractions IVa, IVb, IVc and VI were acetylated, permethylated and examined in the usual manner. The sequence data obtained are given in Table 2(*a*). Both fractions IVa and IVc gave no sequence information, presumably because the component peptides were too large and/or involatile.

Peptide IVb gave a spectrum (Fig. 1), which we interpret as being derived from the sequence Asn-Ala-Phe-Ala-Leu-Gln, at a higher source temperature than would have been expected for this sequence (290°C). Attention is drawn to the signals at *m/e* 126 and 98. These would normally be interpreted as the derivative of the *N*-terminal cyclization product of glutamic acid or glutamine (Scheme 1) (Agarwal *et al.*, 1969; Morris *et al.*, 1971).

However, we have examined a large number of peptides, of both known and unknown structure, having this *N*-terminus and in every case the activation energy for the decomposition of *m/e* 126 → *m/e* 98 (loss of CO) is very low, and the *m/e* 126 species is thus of very low abundance in the spectra. It is therefore noteworthy that the *m/e* 126 species is of appreciable abundance in Fig. 1, thus suggesting a different structure for this ion. Derivative formation of sample IVb with deuteromethyl iodide shifted the *N*-terminal signal to *m/e* 132, thus showing the presence of two methylatable sites in the structure

Table 2. *Analysis of digests TM and TMP*

'One paper clean up' indicates that the corresponding fractions were purified by electrophoresis at pH6.5. For details see the text.

(*a*) Sequences derived from peptides in digest TM:

Peptide fraction	Method of analysis	Sequence deduced
IVa		—
IVb		Asn-Ala-Phe-Ala-Leu-Gln
IVc	Direct examination	—
V		Leu-Val-Leu-Leu-Pro-Asn-Ser
VI		Leu-Val-Ala-Glu-Leu
		Val-Gly-Ala-Val-Leu-Pro-Gly-Pro-Val
		Asn-Ala-Phe-Ala-Leu-Gln

(*b*) Sequences derived from peptides in digest TMP:

IVa	Direct examination	Asn-Leu-Asn-Ala-Ala
		Leu-Gln-Gly-Leu-Leu
IVb	Direct examination	Leu-Gly-Gln-Asn
		Leu-Ala-Gln-Lys
IVc	One paper clean up	Asn-Leu-Asn-Ala
		Leu-Val-Leu-Leu
		Phe-Ala-Val
		Pro-Val-Tyr
		Asn-Val-Thr-Val
VI	One paper clean up	Val-Asp-Leu-HSe
		Leu-Val-Ala-Glu-Leu
		Val-Gly-Ala-Val-Leu-Pro-Gly-Pro-Val
		Leu-Val-Leu-Leu-Pro
		Val-Thr
		Asn-Val-Thr-Val

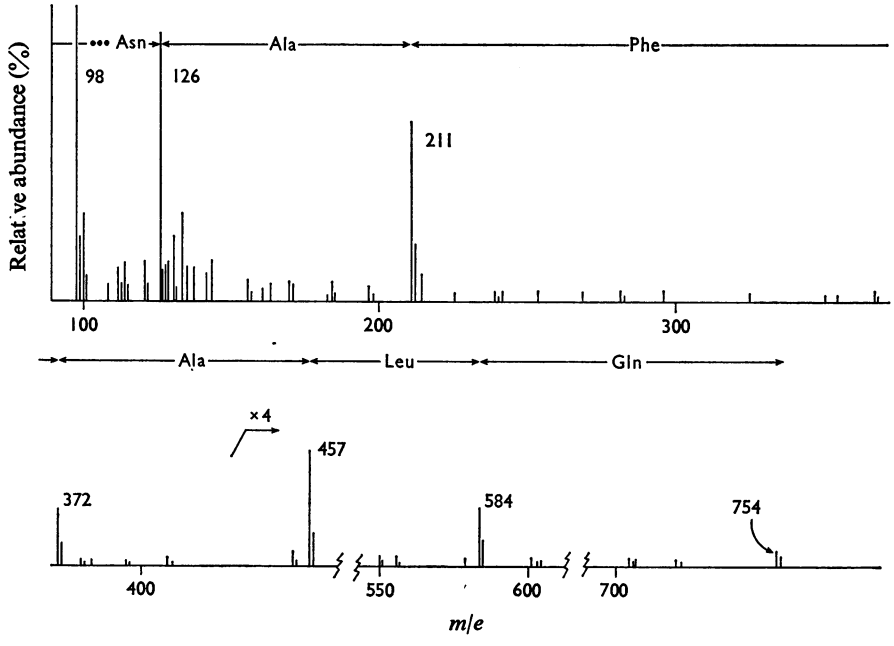
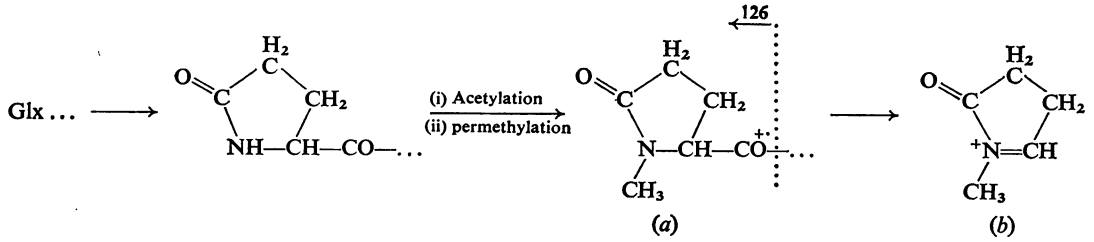
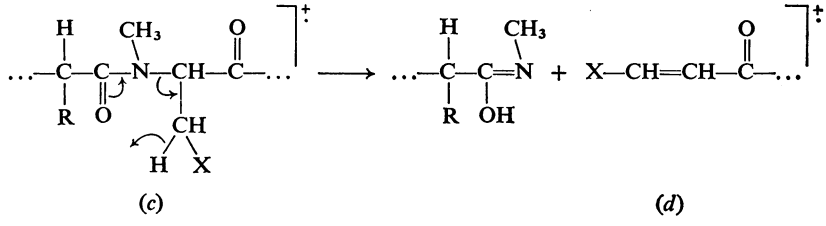


Fig. 1. *Partial mass spectrum, above m/e 95, of a peptide isolated from digest TM*

Source temperature 290°C. For clarity, less abundant signals are increased by a suitable factor, e.g. ×4. For details of interpretation see the text.



Scheme 1. *Normal derivation of m/e 126 and 98*



Scheme 2. *Newly observed N–C cleavages*

¬ means charge not localized. (*d*), *m/e* 126 [Asn, X = $CON(CH_3)_2$]; *m/e* 113 (Asp, X = $COOCH_3$); *m/e* 184 (Trp, X = Me-indole).

giving rise to *m/e* 126. We assign this fragment to the product of N–C cleavage with hydrogen transfer at asparagine (Scheme 2), which we draw thus. . .$(Me)_2$-Asn.

N–C cleavages have been observed previously in the spectra of acetylated and esterified peptides containing phenylalanine, tyrosine, tryptophan and histidine (Shemyakin *et al.*, 1966) and in permethylated peptides containing tyrosine (Morris *et al.*, 1971), phenylalanine (A. Dell, H. R. Morris, W. V. Shaw & B. S. Hartley, unpublished work), and histidine (Morris, 1972). The cleavage has not previously been reported as occurring at asparagine, or at aspartic acid or tryptophan in permethylated peptides (see below). Although peptide IVb was examined as an unknown, no difficulty was encountered in deducing the sequence given, once the origin of the signal at *m/e* 126 had been ascertained. Fraction VI yielded three major sequences, one of which appeared to be a thermal breakdown product showing the same Asn-Ala-Phe-Ala-Leu-Gln sequence as in fraction IVb.

Digest TMP

Mass spectrometry on digest TM indicated that much of the material was of high molecular weight, and therefore some samples were treated with pepsin to form smaller fragments amenable to mass spectrometry. A considerable amount of sequence data was revealed by this non-specific digest, again bearing in mind our inability to sequence peptides containing histidine, methionine, arginine or cysteine at this stage in the work. Results are presented in Table 2(*b*).

Digest L

Because of the limitations (imposed by volatility) on the size of the peptides that can be sequenced by the mass-spectrometric method, it was decided to make use of a relatively non-specific enzyme, thermolysin, in an effort to generate large numbers of small fragments. This enzyme has been used successfully in classical approaches, but unsuspected difficulties arose from its application in our sequencing strategy. As a direct consequence of the specificity of the enzyme serious difficulties arose in extracting unambiguous sequences from the peptide mixtures examined (see the Discussion section for a detailed explanation). These difficulties were exacerbated by the large number of component peptides, since we were looking at a whole-protein digest. Because of these problems, the thermolysin digest was abandoned as a method of choice, but not before some useful unambiguous sequences had been obtained (Table 3).

Digest C

This digest proved to be one of the most successful from the mass-spectrometric viewpoint for three

Table 3. *Sequences derived from peptides in digest L*

Leu-Leu-Phe-Thr
Val-Ala-Gln
Val-Tyr
Ala-Phe
Val-Leu-Leu
Val-Asp-Leu
Leu-Gly-Glu
Val-Gly
Leu-Ala-Gln-Lys

Table 4. *Analysis of fractions from digest C*

$\lrcorner$Glu indicates the assignment of the *N*-terminal residue of the peptide to pyrrolid-2-one-5-carboxylic acid. CmCys indicates the assignment of a carboxymethylcysteine residue in a sequence.

(*a*) Sequences derived from peptides in digest C:

Fraction IV
Asn-Leu-Leu-Glu-Gly-Leu
Ala-Leu-Gln-Val-Asp
Pro-Asp-Val
Leu-Leu-Ala-Gln-Lys
Asn-Lys-Leu-Val-Ala-Glu-Leu
Lys-Leu-Val-Ala
Lys-Leu-Val-Ala-Glu-Leu
Gly-Ala-Gly-Ala-Lys-Val-Val-Leu
Thr-Ala-Ser-Lys-Phe

Fraction V
$\lrcorner$Glu-Gly-Leu
$\lrcorner$Glu-Leu-Thr-Gly
Ala-Leu-Gln-Val-Asp-Leu
Val-Ala-Glu-Leu-Gly
Leu-Asn-Leu-Asn-Ala-Ala-Phe
Asn-Leu-Asn-Ala-Ala
Ala-Val-Gln-Ala-Phe
Leu-Asn-Ala-Ala
Glu-Pro-Val-Tyr
Val-Ala-Gln-Tyr
Val-Ala-Gln
Ala-Ala-Phe

(*b*) Samples containing arginine or histidine sequenced after (1) hydrazinolysis, (2) acetylation and (3) short permethylation:

Fraction IV
9a Leu-Gly-Gly-Pro-Val-Ala
Val-Ala-Glu-Leu-Gly
9b Ala-Leu-Gln-Val-Asp-Leu-Met
Met-Gln-Pro-Leu-Glu-Val-Ala
Ser-Val-Leu
Ala-Orn-Thr
20a Ala-Lys-Val-Val-Leu
23b Orn-CmCys
Val-Orn-Ser-Val-Leu
26a Met-Lys-His
Ser-Val-Ser-Ser
29a Val-Orn-Ser-Val-Leu-Pro
Orn-CmCys-Val-Orn
30a Orn-Glu-Gly-Glu-Lys

main reasons. (*a*) The number of proteolytic cleavages was such that a greater number of peptides were amenable to analysis by mass spectrometry. (*b*) *N*-Termini were quite random, and the interpretation problem experienced with digest L (see the Discussion section) was not encountered. (*c*) During the latter part of the study on this digest, we developed new procedures for the preparation of peptide derivatives (Morris, 1972; Morris *et al.*, 1973), which subsequently brought many of the peptides containing problem amino acids into the scope of our method.

Most of the samples were analysed as impure mixtures, and the sequences deduced are given in Table 4(*a*).

At this point, samples known to contain histidine or arginine (from specific staining of the analytical electrophoretograms; see the Experimental section) were further purified by preparative electrophoresis, eluted and subjected to the new procedures for the formation of derivatives. Valuable sequence information was produced on examination of these derivatives in the mass spectrometer, and a number of samples were apparent as mixtures (of two to four components) even after purification by paper electrophoresis (see Table 4*b*). The results in Table 4(*b*) are of particular interest, since the peptides, which were sequenced as unknowns, contain examples of the problem amino acids and had previously given little or no sequence information. A mass spectrum of the two major components of the fraction IV9b mixture is shown in Fig. 2.

At a source temperature of 255°C (Fig. 2*a*) a strong *N*-terminal alanine signal is present at *m/e* 128. This is followed by a mass interval corresponding to leucine to *m/e* 255. The next most abundant signal in the spectrum is at *m/e* 482, a mass difference of 227, which does not fit any previously observed amino acid derivative. However, beyond this signal a sequence Val-Asp-Leu could be assigned to *m/e* 595, 738 and 865. It was thus apparent that some unusual modification had taken place, probably during treatment with hydrazine, since this sample had also contained an arginine peptide. The mass difference agrees with the formation of an amide hydrazide of glutamine, as shown in Scheme 3. This has now been observed for both glutamine and asparagine, and we have confirmed the identity of the product by deuterium-labelling studies. The sequence observed at 255°C is thus Ala-Leu-Gln-Val-Asp-Leu-.

At a source temperature of 270°C (Fig. 2*b*) the spectrum of the above peptide had become stronger, revealing a methionine sequence ion, beyond the leucine signal, at *m/e* 1010. The loss of CH_3SH should be noted, which we find characteristic of peptides containing methionine, at *m/e* 962. In addition, a new sequence has appeared containing an *N*-terminal methionine residue at *m/e* 188. This can be followed up the spectrum, each sequence ion showing a loss

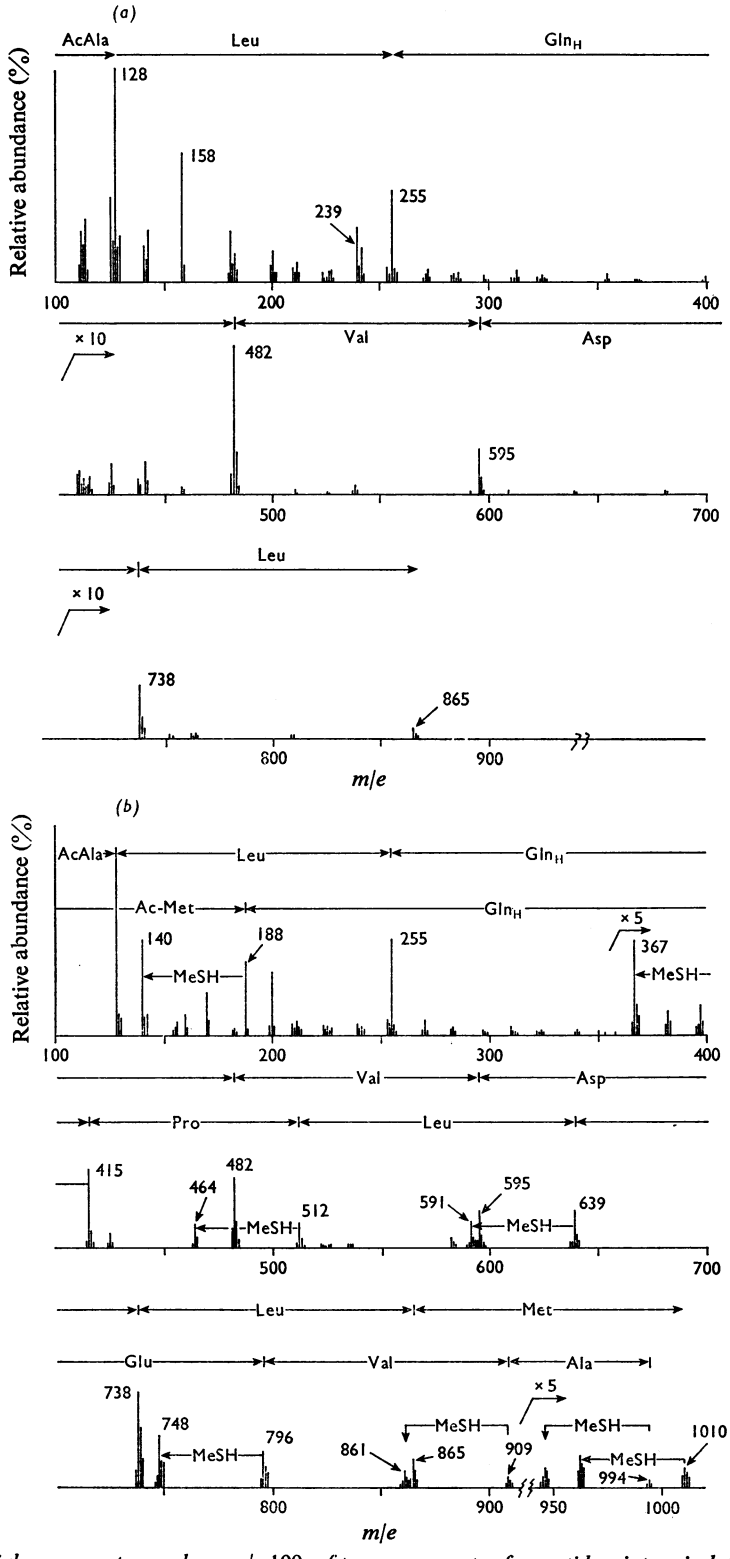
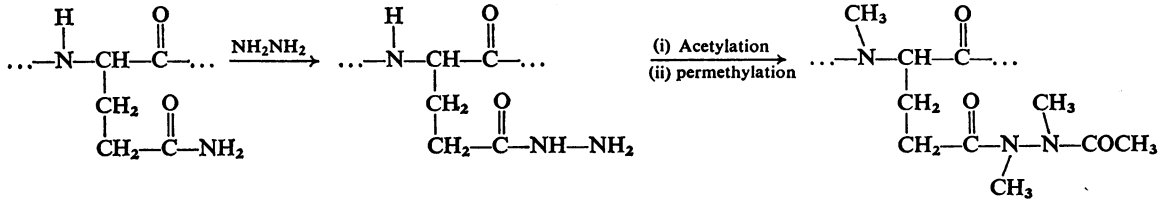


Fig 2. *Partial mass spectrum, above m/e* 100, *of two components of a peptide mixture isolated from digest C* (*a*) shows the spectrum at a source temperature of 255°C, (*b*) at a source temperature of 270°C. The abbreviation Gln_H is used to denote the formation of the hydrazide glutamine. For detailed interpretation see the text.



Scheme 3. *Hydrazide formation on amides*

of CH_3SH, via ions at *m/e* 415, 512, 639, 796, 909, and 994. The major components of this mixture at high source temperature are therefore shown to be Ala-Leu-Gln-Val-Asp-Leu-Met- and Met-Gln-Pro-Leu-Glu-Val-Ala. Neither sequence gave a *C*-terminal signal.

The peptides containing carboxymethylcysteine were sequenced by assuming that decomposition to give dehydroalanine had taken place. The spectra showed a mass difference of 83 mass units, which could also correspond to $(Me)_2Ser-CH_3OH$, but a knowledge of the source temperature, absence of a serine sequence ion or metastable for a loss of methanol from it (Morris *et al.*, 1971) convinced us that serine was not present. This interpretation has been borne out by the classical sequence study. The new mild hydrazinolysis procedure used (Morris *et al.*, 1973) appears not to hydrolyse the various peptide linkages *C*-terminal to acidic amino acids, but the mobility on paper of peptide 26A (+0.36 relative to aspartic acid) suggests that the two sequences observed, Met-Lys-His- and Ser-Val-Ser-Ser-, are produced by hydrolysis of the His-Ser bond probably during hydrazinolysis. Several important 'tryptic overlaps' were obtained in this brief study.

Digest TC

Table 5(*a*) shows the sequences obtained from mass-spectrometric examination of the crude mixtures obtained from ion-exchange fractionation of this digest. A considerable amount of sequence information was obtained from this digest, but also an unusual fragmentation phenomenon was observed for the tryptophan-containing peptides, which made interpretation as unknowns more difficult. This problem seems only to be apparent in mid-chain tryptophan sequences; *N*-terminal and *C*-terminal tryptophan residues behave normally, giving characteristic 'benzylic' and N–C cleavage ions at *m/e* 144 and *m/e* 215 (the latter indicating that the tryptophan is *C*-terminal).

Table 5. *Analysis of fractions from digest TC*

See the legend to Table 4.

(*a*) Sequences derived from peptides in digest TC:

Ala-Leu-Gln-Val-Asp-Leu
Leu-Val-Ala-Glu-Leu-Gly
Leu-Gly-Gly-Pro-Val-Ala
Gln-Gly-Leu
Leu-Val-Ala-Glu-Leu
Asn-Leu-Leu-Gln-Gly-Leu
Glu-Val-Asp-Leu
Pro-Asp-Val
Met-Asp-Glu-Ala-Leu
Glu-Ala-Leu-Ala
Ser-Gly-Asp-Leu-Leu
Ala-Val-Gln-Ala-Phe
Ser-Val-Asp-Leu
Leu-Asp-Leu-Phe
Glu-Pro-Val-Tyr
Val-Ala-Gln
Leu-Val-Leu
Ala-Ala
Leu-Leu-Ala-Gln-Lys
Thr-Ala-Ser-Lys-Phe
Ala-Leu-Leu-...Trp-Pro-Lys
Trp-Pro-Lys
Ser-Val-Leu-Pro-His

(*b*) Samples containing arginine or histidine sequenced after (1) hydrazinolysis, (2) acetylation and (3) short permethylation:

Fraction	
14a	Asn-Val-Thr
	Ser-Val-Leu-Pro-His
22b	Asp-Orn-Val-Leu
	Thr-Gly-Orn-Leu-Asp-Leu-Phe
22c	Asn-Val-Thr
23c	Met[O]-Asn-Thr
24b	His-Ala-Asn-Ala-Gly-Ala-Tyr
26c	Val-Leu-His-Leu
30a	Met-Val-Thr-Orn
30b	Orn-Leu-Asp-Leu-Phe
	Met-Val-Thr
33b	Orn-CmCys-Val
35	Val-His-Thr-Thr-Orn

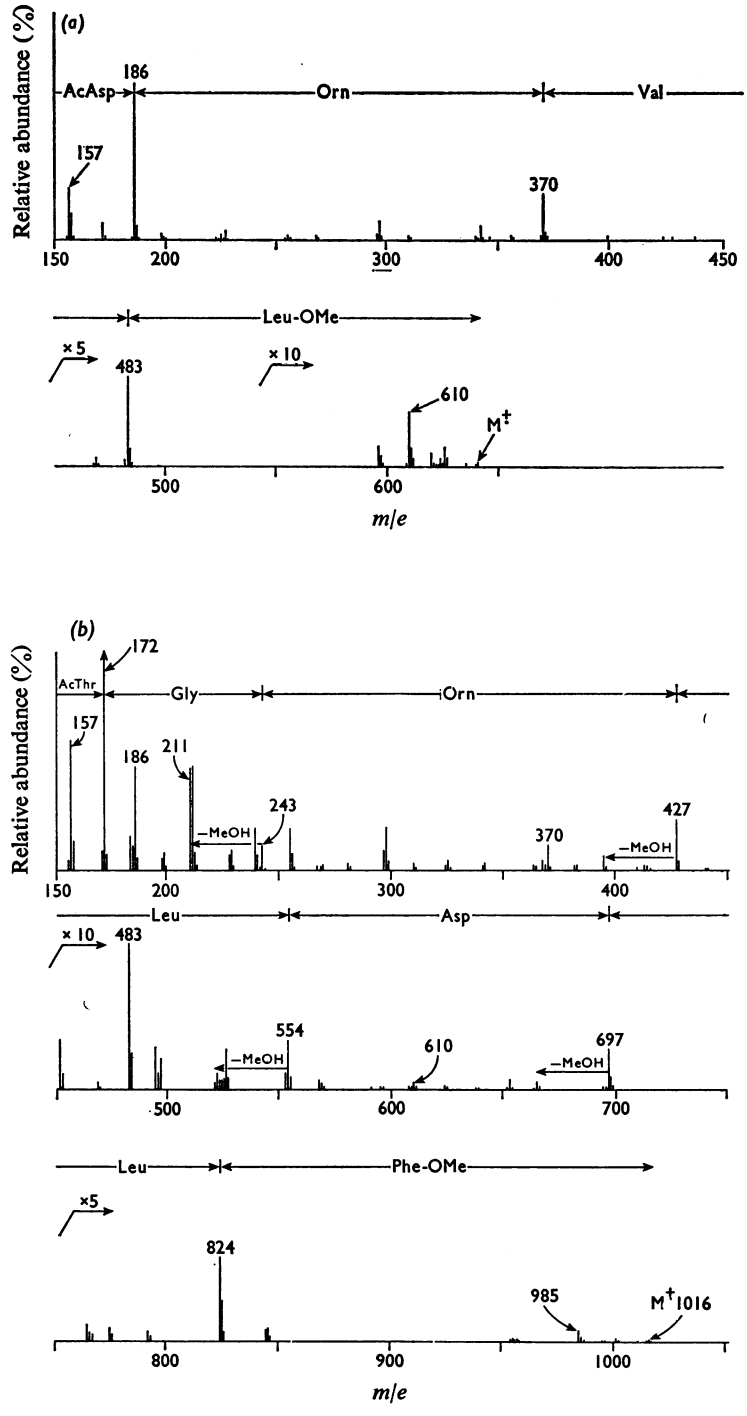
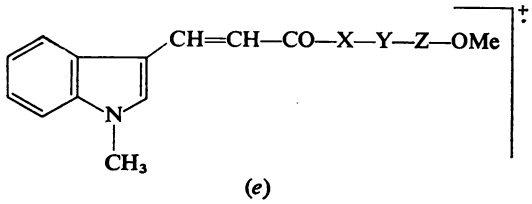


Fig. 3. *Partial mass spectrum, above m/e* 100, *of two components of a peptide mixture isolated from digest TC*

(*a*) shows the spectrum at a source temperature of 220°C, (*b*) at a source temperature of 250°C. For detailed interpretation see the text.

With other mid-chain amino acids, e.g. -Tyr- (Morris *et al.*, 1971), the N–C cleavage product gives rise to a series of normal sequence ions by fragmentation at the peptide bonds, thus confirming the sequence at the *C*-terminal side of the tyrosine. Mid-chain tryptophan peptides behave differently. To take the general case of a peptide A-B-C-Trp-X-Y-Z, the spectrum is dominated by the fragment



(*e*)

(*e*, i.e. . . . TrpXYZ). Thus for one of the unknown peptides shown in Table 5(*a*), the low-temperature spectra showed sequence ions at *m/e* 128, 255 and 382, corresponding to the sequence Ala-Leu-Leu. Signals at *m/e* 144 and 184 indicated the presence of tryptophan in the peptide, but higher-temperature scans were dominated at high mass by one signal at *m/e* 510. Weaker signals were present at *m/e* 184 and *m/e* 281, indicating . . . (Me)Trp and a possible . . . (Me)Trp-Pro. The signal at *m/e* 510 could thus be rationalized as originating from . . . (Me)Trp-Pro-(Me_2Ac)Lys-OMe, although the conventional losses of $\cdot CH_3$ and $\cdot OCH_3$ from the ester were not present. This behaviour of mid-chain tryptophan-containing peptides has been confirmed by deuterium-labelling experiments.

The new procedures for the preparation of derivatives were again applied to samples thought to contain histidine or arginine, which had been further purified on paper. Again a considerable amount of information was obtained from these fractions, including some overlapping regions presumed lost in earlier experiments using different derivative-forming procedures. The sequence data obtained are given in Table 5(*b*). An arginine-staining peptide eluted from the neutral band of an electrophoretogram gave the spectra shown in Fig. 3. The spectrum at a source temperature of 220°C (Fig. 3*a*) clearly showed the presence of a single peptide having the sequence Asp-Orn-Val-Leu (derived from Asp-Arg-Val-Leu) via the sequence ions at *m/e* 186, 370, 483, and 641. At a source temperature of 250°C (Fig. 3*b*) this peptide was still present in the spectrum, but a new component had arisen, giving sequence ions at *m/e* 172, 243, 427, 554, 697, 824 and a molecular ion at *m/e* 1016. With the aid of the characteristic losses of methanol from threonine this component can be sequenced unambiguously as Thr-Gly-Orn-Leu-Asp-Leu-Phe (derived from Thr-Gly-Arg-Leu-Asp-Leu-Phe).

Discussion

All of the mass-spectrometric sequence data were interpreted without reference to the sequence determined by classical methods, and many of the sequences predated the classical data. We have found no conflicts between the sequences determined by these techniques, so this mass-spectrometric investigation has provided valuable confirmatory evidence for the complete sequence shown in Fig. 4. About 80% of the sequence has been screened in this way. The work was carried out over a period of about 18 months, during which the development of suitable peptide-digestion and fractionation techniques and new derivative-formation procedures greatly speeded the acquisition of useful data. Had these been available initially we believe that almost all of the sequence would have been determined, although not necessarily fully overlapped, during this period.

Of particular importance to the interpretation of many of the spectra has been the N–C cleavage reaction with associated hydrogen rearrangement. The abundance of a signal is an important factor in interpretation (see below) and the newly observed cleavages with their associated *m/e* values and abundances are shown in Table 6 (see Scheme 2 for plausible ion structures). No difficulty should be found in observing the . . . $(Me)_2$Asn and . . . (Me)-Trp cleavage ions but, as Table 6 indicates, the . . . (Me)Asp sequence ion is very often weak. This being so, the presence of an unidentified signal in the *m/e* 184 [. . . (Me)Asp-(Me)Gly] to *m/e* 327 [. . . (Me)Asp-(Me)Trp] range of the spectrum should always be checked for an ion corresponding to . . . (Me)Asp as well as other genuine *N*-termini.

Certain ambiguities in the interpretation of low-resolution spectra, such as two possible sources for the *m/e* 184 signal mentioned above, can be simply overcome. This signal, for example, can be due to . . .(Me)Trp or . . .(Me)Asp-(Me)Gly. First, one asks, are other tryptophan-associated signals present, particularly *m/e* 144? If not, then *m/e* 184 definitely does not originate from tryptophan. However, if *m/e* 144 is present one could have a mixture of . . .(Me)Trp and . . .(Me)Asp-(Me)Gly sequences. This can be simply resolved by deuteropermethylation, since . . .(Me)Trp, having only one replaceable group, shifts to *m/e* 187, whereas . . .(Me)Asp-(Me)Gly shifts to *m/e* 190. Other associated sequence ions will of course shift accordingly.

Another possible source of error not previously recognized is the interpretation of the signals at *m/e* 126 and *m/e* 98 (Schemes 1 and 2). As we have shown (Scheme 2), we now find that *m/e* 126 can be derived from peptides containing asparagine. In all, four possibilities exist for this signal: (*a*) *N*-terminal pyrrolid-2-one-4-carboxylic acid (Scheme 1); (*b*)

Fig. 4. *Complete sequence of ribitol dehydrogenase, as determined classically*

Sequences determined in this mass-spectrometric study are underlined, and marked with the code for the digest from which the peptide was isolated. Lines not terminated by an arrow indicate that no *C*-terminal sequence ion was observed in the mass spectrum. Lines terminated by an arrow indicate that the full sequence of the peptide was deduced from the mass spectrum. The identities of residues marked with an asterisk have been determined solely by mass spectrometry. For details see the text.

Table 6. *Observations on the newly assigned fragment ions*

Amino acid	Mass of fragment	Normal abundance	Characteristic-associated losses
Asp	113	Weak	——
Asn	126	Moderate	CO → *m/e* 98
Trp	184	Strong	——

N-terminal serine minus methanol [i.e. $Ac(Me)_2Ser-MeOH$]; (*c*) *N*-terminal dehydroalanine (which we have observed is derived from carboxymethylcysteine; see the Results section); (*d*) N–C cleavage at asparagine.

The question arises, therefore, can we differentiate between these signals? We have had no difficulty in doing so in the following manner.

Dehydroalanine may be differentiated from serine by the absence of the signal at *m/e* 158 in the former case. Alternatively, a metastable peak for the decomposition *m/e* 158→126 is observed if serine is present. Deuteropermethylation allows differentiation of pyrrolid-2-one-5-carboxylic acid and dehydroalanine (shift to *m/e* 129) from ...$(Me)_2$Asn (shift to *m/e* 132), and deuteroacetylation allows the

differentiation of the remaining ambiguity, i.e. pyrrolid-2-one-5-carboxylic acid does not shift, but *N*-terminal dehydroalanine is found at *m/e* 129, with an accompanying signal at *m/e* 101 (loss of CO). Most important in the analysis of a mixture is the observation of the relative movements of all the signals in the mass spectrum as a temperature gradient is applied at the probe, which indicates which *N*-terminal signals belong to which sequence ions. In addition, a qualitative amino acid analysis or the known presence of a carboxymethylcysteine residue from radiolabelling data (a very common procedure in modern protein chemistry) may be helpful in very difficult cases. Although the above procedure may seem complex, it is worth noting that in the hundreds of peptide spectra examined in this study, deuterium labelling was only needed for three samples, and these labelling studies were confirming our interpretation of previously unreported fragmentations.

The concept of using a relatively non-specific proteolytic enzyme for digestion in order to provide smaller fragments more amenable to mass spectrometry (see the Results section) is certainly a sound one. Our choice of thermolysin for a whole-protein digest, however, gave rise to unsuspected difficulties. These did not arise from incomplete digestion, for indeed the gel-filtration data showed the expected preponderance of low-molecular-weight materials. The problem arose as a result of the specificity of the enzyme coupled with our strategy for sequencing by mass spectrometry. The majority of peptides released by thermolytic digestion have *N*-terminal leucine, isoleucine, valine or alanine and very often another hydrophobic residue in the second position (leucine, valine, alanine, glycine, phenylalanine etc.). If the products were examined as single peptides, this would create no problems, but in the complex mixtures analysed, ambiguity of sequences was quickly apparent. This arose because of the numerical relationships between some of the derivatized amino acids.

Consider the low-mass end of a hypothetical spectrum. The following observations may be made: *m/e* 128 may be Ac(Me)Ala or Ac(Me)Val−CO; *m/e* 156 is Ac(Me)Val; *m/e* 170 is Ac(Me)Leu. Now *m/e* 241 may be due to any or all of four possible sequences, (i) Ala-Val, (ii) Val-Ala, (iii) Leu-Gly or (iv) Lys. Similarly, *m/e* 326 may be due to any or all of the following sequences, (i) Ala-Val-Ala, (ii) Val-Ala-Ala, (iii) Leu-Gly-Ala, (iv) Lys-Ala, (v) Ala-Lys, (vi) Val-Gln, (vii) Leu-Asn, and so on; the ambiguities multiply as we progress up the spectrum. Now obviously some of these possibilities can be discounted either by (a) the normal mixture analysis practice of comparing relative intensities of consecutive scans; this method indeed proved to be useful in some cases, as evidenced by the sequences obtained from the thermolysin digest shown in Table 3; or (*b*) using isotopic derivatives or mass measurement, e.g. the possibility of *m/e* 241 being due to lysine could be discounted if this signal did not shift by 6 mass units on deuteroacetylation. Unfortunately, neither isotopic derivatives nor mass measurement will differentiate the first three possibilities for *m/e* 241. Thus, if the partial vaporization of the mixture fails, these combinations cannot be differentiated.

Because of the similarity of the *N*-termini of all the peptides present in the complex mixtures derived from the thermolysin digest, most of the mixtures did not yield unambiguous sequences and the digest was abandoned.

We have met this problem only once before, in a study of the partial acid-hydrolysis products of a peptide of most unusual composition ($Gly_{29}Ala_{20}Ser_9Tyr$) isolated from silk fibroin (Morris, 1970; H. R. Morris & F. Lucas, unpublished work). Here again the preponderance of like *N*-termini, followed in this case by alternating sequences of the two major amino acids, led to ambiguous sequence assignments; further purification to give more simple mixtures had to be carried out in order to sequence the components and suggest a total structure for the peptide. Needless to say, in the vast majority of proteins, this problem would not arise, but one must be careful to avoid thermolytic digestion for the reasons given above.

More recently, we have examined the use of another enzyme, elastase, for generating small fragments suitable for mass spectrometry (Morris *et al.*, 1974). This enzyme has yielded the required small fragments, and none of the problems associated with thermolysin have been apparent.

Turning now to the question of overall strategy; previously we have demonstrated the success of a low-resolution mass-spectrometer approach on individual components, and on mixtures of peptides of unknown sequence. In the present paper we have demonstrated the use of this strategy when applied to the sequence analysis of a protein of mol.wt. 27000.

Recently it has been suggested that low-resolution mass spectrometry is inadequate for peptide mixture analysis, and that it is necessary to use a composite approach involving high-resolution mass spectrometry, metastable analysis and chemical-ionization mass spectrometry (Wipf *et al.*, 1973). This statement is supported by the analysis of several known mixtures of synthetic peptides, only three of which were permethylated. Although we also have been investigating the possible benefits of chemical ionization, it will be apparent from our results, obtained in general on more complex mixtures (unknown also) than those quoted by these authors, that such elaborate procedures are not necessary as a general routine method.

We would favour the use of a computer (if available) for the analysis of low-resolution mixture

spectra, but can see no specific advantage in coupling this with a high-resolution approach. Further, we do not believe that with the conventional high-resolution mass spectrometer the same degree of sensitivity can be achieved as with our low-resolution method. Suppression phenomena associated with the introduction of a reference compound, together with the decreased slitwidths needed for high resolving power, have in our laboratory led to a decrease in sensitivity of up to a factor of 100. This can correspond to a loss of one or two sequence signals per peptide at high mass, i.e. two or three amino acid residues not able to be placed in the sequence. If one is studying a peptide mixture, obviously the total information forfeited at high resolution may be greater.

We thank the M.R.C. and S.R.C. for financial support. H. R. M. is on the staff of the M.R.C. Laboratory of Molecular Biology.

References

Agarwal, K. L., Johnstone, R. A. W., Kenner, G. W., Millington, D. S. & Sheppard, R. C. (1969) *Nature (London)* **219**, 498–499

Geddes, A. J., Graham, G. N., Morris, H. R., Lucas, F., Barber, M. & Wolstenholme, W. A. (1969) *Biochem. J.* **114**, 695–702

Hartley, B. S. (1974) in *Evolution in the Microbial World* (Carlile, M. J., & Skehel, J. J., eds.), pp. 151–182, Cambridge University Press, Cambridge

Hartley, B. S., Burleigh, B. D., Midwinter, G. G., Moore, C. H., Morris, H. R., Rigby, P. W. J., Smith, M. J. & Taylor, S. S. (1972) *Proc. FEBS. Meet.* **29**, 151–176

Morris, H. R. (1970) Ph.D. Thesis, University of Leeds

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., Batley, K. E., Harding, N. G. L., Bjur, R. A., Dann, J. G. & King, R. W. (1974) *Biochem. J.* **137**, 409–411

Roepstorff, P., Spear, R. K. & Brunfeldt, K. (1971) *FEBS Lett.* **15**, 237–240

Shemyakin, M. M., Ovchinnikov, Yu. A., Kiryushkin, A. A., Vinogradova, E. I., Miroshnikov, A. I., Alakhov, Yu. B., Lipkin, V. M., Shvetsov, Yu. B., Wulfson, N. S., Rosinov, B. V., Bochkarev, V. N. & Murikov, V. M. (1966) *Nature (London)* **211**, 361–366

Shotton, D. M. & Hartley, B. S. (1973) *Biochem. J.* **131**, 643–675

Taylor, S. S., Rigby, P. J. W. & Hartley, B. S. (1974) *Biochem. J.* **141**, 693–700

Wipf, H.-K., Irving, P., McCamish, M., Venkataraghavan, R. & McLafferty, F. W. (1973) *J. Amer. Chem. Soc.* **95**, 3369–3375