With large numbers of hydroxyl groups, the trimethylsilyl derivatives become too large in mass for the range of many quadrupole mass spectrometers, so methylation is used. However, this would cause problems in structural elucidation since those compounds in which demethylation of the methoxyl or dimethylamino group had taken place would be remethylated, thus obscuring these demethylating processes. This problem can be overcome by the use of deuteriomethylation. We originally carried this out by equilibrating the extract with $O$-deuteriomethanol, to exchange all hydroxyl and amino hydrogens, and carrying out the methylation with diazomethane in deuteriomethanol. This, however, allows the formation of some undeuterated methoxy products. Commercial kits are now available for producing dideuteriodiazomethane of over 99\% purity.

When monitoring of the methoxybenzyl ion at m/e 121 was carried out on the deuteriomethylated extract at pH 9 from a volunteer taking mepyramine, the g.l.c.-m.s. run gave only one response, at the correct retention time for mepyramine. However, when m/e 124 was monitored a response was again obtained at the same retention time as for mepyramine, and was about 10\% of the response for m/e 121. Since m/e 124 corresponds to trideuteriomethoxybenzyl, this must have been formed from $O$-demethylated mepyramine, which was present to the extent of about 10\% of the level of unchanged drug.

It was also expected that a metabolite would be formed in which the methoxybenzyl ring of compound (9) had become further hydroxylated. This hydroxylation could occur either before or after demethylation of the methoxyl group occurred. In the first case deuteriomethylation would lead to a methoxy deuteriomethoxybenzyl substituent, giving an ion of m/e 154 whereas the second alternative would yield a bis deuteriomethoxy substitution on the benzyl ring, giving an ion of m/e 157. Multiple-ion monitoring of both these ions on a g.l.c.-m.s. run gave a response only at m/e 154. A mass spectrum obtained on this metabolite was identical with the mass spectrum of deuteriomethylated compound (12), so that it can be assumed that hydroxylation has occurred at the 3 position of the methoxybenzyl substituent.

Although mepyramine suffers $O$-demethylation, no $N$-demethylation was seen to occur. Side-chain fission produces an ion of m/e 58 due to $\text{CH}_2=\text{NMe}_2$. Deuteriomethylation of an $N$-demethylated metabolite would shift this ion to m/e 61, whereas deuteriomethylation of an $N$-didemethylated metabolite would shift this ion to m/e 64. A g.l.c.-m.s. run monitoring these three ions produced responses only for m/e 58, thus ruling out any N-demethylation.

Waddel, W. J. (1965) J. Pharm. Exp. Ther. 149, 23–28

Protein Sequence Analysis and the Discovery of a New Amino Acid in Prothrombin

HOWARD R. MORRIS
Department of Biochemistry, Imperial College, London SW7 2AZ, U.K.

A number of methods of attacking the problems of protein structure by mass spectrometry are now finding application in the field. Techniques fall broadly into two classes: those involving the study of small molecules, such as amino acid thiohydantoins, or di-, tri- and tetra-peptides, and studies of molecules larger than this. The former approach has not been widely applied as yet, but brief accounts of application to protein-derived peptides have been given. In contrast, a number of major studies of larger peptides isolated from protein digests have now been reported.

Vol. 3
Considering the small-molecule work first, one aspect of this has been the development of methods for examining the products of either manual or automatic Edman procedures, i.e. thiazolinones or thiohydantions. Some of the amino acid residues in the sequence of ovine hypothalamic luteinizing-hormone-releasing factor have been confirmed by using a g.l.c.-mass spectrometry study of the phenylthiohydantoin derivatives (Burgus et al., 1972). The thiourea obtained by reaction of a peptide with methylisothiocyanate has been shown to rearrange and cleave to give the thiohydantoin in the mass-spectrometer ion source (Fairwell et al., 1973), thus obviating the necessity for chemical transformation. This can form the basis of an Edman-mass spectrometric method in which a sample is taken for mass-spectrometric analysis while the bulk of the material is treated with trifluoroacetic acid and the cycle repeated. Problems associated with the identification of phenylthiohydantoin derivatives from the sequenator (normally carried out by g.l.c., t.l.c. and amino acid analysis) can be alleviated by examination in the mass spectrometer. The results can be rapidly quantified by concurrent analysis of $^{15}$N (Richards & Lovins, 1972), brominated (Weygand & Obermeier, 1971) or deuterated (Tschesche et al., 1972) analogues.

The analysis of dipeptides produced by aminodipeptidase digestion of polypeptides continues. The method has been successfully, if modestly, tested on the insulin A chain, and relies on complete digestion followed by g.l.c.-mass-spectrometry identification of all dipeptides produced (Caprioli et al., 1973). The procedure is repeated after one step of Edman degradation, and overlap peptides are constructed. Methods for overcoming enzyme specificity and derivative problems are now being developed.

A similar 'small-molecule' approach involves partial acid hydrolysis and g.l.c.-mass spectrometry of the peptide derivatives. Problems such as deamidation exist, but the method has been successfully demonstrated on a 20-residue peptide from rabbit actin (Nau et al., 1973).

Clearly the above methods show promise as rapid, sensitive procedures for examining relatively small polypeptides. An extension to protein sequence analysis may suffer from the inability to overlap the many di-, tri- or tetra-peptides produced. This problem decreases with peptide size, and the mass-spectrometric study of larger peptides has now proved valuable in a number of protein sequence studies.

All protein studies reported to date have made use of the acetyl permethyl derivative, methods for the preparation of which are now rapid (Morris, 1974) and applicable to all known amino acids (Morris et al., 1973).

In studies of $\alpha$-lactalbumin (Bacon & Graham, 1972), triose phosphate isomerase (Priddle & Offord, 1974), and a cow-pea tobacco-mosaic-virus protein (Rees et al., 1974), mass spectrometry has been used to confirm the sequence of certain regions of these proteins. Methods used in the above studies vary from analysis of peptides of known amino acid composition and length, to analysis of peptide mixtures of unknown composition (Morris et al., 1971).

Two mass-spectrometric protein studies undertaken without a knowledge of the classically determined sequence have now been reported for ribitol dehydrogenase (Morris et al., 1974b) and chloramphenicol transacetylase (Dell & Morris, 1974). In both cases more than 80% of the structures have been deduced by using a random-mixture-analysis approach. Other mass-spectrometric work is also progressing on dihydrofolate reductase from Lactobacillus casei (Morris et al., 1974a), a protein of unknown structure, and on a Pseudomonas azurin, where a homologous sequence is used to align the rapidly determined mixture-analysis data (A. Dell & H. R. Morris, unpublished work).

The new ionization techniques have not been applied to genuine protein problems as yet. Field desorption leads to a loss of sequence information in the spectrum, and chemical ionization can give rise to a more complex spectrum than electron impact. Despite these problems, specific examples of their usefulness are given.

A convincing demonstration of the power of the low-resolution/mixture-analysis mass-spectrometric approach to protein sequencing has arisen recently in connection with studies on the vitamin K-dependent part of prothrombin. Classical dansyl-Edman
and automatic-sequenator techniques had failed to uncover the nature of a 'prosthetic group' giving rise to abnormal electrophoretic mobility in peptides isolated from the N-terminus of prothrombin. The nature of this anomaly has been independently attributed to a new amino acid, γ-carboxyglutamic acid (Gla), in studies involving mass-spectrometric analysis of a pure peptide (Stenflo et al., 1974), and of peptide mixtures (Magnusson et al., 1974) isolated from prothrombin. In the latter study it has proven possible to identify the new amino acid in ten positions in the N-terminal sequence, by examination of crude samples containing several components, and in one case a peptide 18 residues in length. The following sequences were deduced:

\[
\text{Gla-Gla-Ala-Phe-Gla-Ala} \\
\text{Ala-Leu-Gla-Ser-Leu} \\
\text{Phe-Leu-Gla-Gla-Val} \\
\text{Leu-Gla-Orn-Gla} \\
\text{Gla-} \Delta \text{Ala-Leu-Gla-Gla}
\]

[Orn is derived from Arg, and ΔAla (dehydroalanine) from CmCys].

Comparison with the classically determined sequence shows that those residues assigned to Glu are in fact Gla (Magnusson et al., 1974). A synthesis of γ-carboxyglutamic acid has been reported in brief (Morris et al., 1975) which corroborates an anomalous reaction in the natural material during derivative formation.