PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Research on Peptides and Glycopeptides

H. R. Morris

Phil. Trans. R. Soc. Lond. A 1979 **293**, 39-51
doi: 10.1098/rsta.1979.0078

Phil. Trans. R. Soc. Lond. A **293**, 39–51 (1979) [39]
Printed in Great Britain

Research on peptides and glycopeptides

By H. R. Morris
Department of Biochemistry, Imperial College of Science and Technology, Imperial Institute Road, London, S.W.7, U.K.

The current state of methodology for protein sequence analysis is discussed with reference to the enzyme dihydrofolate reductase. Studies on biologically active compounds of unusual structure are shown, covering data on peptide hormones and factor X, the zymogen of a serine protease involved in blood coagulation. Recent research on glycopeptides from biological sources illustrates the development of chemical and mass spectrometric methods for concurrent determination of saccharide and peptide structure. Details of future trends are discussed by reference to present work in the areas of negative chemical ionization, computerized automation and high-field magnet studies.

Protein sequence analysis

It is now 20 years since studies on small synthetic peptides first indicated the possibility of sequence analysis by mass spectrometry (Andersson 1958; Biemann *et al.* 1959), and 10 years since the mass spectrometric analysis of protein-derived peptides and peptide mixtures was first demonstrated (Morris *et al.* 1969). Since then, considerable progress has been made, resulting in combined classical – mass spectrometric attacks on total protein sequences, for example on ribitol dehydrogenase (Morris *et al.* 1974) and chloramphenicol transacetylase (Dell & Morris 1976; A. Dell, H. R. Morris, W. Shaw & D. B. Burleigh, unpublished work). Only this year, however, are we able to announce the first total protein sequence assignment by mass spectrometry, independent of classical methodology, in work on dihydrofolate reductase. Why has it been so difficult and taken so long to reach this stage? The short answer is that the mass spectrometer is not an obvious method of choice for the study of high molecular mass complex polar molecules such as peptides and proteins; it is best suited to the analysis of low molecular mass volatile substances. The problems therefore have been the reconciling of these apparently irreconcilable differences, to produce a methodology that has clear advantages over classical methodology whether manual or automated.

Our contribution to this research has been in the three areas of derivatives, fragmentation and strategy discussed below.

Derivatives

An early problem in the application of mass spectrometry to protein sequencing was the production of suitably volatile derivatives in high yield. Acetylation of primary amino functions (Thomas *et al.* 1968) is the best first step, with the use of acetic anhydride in methanol. Side reactions such as ester formation can be minimized by the finding that the reaction is practically instantaneous for α-amino groups, thus enabling a reduction in reaction time from the standard 3 h to 1 min (Morris 1974; H. R. Morris, D. Mak, K. E. Batley & A. Dell, unpublished work). The deprotonation of δ- or ε-amino functions can be aided by drying the peptide down from triethylamine or sodium bicarbonate before acetylation.

Permethylation was introduced as a second and final derivative-forming step some 10 years ago (Das *et al.* 1967; Vilkas & Lederer 1968), but under the suggested conditions (either Kuhn or Hakamori) permethylation was not usually successful with peptides containing cysteine, histidine or methionine, thus precluding the general application to protein sequence analysis.

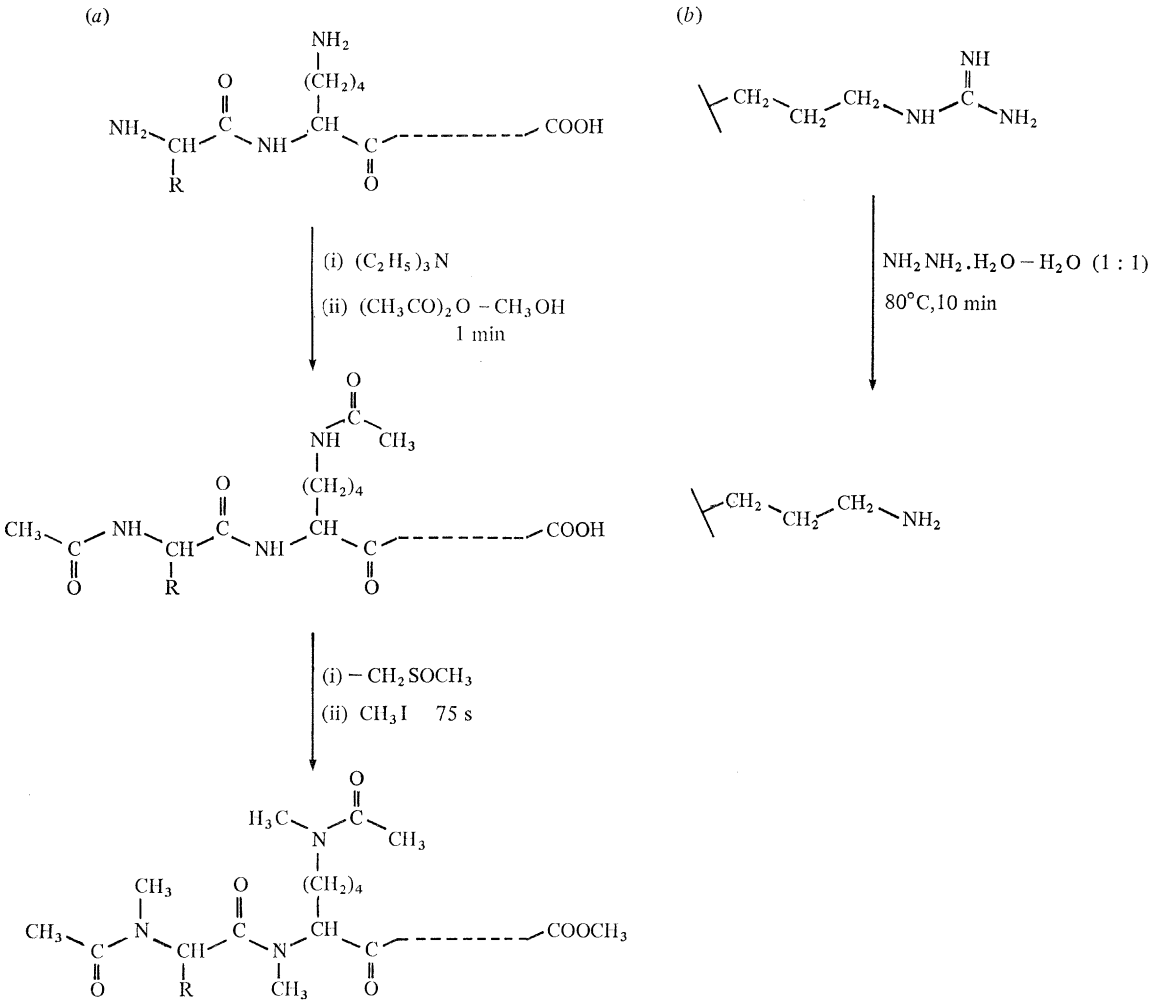


FIGURE 1. (*a*) General derivative-forming reactions applicable to protein-derived peptides. Arginine-containing peptides require an additional hydrazinolysis step (*b*).

The problem was rationalized as one of 'salt' formation (sulphonium and quaternary ammonium), and in 1972 a study of alkylation reaction rates at various functional groups showed that peptides could be practically fully permethylated without extensive 'salt' formation by restricting reaction times for alkylation to the order of 1 min (Morris 1972; Morris *et al.* 1973).

Only one other amino acid, arginine, requires additional chemical modification before mass spectrometric study. This is accomplished by the conversion of arginine to ornithine under the mildest possible conditions (Morris *et al.* 1973). This reaction is carried out only on those peptides showing a positive stain test for arginine.

A summary of current reaction procedures is given in figure 1.

Fragmentation

The one feature of the early work (Andersson 1958) that suggested that mass spectrometry may be useful for peptide sequencing was the observation of fragmentation at the peptide bond. This 'sequence ion' formation first observed for acyl peptide esters is an even more pronounced fragmentation pathway when the amide bond is methylated and attempts have been made to rationalize this (Morris & Dell 1975). Recognition of this important fragmentation pathway is not, however, the only thing needed for the successful interpretation of the mass spectrum. Often the true molecular ion is not present, and we must be aware of other fragmentation pathways including radical or neutral losses from the molecular ion or other fragment ions, and most importantly the formation of N–C cleavage fragments (see figure 2). These fragments are observed whenever one or more of six particular amino acids (Asp, Asn, Phe, Tyr, Trp or His) are present in a peptide, and are very useful for confirming or extending a sequence assignment from normal 'sequence ions' (Morris *et al.* 1971; Morris *et al.* 1974*a*). Occasionally, the juxtaposition of two of these amino acids can be recognized by a low mass fragment ion, even when the sequence ions are not observed at higher masses (Dell & Morris 1975). The exact nuances of interpretation are too detailed for discussion here, but suffice it to say that current ideas on peptide sequence assignment go well beyond simple recognition of 'sequence ions', to an understanding and recognition usually of every signal in the spectrum. This degree of sophistication has been generated by the study of hundreds of peptides in protein sequence analyses, and means that valuable sequence information can now be obtained from spectra, the interpretation of which might well have been abandoned 10 years ago because of 'ambiguous' assignments.

FIGURE 2. Production of an 'N–C cleavage' fragment from the peptide backbone.

Strategy

A most important aspect of this research is which strategy should be developed to make the method successful. Should we use 'classical' enzymes, e.g. trypsin, for initial digestion of the protein, then purifying all the peptides and redigesting the larger ones? What purification methods will be compatible with mass spectrometry? Should we purify the peptides or the peptide derivatives, by, for example, ion exchange in non-salt buffers or g.c. or h.p.l.c. of derivatives? What type of derivative? How can we best obtain information from the mass spectrometer, e.g. by high resolution, low resolution, e.i. or c.i.? Some of these questions are interrelated; for example, the initial choice of say a dipeptidyl aminopeptidase digest would lead to a certain choice of derivatives (not permethylation) and probably g.c.–m.s. separation and identification of dipeptides. Such procedures have been used on modest sized problems (Krutzsch 1978) but as yet have not been demonstrated on unknown proteins, normally defined as a polypeptide chain greater than 100 amino acids in length.

Faced with this question of strategy some years ago, I decided that low resolution direct probe analysis would be the approach to develop. A proper justification for this choice would

involve many detailed comparisons, outside the scope of the present discussion, but some advantages of this approach can be stated briefly. A key word is 'mixtures'. The ability to sequence unambiguously a mixture of peptides (Morris *et al.* 1969; Geddes *et al.* 1969) is unique to mass spectrometry, but the significance of this statement is only grasped when one realizes that the isolation and preparation of individual purified peptides is the rate determining step in any classical protein sequencing method; even when using an automated spinning cup sequencer, rarely more than 50–100 amino acids can be placed in sequence by studying the intact protein, and then the laborious digestion, isolation and purification of individual peptides must be resorted to in order to finish the sequence. Mixture analysis provides us with a much needed way of bypassing this rate limiting step in the overall strategy.

TABLE 1. MASS SPECTROMETRICALLY DETERMINED AMINO ACID SEQUENCE OF DIHYDROFOLATE REDUCTASE

Thr–Ala–Phe–Leu–Trp–Ala–Gln–Asp–Arg–Asp–Gly–Leu–Ile–Gly–Lys–Asp–Gly–His–Leu–Pro–
Trp–His–Leu–Pro–Asp–Asp–Leu–His–Tyr–Phe–Arg–Ala–Gln–Thr–Val–Gly–Lys–Ile–**Met**–Val–
Val–Gly–Arg–Arg–Thr–Tyr–Glu–Ser–Phe–Pro–Lys–Arg–Pro–Leu–Pro–Glu–Arg–Thr–Asn–Val–
Val–Leu–Thr–His–Gln–Glu–Asp–Tyr–Gln–Ala–Gln–Gly–Ala–Val–Val–Val–His–Asp–Val–Ala–
Ala–Val–Phe–Ala–Tyr–Ala–Lys–Gln–His–Leu–Asp–Gln–Glu–Leu–Val–Ile–Ala–Gly–Gly–Ala–
Gln–Ile–Phe–Thr–Ala–Phe–Lys–Asp–Asp–Val–Asp–Thr–Leu–Leu–Val–Thr–Arg–Leu–Ala–Gly–
Ser–Phe–Glu–Gly–Asp–Thr–Lys–**Met**–Ile–Pro–Leu–Asn–Trp–Asp–Asp–Phe–Thr–Lys–Val–Ser–
Ser–Arg–Thr–Val–Glu–Asp–Thr–Asn–Pro–Ala–Leu–Thr–His–Thr–Tyr–Glu–Val–Trp–Gln–Lys–
Lys–Ala

Heterogeneity was discovered at some positions in the sequence and this, together with details of the determination, will be reported elsewhere.

The principle is one of fractional distillation of the individual peptides from the mixture on the probe tip within the mass spectrometer with the use of either the probe or ion source heater; the signals associated with any one peptide in the mass spectrum will rise and fall together, and, depending upon the separation achieved, at different rates to other components in the mixture. The result is that each component can normally be sequenced unambiguously, the method being best suited to mixtures of two to five peptides.

To complete the strategy we need a method of producing peptides within the effective mass range of mass spectrometry, i.e. less than *ca.* 1200 molecular mass. Cathepsins can be used, but by generating such small fragments (dipeptides) we would be forfeiting the valuable 'overlap' information present in larger peptides. In contrast, trypsin or chymotrypsin may produce fragments too large for the mass spectrometer. The successful solution is to choose a non-specific protease (Morris *et al.* 1974*a*, *b*) such as thermolysin, elastase or subtilisin (Dell & Morris 1977).

The overall strategy developed is well illustrated by our study of the sequence of dihydrofolate reductase from MTXR-resistant *L. casei* (Batley & Morris 1977*a*, *b*; H. R. Morris, K. E. Batley, R. Inglis & A. Dell, unpublished work). Here several digests were employed, including

elastase and subtilisin, and the large cyanogen bromide fragments were isolated to overlap and place the many peptides produced into a final sequence (table 1).

Peptide digests were in general separated by only one cation exchange column step, the effluent being screened analytically by high voltage paper electrophoresis before pooling of peptide mixture fractions, and derivatization. A typical spectrum from the elastase digest is seen in figure 3.

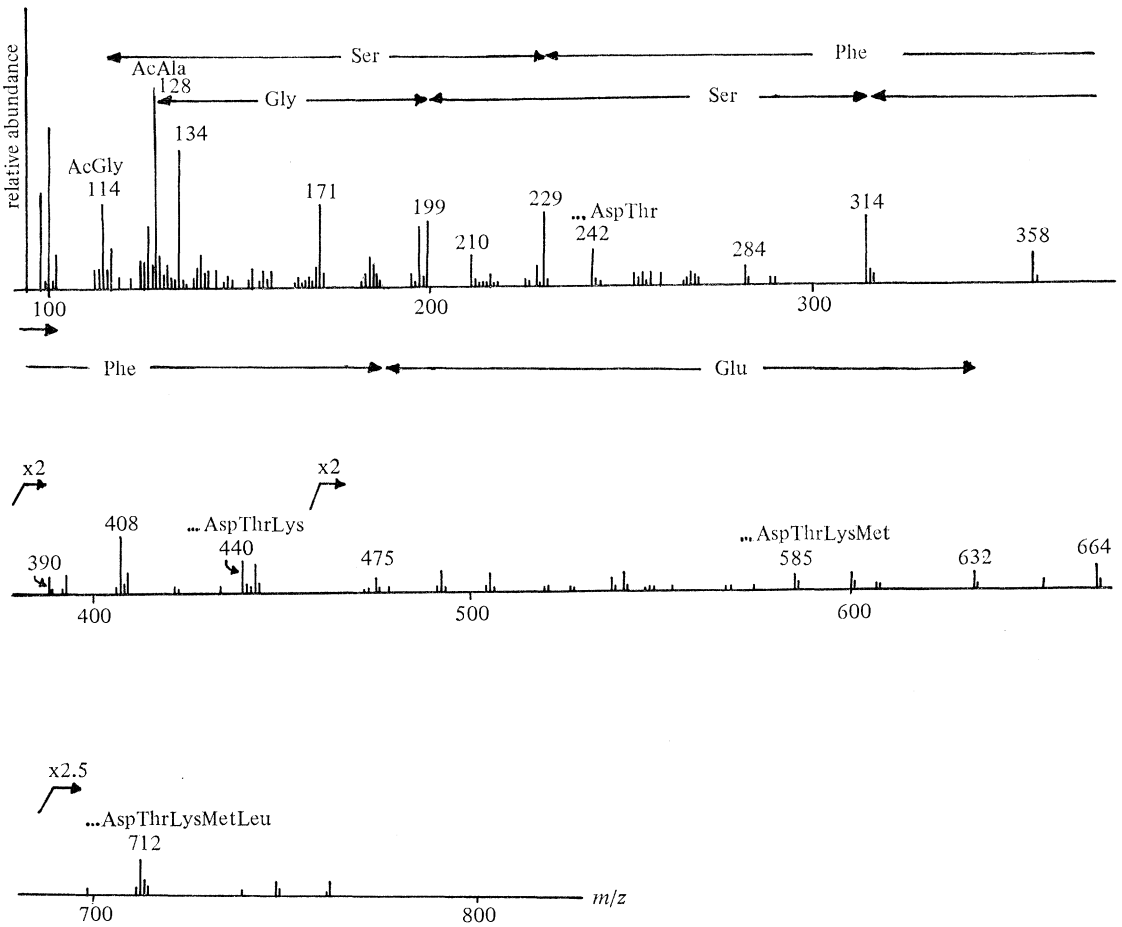


FIGURE 3. Mass spectrum of a derivative of a peptide fraction obtained by ion exchange chromatography of an elastase digest of dihydrofolate reductase.

We see two sequences emanating from normal sequence ions at m/z 114 (Gly) and m/z 128 (Ala) but also other signals at m/z 242, 440 and 585 which do not belong to any obvious *N*-terminus. Working backwards from m/z 585 we can see that m/z 440 corresponds to a methionine mass difference (145 u) and similarly m/z 242 corresponds to a lysine mass difference from m/z 440 (198 u). Both m/z 242 and m/z 440 can be seen to lose methanol indicating that they must be preceded by a hydroxyamino acid, either serine or threonine. Ion m/z 242 minus serine (115 u) gives m/z 127; a meaningless number for an '*N*-terminal' mass. However, m/z 242 minus threonine (129 u) gives m/z 113 which we assign to N–C cleavage at aspartic acid; from experience we know that m/z 113 is not a well stabilized fragment and is normally of very low abundance. In addition to the two obvious sequences we can in this way assign a

valuable sequence Asp–Thr–Lys–Met. At higher source temperatures, the sequences shown were extended and gave the sequence later assigned as residues 119–129 in table 1.

Dihydrofolate reductase (table 1) is the first protein of unknown sequence to be fully sequenced by mass spectrometry without the aid of classical methods. Interestingly, with a molecular mass of 18000, this must be the largest molecule ever to have had its structure fully determined by mass spectrometry!

Present developments in this laboratory include refinements to the method outlined above. One important improvement will be the automatic interpretation of mass spectra by computer, thus removing one of the skills of the procedure and making it a more viable alternative for the protein chemist. Good mass spectra can be produced from 50–100 nmol of peptide, and in more experienced hands 10–30 nmol is all that is required for sequence assignment. It is still normal to purify several micromoles of protein before a sequence study, but studies now in progress on derivative-forming techniques and new ionization methods could reduce this requirement even further. The dihydrofolate reductase work described above took place over several years, mainly because of difficulties in protein availability and purification; the total time spent during this interval amounted to some 24 man months. This compares quite favourably with the time needed for classical study. With the current developments in computer automation we expect to reduce dramatically the time needed for the sequence analysis of an enzyme of this size to approximately 2 man months, by using the strategy and techniques described in this section.

Glycopeptide analysis

An increasingly important group of substances to the biochemist are glycoproteins, providing cell surface recognition sites and other important functions. Structure–function studies pose an especially difficult problem at the structural level, so much so that few glycoproteins have been fully characterized.

The problem is basically fivefold: we need information on (1) protein sequence, (2) carbohydrate sequence, (3) nature of protein–carbohydrate linkage, e.g. *O*-glycosidic to serine, (4) nature of saccharide–saccharide linkage, e.g. $1 \rightarrow 4$, $1 \rightarrow 3$, (5) configuration of linkage(s).

We have been studying ways of using mass spectrometric methods in a minimum number of experiments to obtain information in the above five areas. Our work can be illustrated by reference to 'antifreeze glycopeptide 8' (AF8) a glycopeptide isolated from Antarctic fish blood plasma, and one of a family of glycoproteins some members of which are able to depress the freezing point of water some 500 times more effectively than NaCl on a molar basis. In this work (Morris & Thompson 1976; Morris *et al.* 1978) we found it possible to identify each of points 1–4 above by the study of just two derivatives.

The peptide sequence was established by studying the electron impact (*N*-terminal) and chemical ionization (*C*-terminal) spectra of an acetyl permethyl derivative of glycopeptide AF8. The spectrum showed evidence of heterogeneity in several positions of the 14 residue peptide sequence, and the points of attachment of carbohydrate were clearly visible by the assignment of dehydroamino acids, formed by β-elimination of carbohydrate during addition of the base in the permethylation reaction. Without any necessity for pre-separation, the eliminated permethyl carbohydrate was observed at a different source temperature (fractional vaporization) and the sequence assigned as hexose–aminohexose. Interpretation of linkage between saccharides is normally based upon relative abundances of fragment ions which are

much more pronounced in pertrimethylsilyl rather than permethyl derivatives (Kamerling *et al.* 1971). In a separate experiment, the carbohydrate was eliminated from the peptide by using sodium hydroxide, and again without any purification or separation, a trimethylsilyl derivative was prepared. Comparison of the spectra produced with those of standard $1 \rightarrow 4$ and $1 \rightarrow 3$ linked carbohydrates allowed an assignment of a $1 \rightarrow 3$ linkage in the amino disaccharide (table 2).

TABLE 2. STRUCTURE OF GLYCOPEPTIDE AF8 AS DETERMINED BY MASS SPECTROMETRY (MORRIS *ET AL.* 1978)

Ala–Ala–Thr–Ala–Ala–Thr–Pro–Ala–Thr–Ala–Ala–Thr–Pro–Ala
X X X X

Ala–Ala–Thr–Ala–Ala–Thr–Ala–Ala–Thr–Pro–Ala–Thr–Pro–Ala
X X X X

Ala–Ala–Thr–Ala–Ala–Thr–Ala–Ala–Thr–Ala–Ala–Thr–Pro–Ala
X X X X

X = β-D-galactosyl(1 → 3)α-*N*-acetylgalactosamine.

Thus one can deduce a remarkable amount of structural information from the two suggested experiments. While the overall structure assigned to glycopeptide AF8 in table 2 is quite complex, it is worth noting that had the carbohydrate portions been composed of, say, just hexoses or larger branched saccharides, then of course the problem would have been much greater. Distinguishing between hexoses is difficult because any single bond cleavage in the ring will result in indistinguishable ion structures, for example for glucose, galactose or mannose permethyl derivatives. The work of Reinhold *et al.* (1974) on monosaccharides suggested the value of stereoselective ring-forming derivatives for distinguishing between isomeric units. We have extended this work to disaccharides, and recent results on mixed benzene boronate trimethylsilyl derivatives show fragmentation differences allowing a distinction between terminal non-reducing glucose and galactose, and interestingly, clear signals apparently diagnostic of α- or β-configuration in, for example, isomaltose or gentiobiose respectively (M. R. Thompson & H. R. Morris, unpublished work). Carbohydrate structure can of course be quite complex, and these results are preliminary, but nevertheless there seems to be good reason for optimism that all of points 1–5 above may be determined in future studies on complex glycopeptides by using the mass spectrometric approach outlined here.

NEUROCHEMISTRY

Another area where mass spectrometry has played and is playing an important role is in studies on neuropeptides isolated from the central nervous system. Imagine that you have just isolated a substance from a brain homogenate which has morphine-like activity. This activity is lost on incubation of the substance with certain proteases, suggesting that it is a peptide. After several purification steps the amino acid analysis suggests that you have a few tens of nanomoles of material, but also that the substance still is not necessarily pure, and ultraviolet fluorescence data suggest that the molecule may contain a 'modified' tryptophan residue which would not have been detected by conventional amino acid analysis. This is the situation

in which J. Hughes and coworkers found themselves after isolating the endogenous ligand for the opiate receptor. Classical dansyl–Edman studies gave only partial data, not correlating well with the amino acid analysis. This situation is common in neurochemical studies, where the confidence of sequence assignment normally associated with protein sequence studies is not available when the sample is not necessarily of protein origin, and therefore may have unusual amino acids, unusual linkages, or may even not be a peptide at all.

The mass spectrometer gives a further degree of confidence in such structural assignments because it can handle 'impure' materials by the mixture analysis principle, and it operates on more exact physico-chemical properties of the compound in question than just a relative mobility on a t.l.c. plate.

The assignment of the definitive structure of the above substance, enkephalin, was made by studying the mass spectrum of an acetyl permethyl derivative, and enkephalin was found to be a mixture of two pentapeptides, Tyr–Gly–Gly–Phe–Met and Tyr–Gly–Gly–Phe–Leu (Hughes *et al.* 1975). The amino acid analysis data had indicated a longer peptide, but once this mass spectrometric assignment had been made and the presence of any tryptophan analogue discounted, then the classical data were found not to be in disagreement with this interpretation.

A number of biologically active peptides have been shown to possess blocked structures, and these pose considerable problems for classical methodology, but fewer problems for mass spectrometry since free peptides would be 'blocked' before study by making derivatives. A recent example of this type of problem was the study of adipokinetic hormone, isolated from locust and responsible for lipid mobilization during flight (Stone *et al.* 1976). Aminopeptidases and carboxypeptidases failed to destroy the biological activity, indicating a blocked peptide, and it was not possible to determine the full structure by classical methods for neither *N*-terminal nor *C*-terminal degradation was possible.

The mass spectrum of a permethyl derivative of adipokinetic hormone was complex but nevertheless interpretable, showing a pyrrolidone carboxylate (m/z 98 and 126) blocking the *N*-terminus, and a signal at m/z 428 interpreted as arising from an N–C cleavage at tryptophan and containing a *C*-terminal amide: ...Trp–Gly–Thr–NH_2. A further small sample of adipokinetic hormone was then derivatized with CD_3I to confirm assignments by comparing mass shifts. The spectra could be interpreted as arising from a fully blocked decapeptide,

PCA–Leu–Asn–Phe–Thr–Pro–Asn–Trp–Gly–Thr–NH_2.

In contrast to the enkephalin story, tryptophan had not been expected in this structure, but was of course observed as a clear (well stabilized) signal and placed in sequence without difficulty!

New amino acids

Another important area of biochemistry in which mass spectrometry is playing a significant role is in enzymic regulation and control in the blood coagulation system.

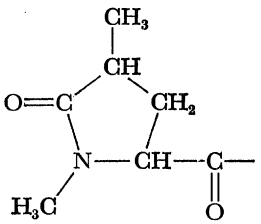
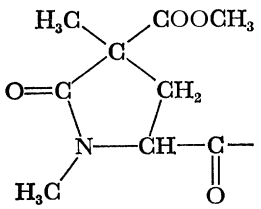
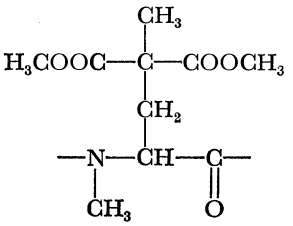
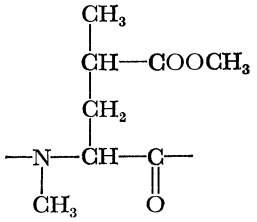
Perhaps the most obvious strength of the mass spectrometric method is in the area of new structures where no reference compounds exist with which to make a classical interpretation based upon, for example, comparative t.l.c. mobility. Just such a situation arose quite surprisingly in 1974 in a sequence study of one of the blood coagulating enzymes, prothrombin. The outcome of this work was the discovery of a new amino acid, γ-carboxyglutamic acid (Gla), and its sequence location in ten positions in the *N*-terminal region of prothrombin

(Magnusson *et al.* 1974; Stenflo *et al.* 1974; Morris *et al.* 1976). This new amino acid was missed in classical sequencing procedures (in fact it was assigned as ordinary glutamic acid) and we can now rationalize this as due to a ready decarboxylation during Edman degradation or hydrolysis in the dansyl procedure. Since this discovery, there has been considerable interest in the function of Gla and, in relation to this, finding whether any other proteins contain the amino acid.

TABLE 3. Gla IN FACTOR X

protein	Gla	basis of assignment	reference
X_1	14?	automated sequencer: poor yields of phenylthiohydantoins	Enfield *et al.* 1975
unseparated X	14	diborane reduction and quantitation of dihydroxyleucine	Howard & Nelsestuen 1975
mixed X_1 and X_2	11	sequencer: mass spectrometry of methyl phenylthiohydantoins	Bucher *et al.* 1976
X_1	7, 8	amino acid analysis of alkaline hydrolysates	Neal *et al.* 1976
X_2	13		

TABLE 4.

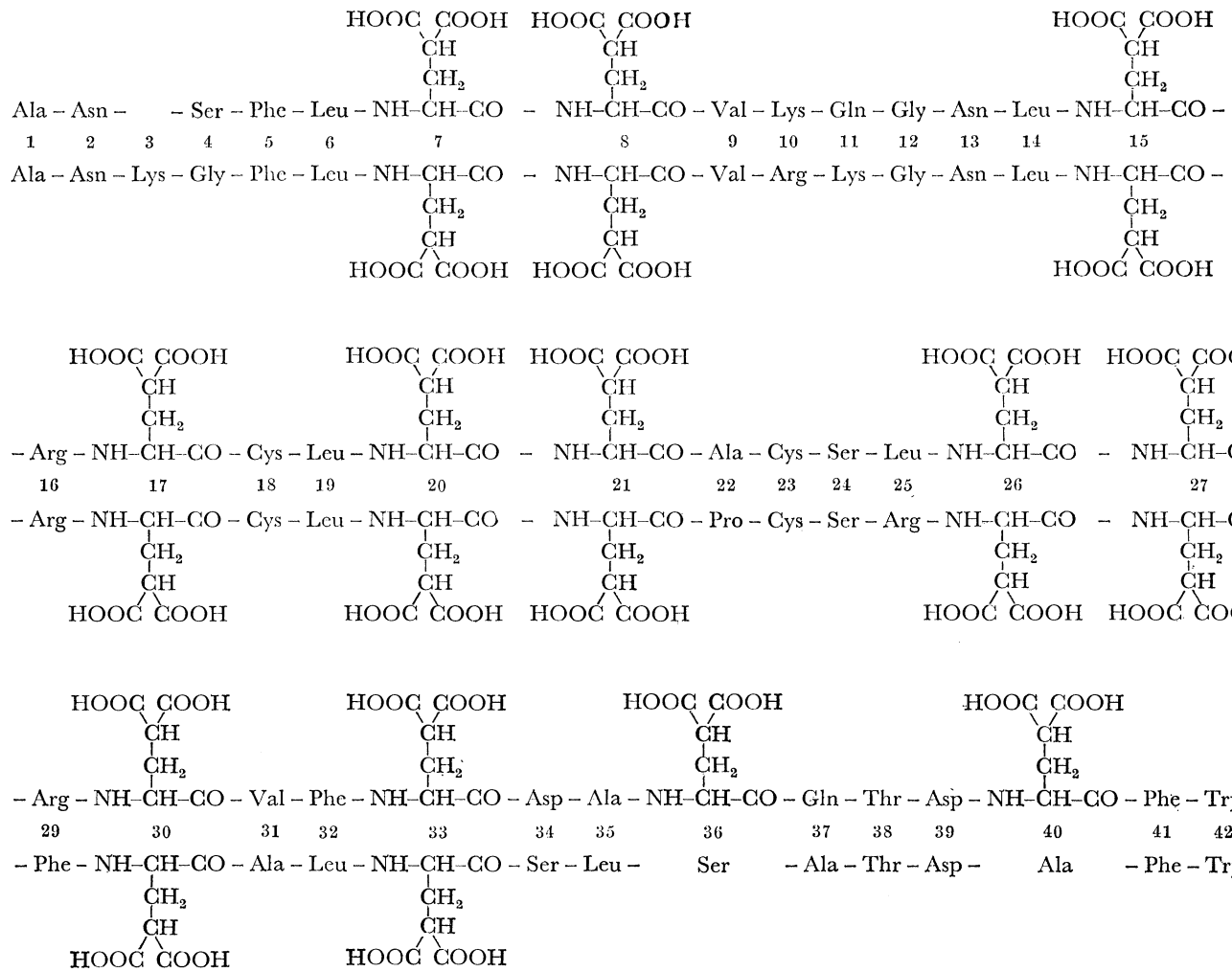
code	structure	occurrence	*N*-terminal mass	in-chain mass difference	common fragment ion
X		at *N*-terminus	140 (not observed)		(X minus CO)$^+$ (112 u)
X_c		at *N*-terminus	198 (not observed)		(X_c minus CO)$^+$ (170 u)
Y		in chain or at *C*-terminus	—	171	—
Y_c		in chain or at *C*-terminus	—	229	—

Factor X is another of the important zymogens in blood, and this protein has been shown to exist in two chromatographically distinct forms, factors X_1 and X_2. Both forms have biological activity, and an explanation for this phenomenon has been sought for a number of years. The discovery of Gla in prothrombin lead to several groups screening for the new amino acid in factor X (see table 3). As the table shows, different groups have obtained conflicting results, and attempts to explain the difference between factors X_1 and X_2 as due to different Gla content are unconvincing.

In our original work on the identification of Gla (Magnusson *et al.* 1974; Morris *et al.* 1976) we established two unique mass spectrometric signals, and two unique mass differences which are diagnostic for the new amino acid, given trivial names X, X_c, Y, Y_c in table 4.

The observation of these signals in a mass spectrum allows a definitive assignment not just of the presence of Gla in a peptide but also its exact location in the sequence. In a recent study of factor X_1 and X_2 we have isolated peptides from the *N*-terminal 42 residues of each protein and sequenced them by mass spectrometry with the methods outlined earlier. Our

TABLE 5. MASS SPECTROMETRIC SEQUENCE OF CALCIUM-BINDING REGIONS OF FACTOR X (*a*) AND BOVINE PROTHROMBIN (*b*)



results, shown in table 5, demonstrate that the data in table 3 are incorrect and that each of factors X_1 and X_2 contain 12 residues of the new amino acid Gla, and their sequences are identical (Thogersen *et al.* 1978). The difference between these two proteins must therefore reside elsewhere in the molecule, possibly in the carbohydrate attachment, and this is under investigation.

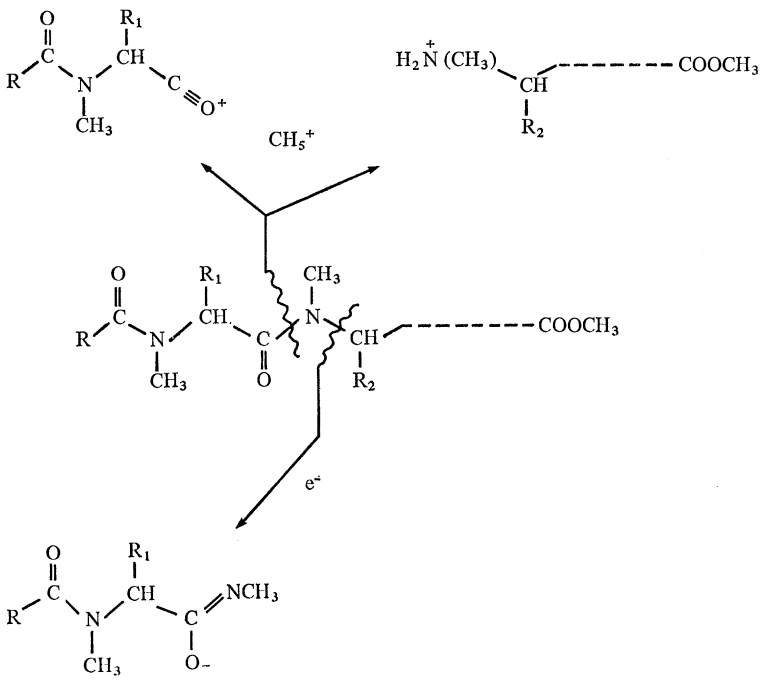


FIGURE 4. Fragmentation pathways observed for peptide derivatives under positive and negative chemical ionization analysis.

It is interesting to note that factor X, for some reason as yet unknown, has two γ-carboxyglutamic acids more than prothrombin (see table 5), but that we find 10 of the 12 residues in identical positions in both sequences. This basic structural information is now being used to formulate hypotheses for the role of Gla in blood coagulation control mechanisms. The formation of Gla is post-ribosomal and dependent on vitamin K, and it is reasonable to assume that one of its functions is to bind calcium. However, a large number of Ca^{2+}-binding proteins exist which do not require Gla in their sequences, making the necessity for this new amino acid more intriguing.

NEW DEVELOPMENTS

An interesting recent development in the peptide area is the study of negative chemical ionization. When acyl permethyl peptides are ionized in a conventional chemical ionization (c.i.) source, the resulting fragmentation pathways differ from those induced by electron impact (e.i.), yielding *C*-terminal fragments (see figure 4) of value to the structure elucidation. The methane gas normally used for c.i. studies produces thermal electrons when ionized and these can be used to generate low internal energy negative ions. The fragmentation of such acyl permethyl peptide negative ions differs from that observed in the positive ion mode as seen in figure 4, but can nevertheless be used for structure elucidation (A. Buko, D. F. Hunt & H. R. Morris, unpublished work). The attractive feature of negative ion c.i. is the possibility

for much enhanced sensitivity over the positive ion spectrum for components with positive electron affinities and large capture cross sections. The capture of thermal electrons can be promoted by 'tagging' the molecule with an efficient electron capturing group, e.g. a fluoroacyl group, and detailed studies of the effects of various blocking groups on peptide negative ion spectra are now under way. In some recent work we have attempted to combine a useful feature of protein chemistry, i.e. tagging the peptide with a fluorescent label, with negative ion c.i. to produce a spectrum of the 1:1 CH_3OH–CD_3OD labelled peptide ester (figure 5). The spectrum of approximately 1 nm of this sample shows a remarkably abundant $M^{-\cdot}$ signal at m/z 510, whereas no molecular ion is visible in the positive c.i. spectrum. The applicability of this technique to real problems in protein chemistry is now under investigation.

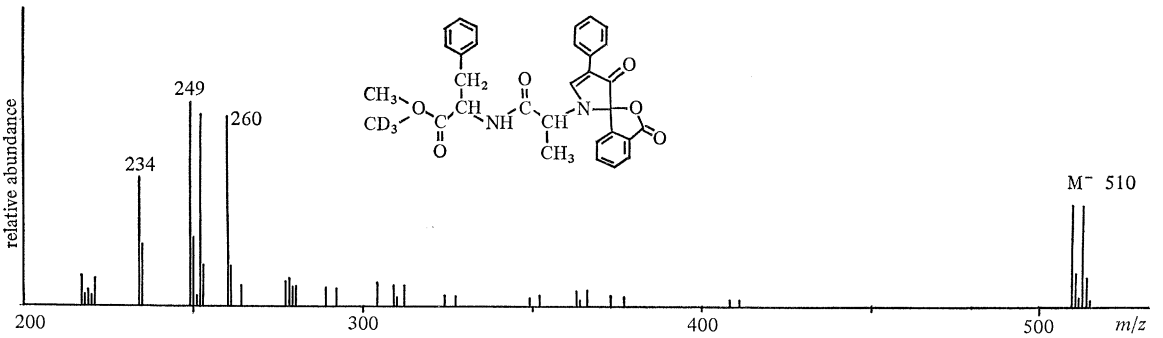


FIGURE 5. Negative ion chemical ionization spectrum of 1 nmol of a fluorescent derivative of Ala–Phe-ester.

Finally, the astute reader will have noted that I have said little about the problems associated with the high-mass nature of the samples studied. Even modest sized (8–12 residue) peptides and especially glycopeptides soon exceed the conventional mass spectrometer mass range of about 1000. To focus the higher mass ions, a reduction in accelerating voltage must be made, sometimes in the middle of the short volatilization time of the peptide, necessitating refocussing and leading to an inevitable loss of sensitivity. With this problem in mind, some years ago, I persuaded A.E.I.–Kratos to build a high-field magnet, which has now been operating at Imperial College for some months. The new magnet now allows us to reach masses of 3000 at full accelerating voltage and full sensitivity, tripling the conventional mass spectrometer mass range. Of course, it has been possible to obtain (usually poor quality) spectra beyond mass 1000 and in at least one special case (a fluorophosphazene reference) beyond mass 3000 on a conventional mass spectrometer by reducing the accelerating voltage. Knowledgeable and experienced mass spectroscopists also appreciate, however, that in the world of real organic or biochemical studies, considering the chemical nature, involatility and nanomolar sample size, it is quite futile to draw the analogy, and expect to generate spectra in the 3000 mass range from a conventional spectrometer.

The high-field magnet allows us to achieve this, and we are presently extending the scope of our electron impact, chemical ionization and field desorption studies to include much larger and more complex peptide, carbohydrate and glycopeptide molecules.

In summary, mass spectrometry research in the peptide and glycopeptide areas has made considerable progress over the past few years. The mass spectrometer is now a useful complement and a viable alternative to classical protein sequencing methods, and the first total protein sequence analysis by mass spectrometry alone has been accomplished. In other areas,

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

e.g. neuropeptides and coagulation, important problems have already been solved, and I have presented evidence that mass spectrometry is the method of choice here. Current work on new ionization–derivative combinations, computer automation and high mass studies promises to give us even greater sensitivity and to allow the solution of further important chemical and biochemical problems beyond the scope of present technology.

References (Morris)

Anderson, C. O. 1958 *Acta chem. scand.* **12**, 1353–1355.
Batley, K. E. & Morris, H. R. 1977*a* *Biochem. biophys. Res. Commun.* **4**, 1010–1014.
Batley, K. E. & Morris, H. R. 1977*b* *Biochem. Soc. Trans.* **5**, 1097–1099.
Biemann, K., Grapp, F. & Seibl, J. 1959 *J. Am. chem. Soc.* **81**, 2274–2276.
Bucker, D. *et al.* 1976 *FEBS Lett.* **68**, 293–296.
Das, B. C., Gero, S. D. & Lederer, E. 1967 *Biochem. biophys. Res. Commun.* **29**, 211–213.
Dell, A. & Morris, H. R. 1974 *Biochem. biophys. Res. Commun.* **61**, 1125–1132.
Dell, A. & Morris, H. R. 1976 In *Advances in mass spectrometry in biochemistry and medicine.* (ed. A. Frigerio & N. Castagnoli), vol. 1, pp. 263–266. New York: Spectrum Publications.
Dell, A. & Morris, H. R. 1977 *Biochem. biophys. Res. Commun.* **78**, 874–880.
Enfield, D. L. *et al.* 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 16–19.
Geddes, A. J., Graham, G. N., Morris, H. R., Lucos, F., Barber, M. & Wolstenholme, W. A. 1969 *Biochem. J.* **114**, 695–702.
Howard, J. B. & Nelsestuen, G. L. 1975 *Proc. natn. Acad. Sci. U.S.A.* **72**, 1281–1285.
Hughes, J., Smith, T. W., Kosterlitz, H. W., Fothergill, L., Morgan, B. A. & Morris, H. R. 1975 *Nature, Lond.* **258**, 577–579.
Kamerling, J. P., Vliegenthart, J. F. G., Vink, J. & de Ridder, J. J. 1971 *Tetrahedron* **27**, 4749–4757.
Krutzsch, H. 1978 In *Methods in enzymology.* (In the press.)
Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Morris, H. R. & Dell, A. 1974 *FEBS Lett.* **44**, 189–193.
Morris, H. R. 1972 *FEBS Lett.* **22**, 257–261.
Morris, H. R. 1974 *Biochem. Soc. Trans.* **2**, 806–809.
Morris, H. R. & Dell, A. 1975 In *Instrumentation in amino acid sequence analysis* (ed. R. N. Perham), pp. 147–191. London: Academic Press.
Morris, H. R. & Thompson, M. R. 1976 In *Advances in mass spectrometry in biochemistry and medicine* (ed. A. Frigerio & N. Castagnoli), vol. 1, pp. 239–245. New York: Spectrum Publications.
Morris, H. R., Geddes, A. J. & Graham, G. N. 1969 *Biochem. J.* **111**, 38P.
Morris, H. R., Dickinson, R. J. & Williams, D. H. 1973 *Biochem. biophys. Res. Commun.* **51**, 247–255.
Morris, H. R., Williams, D. H. & Ambler, R. P. 1971 *Biochem. J.* **125**, 189–201.
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.
Morris, H. R., Thompson, M. R., Osugo, D. T., Ahmed, A. I., Chan, S. M., Vandenheede, J. & Reeney, R. 1978 *J. biol. Chem.* **253**, 5155–5162.
Morris, H. R., Dell, A., Petersen, T. E., Sottrup-Jensen, L. & Magnusson, S. 1976 *Biochem. J.* **153**, 663–679.
Morris, H. R., Williams, D. H., Midwinter, G. G. & Hartley, B. S. 1974 *Biochem. J.* **141**, 701–713.
Neal, G. G. *et al.* 1976 *FEBS Lett.* **66**, 257–260.
Reinhold, N., Wirtz-Peitz, F. & Biemann, K. 1974 *Carbohydr. Res.* **37**, 203–221.
Stenflo, J., Fernlund, P., Egan, W. & Roepstorff, P. 1974 *Proc. natn. Acad. Sci. U.S.A.* **71**, 2730–2733.
Stone, J. V., Mordue, W., Batley, K. E. & Morris, H. R. 1976 *Nature, Lond.* **263**, 207–211.
Thomas, D. W., Das, B., Gero, S. D. & Lederer, E. 1968 *Biochem. biophys. Res. Commun.* **33**, 519–527.
Thogersen, H. C., Petersen, T. E., Sottrup-Jensen, L., Magnusson, S. & Morris, H. R. 1978 *Biochem. J.* **173**, 613–627.
Vilkas, E. & Lederer, E. 1968 *Tetrahedron Lett.* **26**, 3089–3091.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES