PRIMARY STRUCTURE DETERMINATION OF PEPTIDES AND PROTEINS BY MASS SPECTROMETRY

M. M. SHEMYAKIN

Institute for Chemistry of Natural Products, USSR Academy of Sciences, Moscow, USSR

A pressing problem in modern organic chemistry is the determination of the structure of polypeptides and proteins, for it is here that the needs and interests of biochemistry have far surpassed present chemical possibilities, while elucidation of many a biological problem on the molecular level requires a clear understanding of the relation between structure and biological function.

I should like to survey here the studies we have carried out in recent years that have led to the proposal of a new, general method for determining the amino acid sequence in peptides and proteins by means of mass spectrometry.

As is well known the primary structure of proteins determines the architecture of the molecule as a whole, including its conformational specificities and in the long run its biological properties. Therefore, determination of the amino acid sequence is one of the basic requisites for further progress in physicochemical and biological investigations of proteins.

What means do we actually possess for making such a determination?

At present the selective chemical or enzymatic degradation of the protein molecule into polypeptide fragments and their separation presents no basic difficulties. However, determination of the amino acid sequence in these polypeptide fragments rapidly and with unequivocal results is not always an easy task, particularly in view of the fact that in the majority of cases one has to deal with microquantities of substance. All methods used for this purpose are based on the stepwise chemical or enzymatic cleavage of the amino acid residues, beginning with the *N*- or *C*-terminus. Usually thereby the positions of not more than 1–3 residues can reliably be established. The only exception is Edman's phenylthiohydantoin method which allows determination of 4–6 *N*-terminal amino acid residues. Possibly, using Edman's automatic "sequenator" the method will permit determination directly in the protein chain of much larger fragments[1]. However, the sequenator has not left the experimental stage.

At the same time the problem of the primary structure elucidation of proteins is so important and the aims are so varied that it is necessary to have available several fundamentally differing methods, not only for the sake of mutual control, but also for the solution of the particular problems. For instance, the sequenator cannot be used to determine the primary structure

of the protein chain in its entirety. Neither can it be used with peptides obtained by partial enzymatic cleavage of native or modified proteins, whereas such peptides could yield valuable information on the secondary or tertiary structure and on the participation of separate groupings or regions of the protein chain in biochemical functions. In this case, as in many others, it is necessary selectively to split the proteins and then determine the structure of the resultant peptides. It is for just such a purpose that we have developed[2-19] a micromethod for determining the amino acid sequence in peptides based on analysis of the mass spectra of their *N*-acyl esters. This method considerably accelerates and simplifies the structural elucidation of proteins.

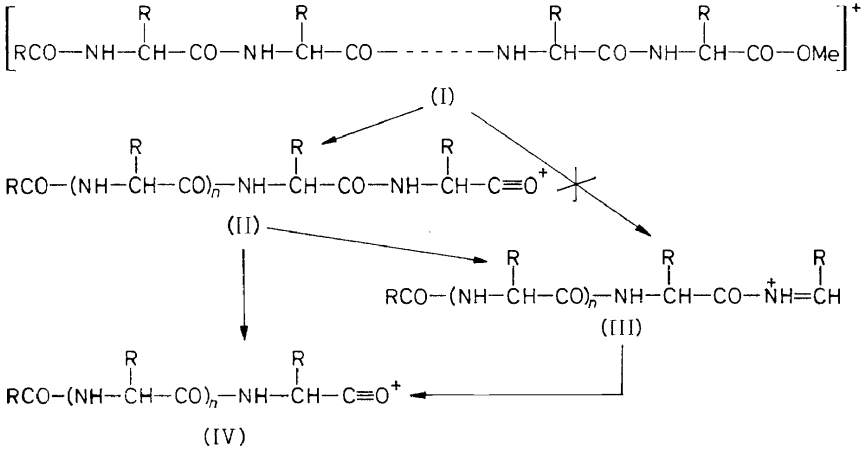
The mass spectrometric method is being very intensively used for determining the structure of steroids, terpenoids, alkaloids, antibiotics, etc., but it has found very little application as yet in structural studies of peptides and proteins. Owing to the very low volatility of free peptides, their mass spectra have to be obtained at high temperatures, at which pyrolysis begins to tell on the results[20, 21]. In order to increase the volatility of the peptides it was proposed to subject them to partial hydrolysis followed by reduction to the polyamino alcohols (Biemann[22-24]) or to convert them into the trifluoroacetyl esters (Weygand[25, 26]); their separation is then achieved by gas liquid chromatography and their analysis by mass spectrometry. However, owing to preparative difficulties neither method found actual use in the structural elucidation of naturally occurring peptides.

New possibilities arose after our discovery in 1965 of the amino (hydroxy) acid type of fragmentation of *N*-acyl esters of peptides and depsipeptides[2], simultaneously observed independently of us on other compounds by Lederer[27-29] and Weygand[30].

Based on an analysis of the mass spectra of a large number of various *N*-acylpeptide esters we drew the following general conclusion: the primary fission of the molecular ion (I) at an ester or amide bond gives rise to the ion (II) with positive charge localized on the *C*-terminus. Following this, stepwise elimination of the amino acid residues from the *C*-terminus occurs, accompanied by migration of the positive charge along the chain. Although fragmentation of the peptide chain in the ion (I) may begin with fission of any one of the amide bonds, this does not distort the general mass spectral picture, affecting only some peak intensities. The resultant fragment of type (II) undergoes further consecutive elimination of the amino acid residues in a single stage (II → IV) or in two stages (II → III → IV). In the latter case intermediate aldimine fragments of type (III) arise which usually are not formed directly from the molecular ion (I).

It is this type of fragmentation which makes possible the amino acid sequence determination in peptides. Therefore, soon after, a number of workers (Biemann[31, 32], McLafferty[33, 34], Barber[35]) proposed to utilize it for programming mass spectral data with the aim of obtaining by means of computers both amino acid composition and sequence. Other workers (Lederer's group[27-29, 36-39], also Weygand's group[40]) directed their attention to employing mass spectrometry for determining the amino acid sequence of relatively simple naturally occurring peptides. To us, however, it was clear that the exceptional possibilities afforded by mass spectrometry for the

rapid determination of the primary structure of peptides and proteins on a microlevel could be realized to the full only after preliminary elucidation of a number of problems that would lead to the creation of a new, sufficiently universal method. It was due to the absence of such a method that chemists and biochemists interested in the amino acid sequence of peptides and proteins had not given mass spectrometry serious attention.



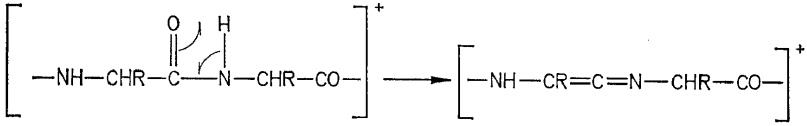
In order for mass spectrometry to achieve the place it deserved in this area, we deemed it necessary first to be able to prepare for the mass spectrometric investigation any peptide built up of the usual amino acids contained in proteins; secondly, to ascertain how the nature and position of the amino acid residues in the peptide chain and also their chemical environment affect the general rule of fragmentation (amino acid type) and also all particular fragmentation rules; thirdly, one had to know the limits of the mass spectrometric method and the possibilities for their extension; and, finally, one had to learn to read correctly the mass spectra, either directly or with the aid of a computer. All this has now been realized.

A study of about 200 *N*-acylpeptide esters of differing amino acid composition and molecular weight (*Tables 1–10*) showed all peptides excepting those containing arginine to give mass spectra with a well expressed pattern of the amino acid type of fragmentation and that the arginine-containing peptides can be also made to undergo this type of fragmentation by modifying the guanidine grouping[2–19].

Fragmentation of the peptide chain is usually accompanied by dehydration of amide bonds (V→ VI). Therefore, the spectra of practically all acylpeptide esters show an M-18 peak. Sometimes this peak is very prominent and in such cases the peaks of the amino acid fragments are often accompanied by 18 m.u. lower satellite peaks. The dehydration process is usually more marked in high molecular and low volatile peptides. Glycine residues or residues of aromatic or heterocyclic acids also facilitate dehydration[16, 18].

Fragmentation of the peptide chain always occurs simultaneously with other processes, mainly connected with the fragmentation of the side chain of each amino acid residue and obeying definite rules. These particular rules

give additional information as to the structure of the peptide, permitting a more precise decoding of the mass spectral data. The basic amino acid type of peptide chain fragmentation and all the particularities observed have been confirmed by studies of the metastable peaks, using isotope labelling, and, when necessary, high resolution mass spectrometry.



(V) (VI)

Turning to the particular rules of fragmentation, first of all it is to be mentioned that in the case of branched chain aliphatic monoamino carboxylic acids (valine, leucine and isoleucine; *Table 1*) a characteristic feature is the elimination of all or part of the side chain as a radical (VII → VIII) or

Table 1. List of aliphatic monoamino carboxylic acid-containing acylpeptide esters investigated mass spectrometrically[5-7, 16-18]

Dec–*Leu*–Ala–Ala–Ala–OMe
Dec–Gly–*Val*–*Leu*–OMe
Dec–Asp(OMe-β)–*Leu*–*Val*–*Val*–OMe
Dec–Gly–Gly–*Leu*–Gly–OMe
Prop–Phe–Gly–*Leu*–Gly–*Leu*–Ala–OMe
Dec–Gly–*Leu*–Gly–*Leu*–Ala–OMe
Dec–Asp(OMe-β)–*Val*–*Leu*–OMe
Dec–Gly–*Val*–Gly–*Leu*–OMe
Dec–Gly–*Ile*–Gly–*Leu*–Ala–OMe
Prop–Phe–Gly–*Leu*–Gly–*Leu*–Ala–OBu^t
Ste–Ala–*Leu*–Gly–*Leu*–*Val*–Pro–Ala–Pro–*Val*–OMe
Hex–*Ile*–*Leu*–Gly–Asp(OMe-β)–*Val*–Phe–OMe
Dec–Asp(OBu^t-β)–*Leu*–Gly–*Leu*–OBu^t
Prop–Phe–*Val*–*Leu*–Ala–OMe
TFA–Try–*Leu*–*Val*–Pro–*Leu*–Ala–OMe
Dec–*Val*–Asn–*Leu*–OBu^t
Prop–Phe–Gly–*Leu*–Ala–OMe
TFA–Phe–Gly–*Leu*–Gly–*Leu*–Ala–OMe
Dec–*Val*–*Val*–Gly–Try–OMe
Ste–Phe–Gly–*Leu*–Gly–*Leu*–Ala–OMe

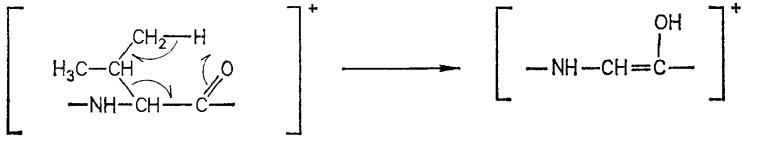
Prop = CH_3CH_2CO; Hex = $CH_3(CH_2)_4CO$; Dec = $CH_3(CH_2)_8CO$; Ste = $CH_3(CH_2)_{16}CO$

olefin (IX → X). Usually the peaks corresponding to the side chain deprived fragments are considerably weaker than those corresponding to the amino acid type of fragmentation. Cleavage of the side chain is much more pronounced in methionine-containing peptides (*Table 2*); here most often the strongest peaks are those corresponding to the elimination of the entire side chain (XI → XII) whereas the C_β—C_γ and C_γ—S bond fissions which occur simultaneously are characterized by peaks of low intensity[5-7, 16-18].

Typical of proline-containing peptides (*Table 3*) is rupture of the hetero-ring (XIII) ← (XIV) → (XV) following formation of the corresponding aldimine fragment in the course of cleavage of the peptide chain[5-7, 16-18].

Fragmentation of petpides containing hydroxyamino acid residues (serine,

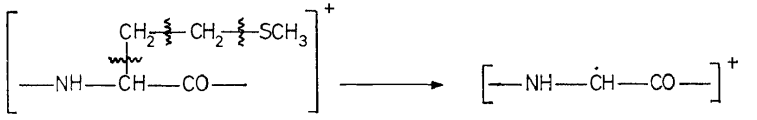
(VII) (VIII)



(IX) (X)

Table 2. List of methionine-containing acylpeptide esters investigated mass spectrometrically[5–7, 16–18]

Dec–Gly–*Met*–Ala–Ala–OMe
Dec–*Met*–Ala–Val–Leu–OMe
Dec–Leu–Gly–Gly–*Met*–OMe
Dec–Leu–Gly–Gly–*Met*–Leu–Gly–OMe
Dec–Gly–Pro–*Met*–OMe
Dec–Gly–Leu–Gly–*Met*–Leu–OMe
Dec–*Met*–Phe–His–Leu–OMe
Dec–*Met*–Try–Val–Leu–Ala–OMe
Dec–Asn–Phe–Ala–Gly–*Met*–OMe
Dec–Gly–*Met*–Gly–OBu^t
Dec–*Met*–Asp(OMe–β)–Val–Gly–Leu–OMe
Dec–*Met*–Pro–Val–Gly–OMe
Dec–Gly–*Met*–Gly–Leu–Ala–OMe
Dec–*Met*–Ala–Pro–Gly–OMe
Dec–*Met*–Gln–Ala–Ile–Gly–OMe
Dec–Phe–*Met*–Glu(OMe-γ)–Phe–Leu–OMe
Dec–Gln–Phe–Ala–Gly–*Met*–OMe



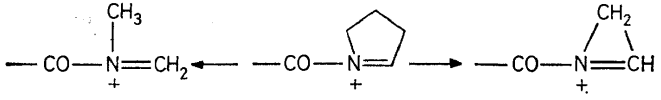
(XI) (XII)

threonine, etc.; *Table 4*) are usually accompanied by dehydration of the side chain (XVI → XVII); also typical of the threonine residue is elimination of the whole side chain accompanied by migration of hydrogen (XVIII → XIX)[5–7, 16–19](cf.[28]).

A very characteristic feature of peptides containing the esters or amides of aspartic or glutamic acid (*Table 5*) is transformation of the ω-substituent into a ketene (XX → XXI ← XXII), elimination of ammonia being a thermal process, while splitting out of the elements of alcohol is mainly the result of electron impact. Sometimes expulsion of the ω-substituent as a whole occurs, an act that can be accompanied by hydrogen loss from the α-carbon atom (XXIII → XXIV → XXV)[4–7, 9, 12, 13, 17 (cf. 37)].

Table 3. List of proline-containing acylpeptide esters investigated mass spectrometrically[5–7, 16–18]

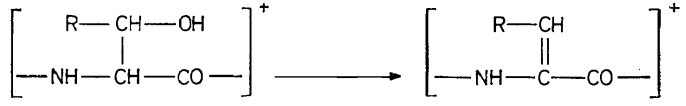
Dec–*Pro*–Ala–Ala–Val–OMe
Dec–Gly–Phe–Ala–*Pro*–Val–Gly–OMe
Dec–Gly–Tyr–*Pro*–Val–Gly–OMe
Dec–*Pro*–Val–Gly–OMe
Dec–Leu–Ala–*Pro*–Gly–OMe
Dec–Val–Ala–*Pro*–Gly–OMe
Dec–Gly–Gly–*Pro*–OMe
Dec–Try–Leu–Val–*Pro*–Ala–*Pro*–Val–OMe
Prop–Try–Leu–Val–*Pro*–Leu–Ala–OMe



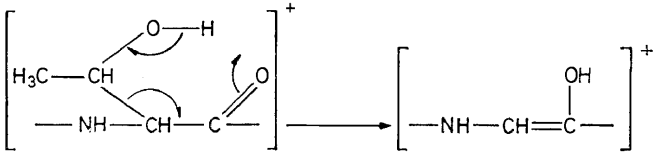
(XIII) (XIV) (XV)

Table 4. List of hydroxyamino acid-containing acylpeptide esters investigated mass spectrometrically[5–7, 16–19]

Dec–Phe–*Ser*–Gly–Ala–OMe
Dec–*Hypro*–Phe–His–Leu–OMe
Dec–*Thr*–Phe–His–Leu–OMe
Dec–*Hypro*–Try–Val–Leu–Ala–OMe
Dec–*Thr*–Pro–Val–Gly–OMe
Dec–*Thr*–Try–Val–Leu–Ala–OMe
Dec–*Hypro*–Leu–Gly–Leu–Ala–OMe
Dec–*Ser*–Try–Val–Leu–Ala–OMe
Dec–*Ser*–Phe–His–Leu–OMe
Dec–Asn–*Ser*–Ile–Gly–OMe
Dec–*Ser*–Phe–Phe–*Ser*–OMe
Dec–Glu(OMe-γ)–*Ser*–Ile–Gly–OMe
Dec–Glu(OMe-γ)–Gly(OMe-γ)–*Thr*–Phe–OMe
Dec–Leu–Gly–Gly–*Homoser*–OMe
Dec–Gly–Ala–Pro–Gly–*Homoser*–OMe
Dec–Ala–Pro–Gly–*Homoser*–OMe



(XVI) (XVII)



(XVIII) (XIX)

Table 5. List of monoamino dicarboxylic acid-containing acylpeptide esters investigated mass spectrometrically[4–7, 9, 12, 13, 17]

Dec–*Asp*(*OMe-β*)–Ala–Val–Phe–OMe
Dec–*Asp*(*OMe-α*)–Ala–Val–Phe–OMe
Dec–*Glu*(*OMe-γ*)–His–Leu–OMe
Dec–*Glu*(*OMe-γ*)–Try–*Glu*(*OMe-γ*)–Phe–Leu–OMe
Dec–*Asn*–Phe–His–Leu–OMe
Dec–*Asn*–Ala–*Glu*(*OMe-γ*)–Phe–Leu–OMe
Dec–Gly–*Glu*(*OMe-γ*)-Tyr–OMe
Dec–*Glu*(*OMe-γ*)–Ala–Ala–Ala–OMe
Dec–*Glu*(*OMe-γ*)–Phe–His–Leu–OMe
Dec–*Asn*–Try–Val–Leu–Ala–OMe
Dec–*Asn*–Ala–Ala–Ala–OMe
Dec–Gly–Phe–Gly–Phe–*Gln*–OMe
Dec–*Asn*–Ile–Ala–Ala–OMe
Dec–Gly–*Glu*(*OMe-γ*)–Ala–Ala–Phe–OMe
Dec–Gly–*Glu*(*OMe-α*)–Ala–Ala–Phe–OMe
Dec–Gly–Leu–Val–*Glu*(*OMe-γ*)–Gly–Ala–Leu–OMe
Dec–Gly–Leu–Val–*Glu*(*OMe-α*)–Gly–Ala–Leu–OMe
Dec–Try–Val–Ala–*Asn*–Leu–OMe
Dec–*Glu*(*OMe-γ*)–Gly–Ala–Leu–OMe
Dec–*Glu*(*OMe-α*)–Gly–Ala–Leu–OMe
Dec–*Glu*(*OMe-γ*)–Ala–Gly–OMe
Dec–*Glu*(*OMe-α*)–Ala–Gly–OMe
Dec–*Asp*(*OMe-β*)–Val–Leu–OMe
Dec–*Asp*–(*OMe-α*)–Val–Leu–OMe
Dec–*Gln*–Phe–Phe–Met–OMe
Dec–Gly–*Asn*–Ala–Ala–Phe–OMe
Dec–*Asp*(*OMe-β*)–Val–Ala–Leu–OBu^t
Dec–*Gln*–Ile–Ala–Ala–OMe
Dec–Ala–Pro–*Glu*$(OMe)_2$
Dec–Ala–*Asp*(*OMe-β*)–Gly–OMe
Dec–Gly–*Asn*–OMe
Dec–Ala–Ala–Ala–*Asp*$(OMe)_2$
Dec–Gly–Gly–Sar–*Gln*–Phe–OMe
Dec–*Gln*–Val–Ala–*Asn*–Leu–OMe

Of considerable interest is the mass spectrometric behaviour of peptides containing aromatic and heterocyclic amino acids (*Table 6*), the more so since they are often found to be *C*-terminal in the chymotrypsin cleavage of proteins. The following three processes are the most characteristic of peptides containing phenylalanine, tyrosine, histidine and tryptophane residues[5–7, 10, 18] (cf. [30, 41]): (a) elimination in the course of the amino acid type of fragmentation of the side chain as the radical $ArCH_2^{\cdot}$ (XXVI → XXVII); (b) fission of the N—C_α bond of the aromatic (heterocyclic) amino acid residue, followed by amino acid type fragmentation of the resultant *C*-terminal fragment (XXVI → XXVIII); (c) elimination of the side chain as the cation $ArCH_2^{\oplus}$ (XXVI → XXIX). All three processes are most strongly expressed in the tryptophane-containing peptides. In addition a characteristic feature of histidine and tryptophane residues is the thermal intermolecular *N*-methylation of the hetero rings (e.g. XXX → XXXI) (cf. [34, 42]).

The presence of sulphur-containing amino acid residues in the peptides, including residues modified in the process of protein cleavage causes no difficulty in mass spectral analysis (*Table 7*). With cysteine derivatives of type (XXXII), in addition to amino acid fragmentation of the peptide

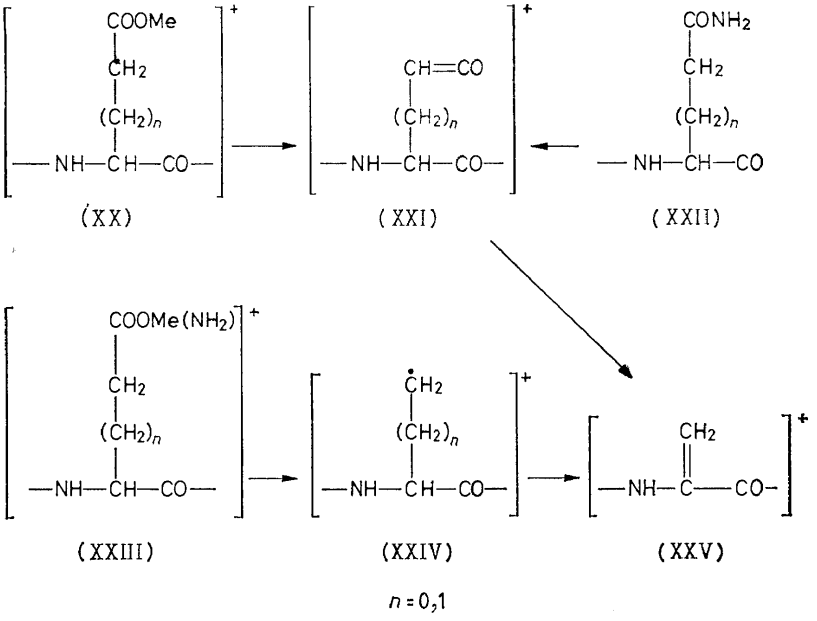


Table 6. List of aromatic and heterocyclic amino acid-containing acylpeptide esters investigated mass spectrometrically[5–7, 10, 18]

Pent–*Try*–Leu–Val–Pro–Leu–Ala–OMe
Oct–*Try*–Leu–Val–Pro–Leu–Ala–OMe
Prop–*Try*–*Try*–Val–Leu–Ala–OMe
Dec–Val–*Phe*–*Try*–*Tyr*–OMe
Dec–*Tyr*–*Try*–Val–Leu–Ala–OMe
Dec–*Tyr*–*Phe*–*His*–Leu–OMe
Dec–*His*–*Try*–Val–Leu–Ala–OMe
Dec–*Tyr*–Ala–Val–Ala–OMe
Ste–*Try*–*Phe*–*His*–Leu–OMe
Ste–*Try*–*His*–Leu–OMe
Dec–*Try*–Leu–Val–Pro–Leu–Ala–OMe
Prop–*Try*–Gly–Leu–Ala–OMe
Ac–*Phe*–Gly–Leu–Gly–Leu–Ala–OMe
Dec–*His*–Leu–Ala–Leu–OMe
Dec–Ala–*His*–Leu–OMe
Dec–*Try*–Val–Gly–Val–OMe
Dec–*His*–Ala–Pro–Val–OBu^t
Dec–*Try*–Leu–Ala–*His*–OMe
Dec–*Phe*–Gly–*His*–Leu–OMe
Ac–*Try*–Val–Leu–Ala–OMe
Ste–*Try*–Leu–Val–Pro–Leu–Ala–OMe
Dec–*His*–Ala–*His*–Leu–OMe
Dec–Val–*Tyr*–Val–Ala–OMe
Dec–Gly–*Try*–Ala–Ala–OMe
Prop–*Try*–Leu–Val–Pro–Leu–Ala–OBu^t
Dec–*Phe*–*His*–Leu–OMe
Dec–*Phe*–Gly–Leu–Gly–Leu–Ala–OMe
Ac–*Try*–Leu–Val–Pro–Leu–Ala–OMe
Dec–*Phe*–Gly–Asn–*Phe*–Val–Leu–OMe
Dec–Glu(OMe-γ)–*Try*–Val–Leu–Ala–OMe

Pent = $CH_3(CH_2)_3CO$; Oct = $CH_3(CH_2)_6CO$

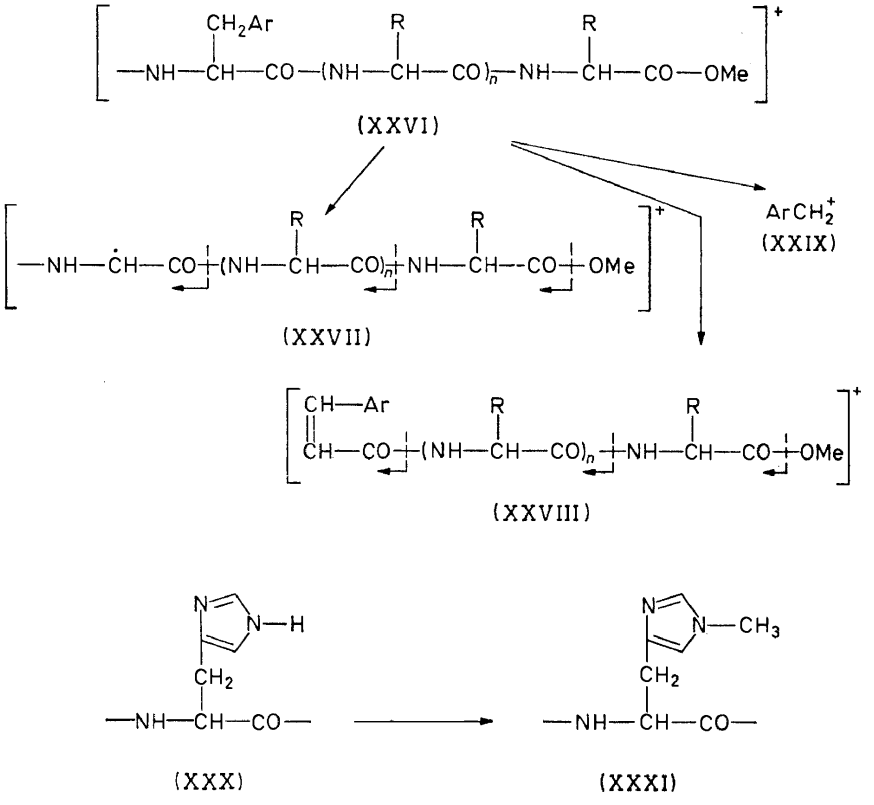
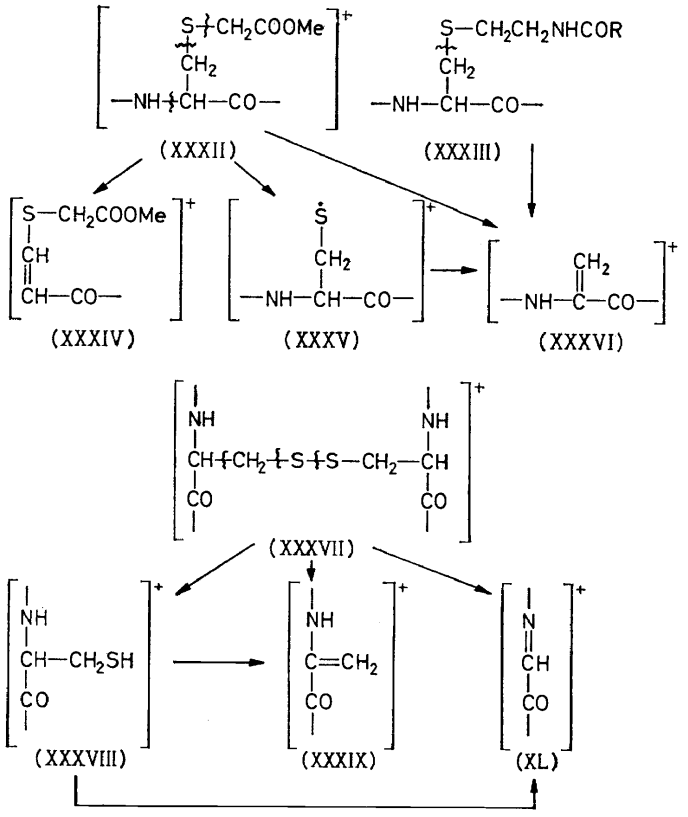


Table 7. List of cysteine- and cystine-containing acylpeptide esters investigated mass spectrometrically[11, 19]

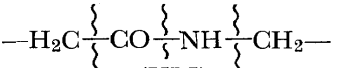
(Dec–Glu(OMe-α)–*Cys*–Gly–OMe)$_2$
(Dec–*Cys*–Phe–Leu–Gly–OMe)$_2$
(Prop–*Cys*–Phe–Leu–Gly–OMe)$_2$
(Dec–Phe–Leu–Gly–*Cys*–OMe)$_2$
Dec–*Cys*(*CH$_2$COOMe*)–*Cys*(*CH$_2$COOMe*)–Gly–Val–OMe
Prop–Phe–*Cys*(*CH$_2$COOMe*)–Gly–OMe
Dec–*Cys*(*CH$_2$COOMe*)–Leu–Lys(Dec)–OMe
Dec–Gly–Phe–Gly–*Cys*(*CH$_2$COOMe*)–OMe
Dec–Phe–Asp(OMe-β)–*Cys*(*CH$_2$COOMe*)–OMe
Dec–*Cys*(*CH$_2$COOMe*)–Gly–Val–OMe
Dec–*Cys*(*CH$_2$COOMe*)–Phe–Leu–Gly–OMe
Dec–Phe–Leu–Gly–*Cys*(*CH$_2$COOMe*)–OMe
Dec–Gly–*Cys*(*CH$_2$COOMe*)–OMe
Dec–Met–*Cys*(*CH$_2$CONH$_2$*)–Leu–OMe
Prop–Phe–Leu–Gly–*Cys*(*CH$_2$CH$_2$NHDec*)–OMe
Dec–*Cys*(*CH$_2$CH$_2$NHDec*)–Phe–Leu–Gly–OMe
Dec–Phe–Leu–Gly–*Cys*(*CH$_2$CH$_2$NHDec*)–OMe

chain, characteristic fragments (XXXIV), (XXXV) and (XXXVI) are formed by fission of the corresponding bonds, while in the case of cysteine derivatives of type (XXXIII) the dehydroalanine fragments (XXXVI)

are prevalent. With cystine-containing peptides (XXXVII), rupture of the corresponding bonds in the side chain leads to the appearance of ions (XXXVIII), (XXXIX) and (XL)[11, 19].



Since the basic amino acids lysine and arginine are widespread in proteins and trypsin cleavage leaves them on the *C*-terminus of the resultant peptide, the mass spectrometric determination of the amino acid sequence in such peptides is of particular interest. Providing the ω-amino group of the lysine (and also ornithine) residues is protected by acylation simultaneously with the *N*-terminal α-amino group, peptides containing these amino acid residues (*Table 8*) normally undergo the amino acid type of fragmentation[5–7, 16]. Here it is of interest to mention only fission of the C—C and C—N bonds of the ω-amide group (XLI) often accompanied by migration of hydrogen. These processes are similar to the well known fragmentation of secondary amides.



(XLI)

A special case proved to be that of arginine-containing peptides, because their thermal lability precludes their direct mass spectrometry. Protection

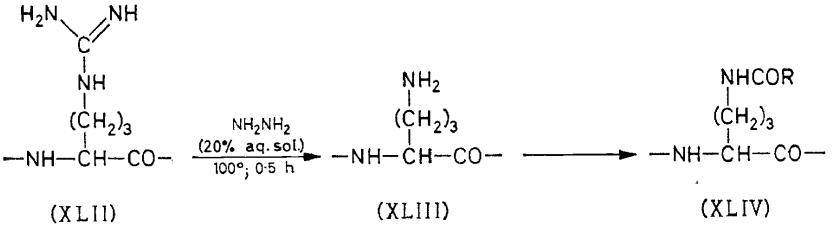
Table 8. List of lysine-containing acylpeptide esters investigated mass spectrometrically[5–7, 16]

Dec–*Lys*(*Dec*)–Phe–Phe–Ser–OMe
Dec–*Lys*(*Dec*)–Thr–Pro–Ala–OMe
Dec–Leu–Gly–Gly–*Lys*(*Dec*)–OMe
Dec–Val–Phe–Try–*Lys*(*Hex*)–OMe
Dec–Glu(OMe-γ)–Phe–*Lys*(*Dec*)–OMe
Hex–*Lys*(*Hex*)–Try–Val–Leu–Ala–OMe
Hex–*Lys*(*Hex*)–Phe–His–Leu–OMe
Dec–*Lys*(*Dec*)–Pro–Val–Gly–OMe

of the guanidine grouping by diacylation is also ineffective; owing to thermal instability of the compounds and predominant localization of the positive charge on the diacylguanidine grouping, the peaks characteristic of the amino acid type of fragmentation are absent in the mass spectra. However, the arginine residue in peptides can be transformed into an ornithine residue by hydrazinolysis, which rarely affects the amide bonds in the peptide chain[7, 8, 14, 16]. After acylation and esterification the resultant esters of the N^{α},N^{ω}-diacylornithine containing peptides (*Table 9*) are quite suitable for

Table 9. List of ornithine-containing acylpeptide esters prepared from arginine-containing peptides and investigated mass spectrometrically[7, 8, 14, 16]

Dec–Try–Phe–*Orn*(*Dec*)–OMe
Dec–Try–*Orn*(*Dec*)–Leu–OMe
Dec–*Orn*(*Dec*)–Leu–Gly–Leu–Ala–OMe
Dec–Phe–*Orn*(*Dec*)–OMe
Dec–Phe–*Orn*(*Dec*)–Leu–Ala–OMe
Dec–Phe–*Orn*(*Dec*)–Leu–OMe
Dec–Ala–Leu–*Orn*(*Dec*)–OMe
Dec–*Orn*(*Dec*)–Leu–OMe

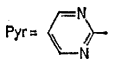


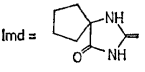
(XLII) (XLIII) (XLIV)

mass spectrometric determination of the amino acid sequence (XLII → XLIII → XLIV). Another method[7, 8, 14, 16] involves condensation of the guanidine group of the arginine residue in N^{α}-acylpeptides with β-dicarbonyl compounds (1,1,3,3-tetra-alkoxypropane or acetylacetone) under acid conditions, or with α-dicarbonyl compounds (for instance cyclohexane-1,2-dione) under weakly alkaline conditions (XLV ← XLVI → XLVII). Both types of compounds—(XLV) and (XLVII) (*Table 10*) are suitable for mass spectrometry, but the first is preferable because of its higher volatility. They give a sufficiently well expressed amino acid type of fragmentation which is little affected by the specificities in the fragmentation

Table 10. List of N^{δ}-substituted ornithine-containing acylpeptide esters prepared from arginine-containing peptides and investigated mass spectrometrically[7, 8, 14, 16]

Dec–Thr–*Orn(Pyr)*–Leu–Phe–OMe
Dec–Tyr–*Orn(Pyr)*–Leu–Phe–OMe
Dec–*Orn(Pyr)*–Leu–Val–*Orn(Pyr)*–OMe
Dec–His–*Orn(Pyr)*–Leu–Phe–OMe
Dec–Gly–Ser–*Orn(Pyr)*–OMe
Dec–Phe–His–Leu–*Orn(Pyr)*–Leu–OMe
Dec–Val–*Orn(Pyr)*–*Orn(Pyr)*–Leu–OMe
Dec–Met–Phe–*Orn(Pyr)*–OMe
Dec–Phe–*Orn(Pyr)*–Lys(Hex)–OMe
Dec–Val–*Orn(Pyr)*–*Orn(Pyr)*–OMe
Dec–*Orn(Pyr)*–*Orn(Pyr)*–Leu–OMe
Dec–Glu(OMe-γ)–Phe–*Orn(Pyr)*–OMe
Dec–Asn–Phe–*Orn(Pyr)*–OMe
Dec–*Orn(Pyr)*–Leu–*Orn(Pyr)*–OMe
Dec–Hypro–*Orn(Pyr)*–Leu–Phe–OMe
Dec–*Orn(Pyr)*–Leu–Ala–OMe
Dec–Phe–*Orn(Pyr)*–Leu–OMe
Dec–*Orn(Pyr)*–Leu–Gly–Leu–Ala–OMe
Dec–Phe–*Orn(Me$_2$Pyr)*–Leu–OMe
Dec–*Orn(Me$_2$Pyr)*–Leu–Ala–OMe
Dec–Phe–Phe–*Orn(Me$_2$Pyr)*–OMe
Dec–*Orn(Imd)*–Leu–Gly–Leu–Ala–OMe
Dec–*Orn(Imd)*–Gly–OMe
Dec–Try–*Orn(Imd)*–Leu–OMe
Dec–*Orn(Imd)*–Leu–Ala–OMe
Dec–Phe–*Orn(Imd)*–Leu–OMe
Dec–Phe–Phe–*Orn(Imd)*–OMe

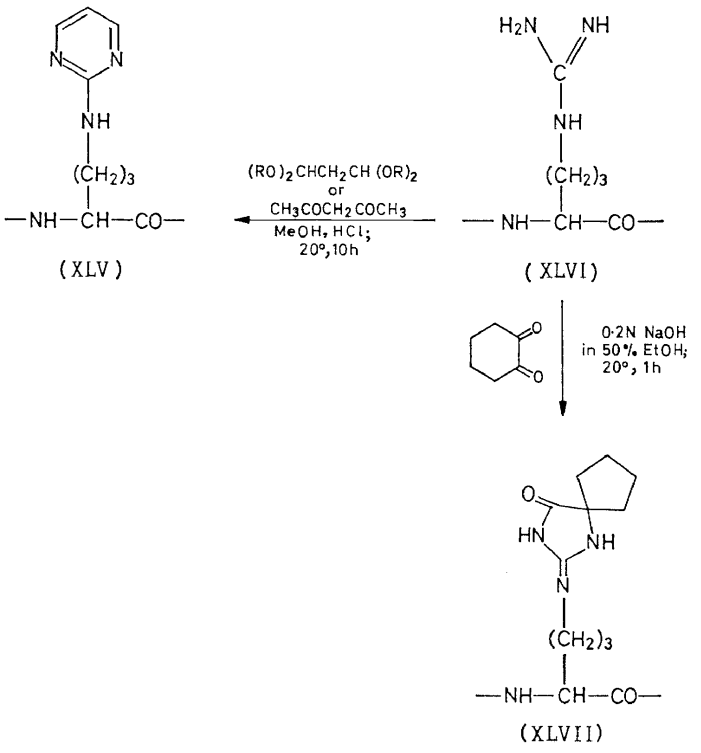




behaviour of the N^{ω}-substituent of the ornithine residue. Concerning these specificities mention should be made of the appearance of a characteristic group of peaks in the low mass region (up to 220 m.u.) permitting unequivocal identification of the N^{ω}-substituted ornithine residue in the peptide. Moreover, its amino group readily undergoes methylation, so that the molecular peaks and the peaks of the amino acid fragments are accompanied by satellites removed by +14 (28) m.u.

The basic law, and all the particular rules, that we found for the fragmentation of peptides have universal application, their validity having been confirmed by extensive data accumulated from studies of many differing series of synthetic peptides of known structure. This opens wide possibilities for the mass spectrometric analysis of the primary structure of polypeptides and proteins. However, in order to obtain maximum information it is essential that the method be adapted to each specific case.

A prerequisite for successful use of the method is preliminary knowledge of the amino acid composition of the peptide (readily obtainable with an amino acid analyzer). From this one can first of all make a reasonable conclusion as to the feasibility of a mass spectrometric study, because the composition will allow one to judge of the volatility of its derivatives; secondly, one can choose the type and conditions of preliminary treatment of the peptide (its modification, the introduction of protective groups); and

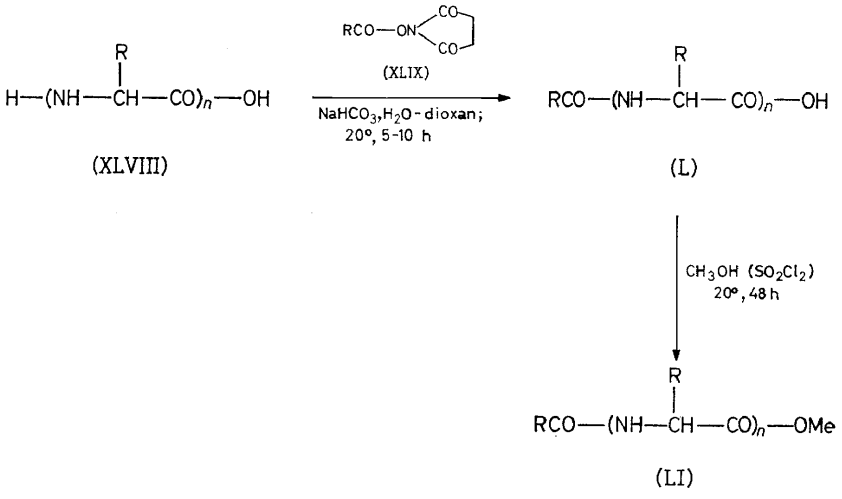


(XLV) (XLVI) (XLVII)

finally, one can select optimum conditions for obtaining the mass spectra. Not less important is the fact that, whatever the peptide, knowledge of its amino acid composition greatly facilitates reading of its mass spectrum, thus making for an unequivocal interpretation of the spectral pattern.

A second requisite for success is the existence of a micromethod for transforming the free peptide formed in the enzymatic or chemical cleavage of proteins into mass spectrometrically convenient derivatives. According to our method[3, 16–19], the peptides (XLVIII) are best acylated with *N*-hydroxysuccinimide esters of carboxylic acids (XLIX), which allows the reaction to be carried out in aqueous or aqueous-dioxane solution. Under such conditions acylation readily takes place of both the α-amino group of any amino acid and the ω-amino group of lysine and ornithine residues, while the guanidine group of arginine is unaffected. If the molecule contains arginine residues these are transformed into an ornithine derivative by one of the aforementioned methods (the choice of the method depending upon the amino acid composition of the peptide). For esterification (L→LI) methanol in the presence of catalytic amounts of sulphuryl chloride is recommended. A highly important point is that the products from all stages, including the final one, do not require purification, the mass spectra of the resultant samples usually being practically identical to those of the pure compounds.

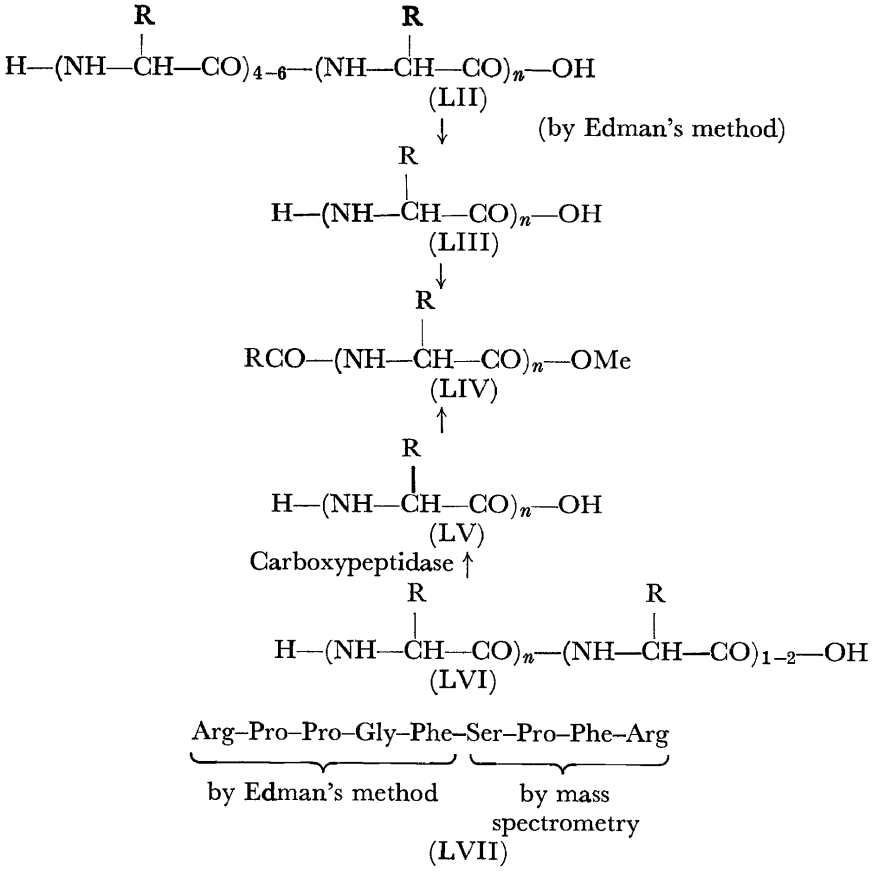
The limits of the mass spectrometric method are mainly determined by the volatility of the acylpeptide esters. This can be augmented by correct



choice of protective groups, primarily the *N*-acyl protection. At present we prefer to acylate the peptide by fatty acids of medium molecular weight (from valeric to decanoic). The volatility is greatly enhanced by the trifluoroacetyl grouping, but present methods for its introduction are too drastic. To increase the volatility of peptides, Lederer recently suggested exhaustive methylation of their amide bonds[43]. Sometimes this procedure gives good results, but it cannot be regarded as a general method. We have found that in the case of histidine, tryptophane, arginine and sulphur-containing peptides such treatment makes more difficult the decoding of the spectra [16, 18, 19].

The mass spectrometric method is applicable to peptides with up to 10–12 amino acid residues. We considerably extended its boundaries by combining it with the stepwise Edman cleavage of several (4–6) amino acid residues from the *N*-terminus[44]. It was shown that the remaining peptide (after conversion to the acylpeptide ester) can be directly subjected to mass spectrometry without purification (LII → LIII → LIV). One of the examples we have investigated is the peptide (LVII). Sometimes one can also carry out the preliminary cleavage of 1–2 amino acid residues from the *C*-terminus, for instance, lysine and arginine by carboxypeptidase B (LVI → LV → LIV). Since certain, for instance aromatic, amino acid residues, not only lower the volatility, but also considerably complicate the mass spectral picture, such chain shortening may also be undertaken in order to facilitate the analysis of the spectrum.

In most cases low resolution mass spectra are sufficient for amino acid sequence analyses, high resolution spectra being required only when doubt as to the assignment of a particular peak requires more precise determination of the mass of the corresponding ion; often a clue to the nature of this ion can be obtained from the metastable ions. In practically all cases the fragmentation laws we have established permit unambiguous reading of the mass spectra. These laws should also form the basis for setting up computer



programmes. Here one cannot rely solely on the amino acid type of fragmentation, or, still less, ignore the amino acid composition of the peptide, as has been often done in recent work[31–35]. It should be borne in mind that peaks corresponding to amino acid and aldimine fragments, can sometimes be absent, as we have observed in the mass spectrometry of peptides (LVIII,

Acylpeptide ester	*Absent Fragment*	
	Amino acyl fragment	*Aldimine fragment*
Dec–Met–Phe–His–Leu–OMe (LVIII)	$(Dec–Met–Phe–His)^{\oplus}$ m/e 570	$(Dec–Met–Phe–His-28)^{\oplus}$ m/e 542
Dec–Thr–Phe–His–Leu–OMe (LIX)	$(Dec–Thr–Phe–His)^{\oplus}$ m/e 540	$(Dec–Thr–Phe–His-28)^{\oplus}$ m/e 512
CH_2COOMe | Dec–Cys–Cys–Gly–Val–OMe | CH_2COOMe (LX)	$[Dec–Cys(CH_2COOMe)]^{\oplus}$ m/e 330	$[Dec–Cys(CH_2COOMe)-28]^{\oplus}$ m/e 302

LIX and LX); molecular ion peaks can also be lacking[18, 19]. Hence, excessive simplification here may lead to serious errors.

However, it is quite obvious to me that the mass spectrometric method can greatly simplify determination of the primary structure of peptides and proteins, reducing this problem to purely routine work. In my Institute the method is being successfully used to elucidate the structure of certain regions of the pepsin and aspartate transaminase molecules.

References

[1] P. Edman and G. Begg. *European J. Biochem.* **1**, 80 (1967).
[2] N. S. Wulfson, V. A. Puchkov, B. V. Rosinov, Yu. V. Denisov, V. N. Bochkarev, M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, and M. Yu Feigina. *Tetrahedron Letters* 2805 (1965).
[3] A. A. Kiryushkin, Yu. A. Ovchinnikov, M. M. Shemyakin, V. N. Bochkarev, B. V. Rosinov and N. S. Wulfson. *Tetrahedron Letters* 33 (1966).
[4] N. S. Wulfson, V. N. Bochkarev, B. V. Rosinov, M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, and A. I. Miroshnikov. *Tetrahedron Letters* 39 (1966).
[5] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, A. I. Miroshnikov, Yu. B. Alakhov, V. M. Lipkin, Yu. B. Shvetsov, N. S. Wulfson, B. V. Rosinov, V. N. Bochkarev, and V. M. Burikov. *Nature, Lond.*, **211**, 361 (1966).
[6] M. M. Shemyakin, Yu. A. Ovchinnikov, and A. A. Kiryushkin, *Peptides*, Proceedings of the Eighth European Peptide Symposium, p. 155. North-Holland Publishing Co., Amsterdam, 1967.
[7] Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, B. V. Rosinov, and M. M. Shemyakin. *Biokhimia* **32**, 427 (1967).
[8] M. M. Shemyakin, Yu. A. Ovchinnikov, E. I. Vinogradova, M. Yu. Feigina, A. A. Kiryushkin, N. A. Aldanova, Yu. B. Alakhov, V. M. Lipkin, and B. V. Rosinov. *Experientia* **23**, 428 (1967).
[9] A. A. Kiryushkin, A. I. Miroshnikov, Yu. A. Ovchinnikov, B. V. Rosinov, and M. M. Shemyakin. *Biochem. Biophys. Res. Commun.* **24**, 943 (1966).
[10] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, Yu. B. Alakhov, V. M. Lipkin, and B. V. Rosinov. *Zh. Obshch. Khim.* **38**, 798 (1968).
[11] A. A. Kiryushkin, V. A. Gorlenko, Ts. E. Agadzhanyan, B. V. Rosinov, Yu. A. Ovchinnikov, and M. M. Shemyakin. *Experientia*, **24**, 883 (1968).
[12] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, A. I. Miroshnikov, and B. V. Rosinov. *Zh. Obshch. Khim.*, in press.
[13] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, A. I. Miroshnikov, and B. V. Rosinov. *Zh. Obshch. Chim.*, in press.
[14] M. M. Shemyakin, Yu. A. Ovchinnikov, A. A. Kiryushkin, E. I. Vinogradova, M. Yu. Feigina, N. A. Aldanova, Yu. B. Alakhov, V. M. Lipkin, and B. V. Rosinov. *Tetrahedron*, in press.
[15] A. A. Kiryushkin, V. M. Burikov, and B. V. Rosinov. *Tetrahedron Letters* 2675 (1967).
[16] Yu. B. Alakhov, Thesis, Moscow, 1968, and unpublished data.
[17] A. I. Miroshnikov, Thesis, Moscow, 1968, and unpublished data.
[18] V. M. Lipkin, Thesis, Moscow, 1969, and unpublished data.
[19] V. A. Gorlenko, Thesis, Moscow, 1969, and unpublished data.
[20] G. A. Junk and H. J. Svec. *Anal. Biochemistry* **6**, 199 (1963).
[21] H. J. Svec and G. A. Junk. *J. Am. Chem. Soc.* **86**, 2278 (1964).
[22] K. Biemann and W. Vetter. *Biochem. Biophys. Res. Commun.* **3**, 578 (1960).
[23] K. Biemann, F. Gapp, and J. Seibl. *J. Am. Chem. Soc.* **81**, 2274 (1959).
[24] K. Biemann. *Chimia* **14**, 393 (1960).
[25] F. Weygand, R. Geiger, and W. Swodenk. *Angew. Chem.* **68**, 307 (1956).
[26] F. Weygand, A. Prox, W. König, and H. H. Fessel. *Angew. Chem.* **75**, 724 (1963).
[27] M. Barber, P. Jolles, E. Vilkas, and E. Lederer. *Biochem. Biophys. Res. Commun.* **18**, 469 (1965).
[28] M. Barber, W. A. Wolstenholme, M. Guinand, G. Michel, B. C. Das, and E. Lederer. *Tetrahedron Letters* 1331 (1965).
[29] G. Laneelle, J. Asselineau, W. A. Wolstenholme, and E. Lederer, *Bull. Soc. Chim., France* 2133 (1965).
[30] F. Weygand, A Prox, H. H. Fessel, and K. K. Sun. *Z. Naturforschung* **20b**, 1169 (1965).
[31] K. Biemann, C. Cone, and B. R. Webster. *J. Am. Chem. Soc.* **88**, 2597 (1966).

[32] K. Biemann, C. Cone, B. R. Webster, and G. P. Arsenault. *J. Am. Chem. Soc.* **88**, 5598 (1966).
[33] M. Senn and F. W. McLafferty. *Biochem. Biophys. Res. Commun.* **23**, 381 (1966).
[34] M. Senn, R. Venkataraghavan, and F. W. McLafferty. *J. Am. Chem. Soc.* **88**, 5593 (1966).
[35] M. Barber, P. Powers, M. J. Wallington, and W. A. Wolstenholme. *Nature, Lond.* **212**, 784 (1966).
[36] E. Bricas, J. van Heijenoort, M. Barber, W. A. Wolstenholme, B. C. Das, and E. Lederer. *Biochemistry* **4**, 2254 (1965).
[37] J. van Heijenoort, E. Bricas, B. C. Das, E. Lederer, and W. A. Wolstenholme. *Tetrahedron* **23**, 3403 (1967).
[38] M. Guinand, M. J. Vacheron, G. Michel, B. C. Das, and E. Lederer. *Tetrahedron, Suppl.* **7**, 271 (1966).
[39] M. Guinand, G. Michel, B. C. Das, and E. Lederer. *Vietnam. Chim. Acta* 37 (1966).
[40] A. Prox and F. Weygand, *Peptides*, Proceedings of the Eighth European Peptide Symposium, p. 159. North-Holland Publishing Co., Amsterdam, 1967.
[41] H. F. Grützmacher and K. Heyns. *Ann. Chem.* **698**, 24 (1966).
[42] D. W. Thomas and K. Biemann. *J. Am. Chem. Soc.* **87**, 5447 (1965).
[43] B. C. Das, S. D. Gero, and E. Lederer. *Biochem. Biophys. Res. Commun.* **29**, 211 (1967).
[44] N. A. Aldanova, E. I. Vinogradova, S. A. Kazaryan, and M. M. Shemyakin. *Zh. Obshch. Khim.*, in press.