Available online at www.sciencedirect.com
ScienceDirect
International Journal of Mass Spectrometry 259 (2007) 1–7
International Journal of Mass Spectrometry
www.elsevier.com/locate/ijms

Laying the groundwork for proteomics Mass spectrometry from 1958 to 1988

Klaus Biemann *

Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

Received 2 August 2006; accepted 2 August 2006
Available online 18 September 2006

Dedicated to Professor Donald F. Hunt at the occasion of his 65th birthday; Don received his training in mass spectrometry in the author's laboratory as a postdoctoral fellow 1967–1968.

Abstract

The development of mass spectrometric methodologies for the sequencing of peptides and proteins are recounted. Early strategies for the determination of very large proteins based on a combination of nucleotide sequencing and mass spectrometric amino acid sequencing are described and their historical significance to the new field of proteomics is outlined.
Published by Elsevier B.V.

Keywords: Mass spectrometry; Proteomics; Peptide sequencing; Protein sequencing; Aminoacyl-tRNA synthetases

1. Introduction

Today a protein can be identified, including its post-translational modifications, from a spot on a 2D gel. This became possible in the late 1990s due to the confluence of developments in DNA sequencing, computer technology and mass spectrometry. But it had been only in 1953 that Sanger completed the first determination of the primary structure of a protein, insulin [1]. It was in the same year that Watson and Crick proposed the double helix structure for DNA and the genetic code had not yet been deciphered. In the 1950s chemists used mass spectrometry mainly for quantitative analysis in the petroleum industry and for the identification of relatively small and volatile organic compounds by matching their spectra with those of known samples.

2. Peptide sequencing by mass spectrometry

In 1958 we began to develop a method for peptide sequencing by mass spectrometry (MS) suitable for the determination of the primary structure of proteins. The major obstacles were the non-volatility of peptides due to their zwitter-ionic character, and the fact that the large number of possible sequences (400 dipeptides, 8000 tripeptides, . . .) made identification by matching impossible. Therefore, we had to convert the peptide to a more volatile derivative that retains sequence information and produces a mass spectrum from which the sequence can be deduced directly. This was achieved by converting the carboxylate anion to a methyl ester and the ammonium ion to an amido group. These reactions eliminated the zwitter ion. Reduction of the carbonyl group of the peptide bonds and methyl esters by lithium aluminum deuteride resulted in an amino alcohol [2](Fig. 1). The newly formed –NH-CHR-CD_2-NH– group not only retains the sequence information but also cleaves preferentially due to the stabilizing effect of the adjacent imino groups. The result is a very simple mass spectrum (Fig. 2) from which the sequence of the parent peptide can be easily deduced directly, without the availability of an authentic sample [3]. The incorporation of deuterium was required to differentiate the side chains of serine and threonine from those formed by the reduction of aspartic acid and glutamic acid. The only ambiguities among the 20 natural protein amino acids were leucine and isoleucine due to their isomeric side chains, and the fact that asparagine and glutamine were partially converted to aspartic and glutamic methyl esters by methanolysis before reduction.

The ultimate aim of this strategy, the sequencing of a protein, required its applicability to the complex mixture of small

* Tel.: +1 617 253 7221.
E-mail address: kbiemann@mit.edu.

1387-3806/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.ijms.2006.08.002

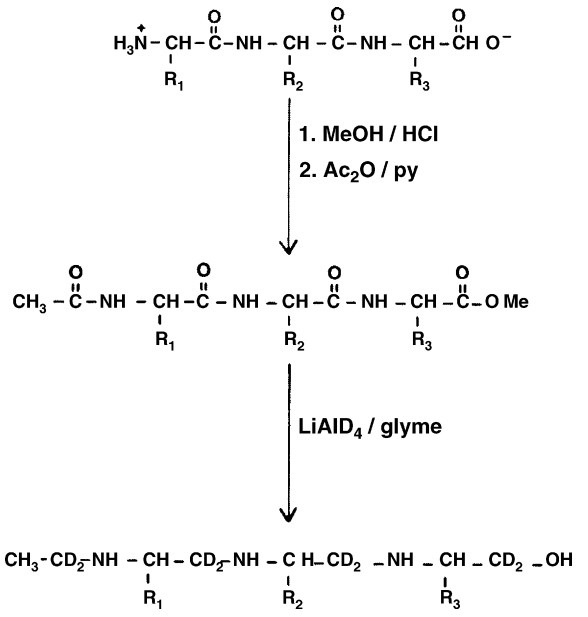

Fig. 1. Reaction scheme for the reduction of peptides to polyamino alcohols.

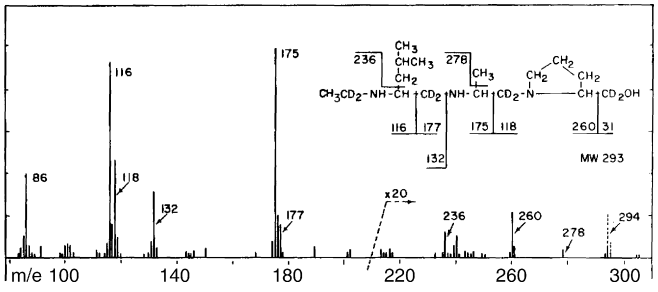

Fig. 2. Mass spectrum of the derivative of a tripeptide (Leu-Ala-Pro) (Reprinted from Ref. [3] with permission from Elsevier).

peptides produced by partial acid or enzymatic hydrolysis. It required efficient separation on a micro-scale after the derivatization of that mixture. The relatively high volatility of the polyamino alcohols allowed their separation by gas chromatography (GC) (Fig. 3), a method developed a few years earlier by James and Martin in the UK. Each fraction could be collected manually and placed into the inlet system of the mass spectrometer (CEC model 21–103C) to produce the spectrum. Thus, the feasibility of this new approach to peptide, and ultimately protein sequencing, had been demonstrated. However,

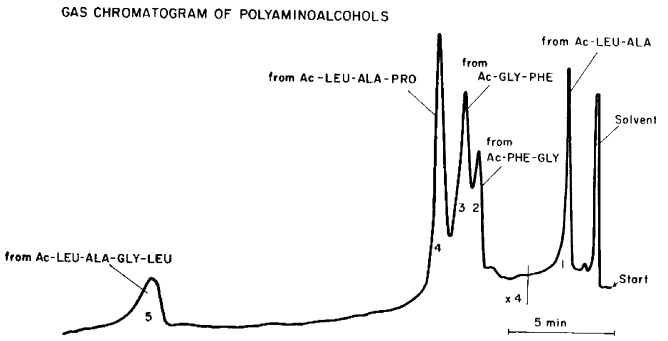

Fig. 3. Gas chromatogram of the polyamino alcohols derived from a mixture of di-, tri- and tetrapeptides (Reprinted from Ref. [3] with permission from Elsevier).

further improvements were necessary to apply it successfully to the very complex mixtures of small peptides consisting of any of the 20 protein amino acids. Trimethylsilylation of the free hydroxyl groups and trifluoroacetylation not only extended the gas chromatographic separability to larger and more polar polyamino alcohols, but also further improved the specificity of the mass spectra (Fig. 4) [4,5]. The development of the direct interface of the GC with the MS [6] eliminated the tedious and time consuming collection of individual fractions; novel computer algorithms facilitated the interpretation of the resulting vast amount of data [5,7].

Another type of peptide derivative suitable for sequencing, acetyl-N,O-permethylated methyl esters, were developed by Morris and Williams at Cambridge (UK). In these derivatives the hydrogen bonding NH-groups had been replaced by $-NCH_3$ but the carbonyl groups were left intact. Therefore, the volatility of these compounds was too low to be amenable to gas chromatography and thus were sublimed into the ion source, which provided some degree of fractionation [8]. These "permethylated" peptide derivatives were also extensively used by Hunt et al. [9].

3. Sequencing proteins

When the final version of our GCMS methodology [4,5] was applied to the partial acid hydrolyzate of subunit 1 of monellin, a small, sweet-tasting protein of unknown structure, 55 di- to hexapeptides could be identified in a single chromatogram (Fig. 5). Because of the extensive overlap, these could be assembled to one unique sequence (Fig. 6) [10].

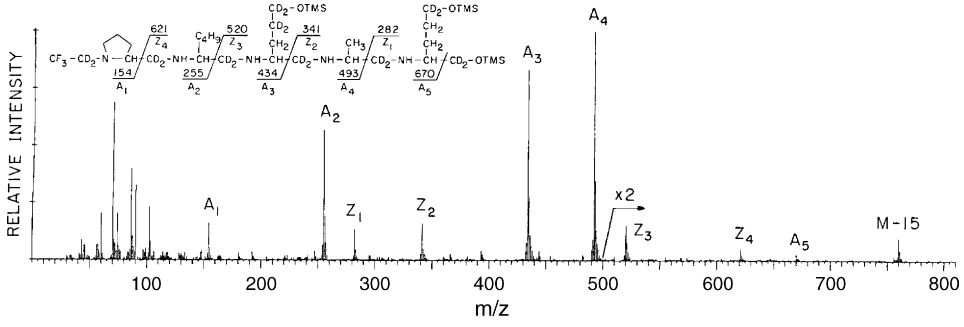

Fig. 4. Mass spectrum of the derivative obtained from a pentapeptide after trifluoroacetylation and trimethylsilylation of the corresponding amino alcohol.

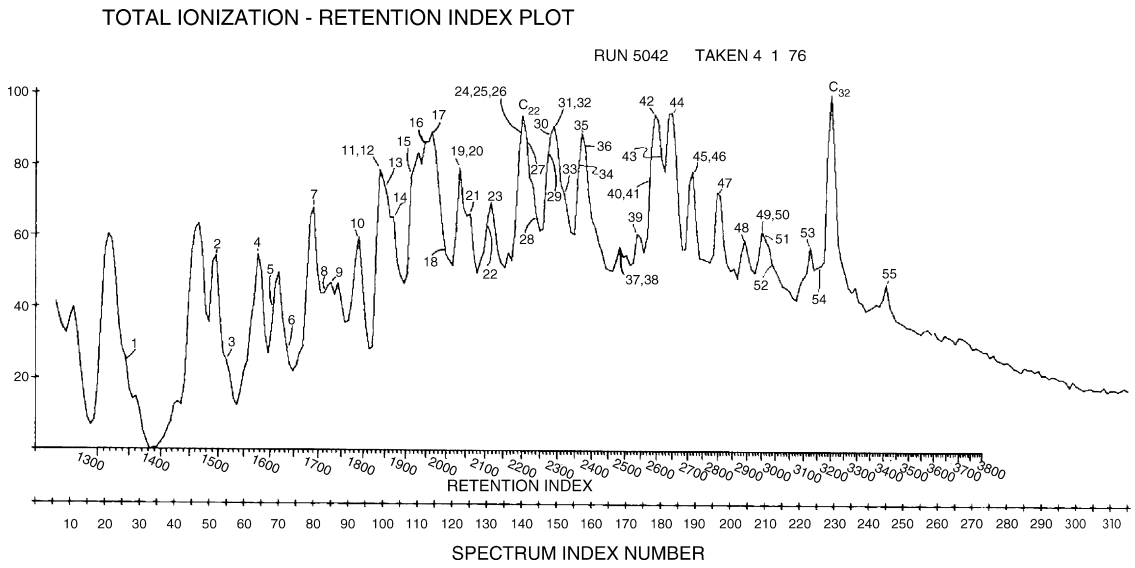

Fig. 5. Total ionization plot (gas chromatogram) of the derivatized partial acid hydrolyzate of subunit 1 of monellin. Numbers indicate elution of peptides identified by their mass spectra. C-22 and C-32 refer to the hydrocarbons added as retention time standards (Reprinted from Ref. [10] with permission from Elsevier).

In the meantime Edman had developed the stepwise degradation and sequencing of peptides and proteins [11]. This methodology became widely used after its automation [12] and commercialization. However, it had some limitations, such as N-terminally blocked and cyclic peptides, highly hydrophobic peptides, peptides containing certain chemically or post-translationally modified amino acids, etc. In such cases our mass spectrometric method provided solutions: the first instance of a N-myristylated peptide [13]; the above mentioned monellin, which has a very hydrophobic C-terminus that caused "wash-out" in the Edman sequencer [14]; peptides containing γ-carboxy-glutamic acid [15], etc.

3.1. Hydrophobic proteins

Because of the reciprocal complementarities of these two very different methods, they were often applied together during the next few years. A good example is the determination of the primary structure of bacteriorhodopsin, the light-sensitive protein from *Halobacterium halobium*. This protein, which loops through the cell membrane seven times, is so hydrophobic that it is insoluble in the aqueous buffers used for enzymatic digestions. In suspension chymotrypsin cleaved at a single peptide bond, producing two polypeptides, C-1 and C-2, which could be separated by gel permeation chromatography. Both were still only soluble in 70% formic acid, the preferred solvent for cyanogen bromide. This reagent cleaves at the C-terminal side of methionine, producing peptides ending in homoserine. For C-2 these could be separated into five fractions, labeled CNBr-1 to 5a,b according to their elution from a reverse phase column, therefore their molecular size. These were sequenced by the Edman degradation as well as by GCMS. As is apparent from Fig. 7, the latter data revealed the sequence of all six peptides, with the exception of a few gaps and missing overlaps in CNBr-1 and CNBr-2. Fortunately, these were covered by the Edman data, which in turn could not reach the hydrophobic C-terminal region of any of these peptides; 5a and 5b were too short to be amenable to the Edman method. In order to assemble these six peptides in the correct order we had to find overlapping peptides containing methionine. This was accomplished by searching [7] the GCMS data set obtained from a partial acid hydrolyzate of intact C-2 for characteristic fragment ions predicted for all potential X-Met-Y sequences. The same strategy revealed the sequence of C-1. Since the N-terminus of CNBr-2 was pyroglutamine the sequence must be C-2–C-1 as shown in Fig. 8, consisting of a linear string of 248 amino acids [16].

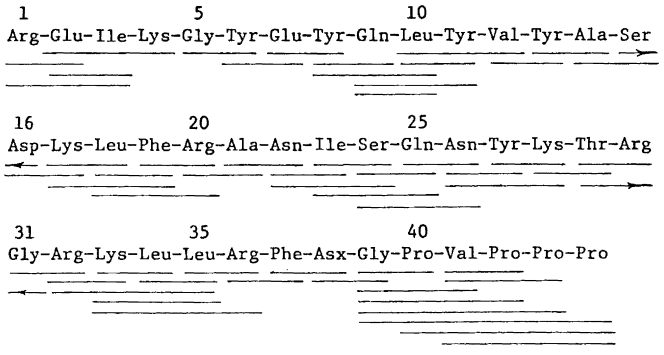

Fig. 6. Amino acid sequence of subunit 1 of monellin. Underlining indicates peptides identified by GCMS (Reprinted from Ref. [10] with permission from Elsevier).

4. The advent of DNA sequencing

In 1978 Fowler and Zabin published a series of six papers describing an 8 year effort at UCLA to determine the 1021 amino acids long sequence of β-galactosidase from *E. coli* using the Edman method. By the time the last [17] of these papers appeared, Gilbert's laboratory at Harvard had developed his DNA sequencing method [18] and had applied it to the gene coding for that protein. Their data confirmed the first 145 amino acids of β-galactosidase. With this development a new vista

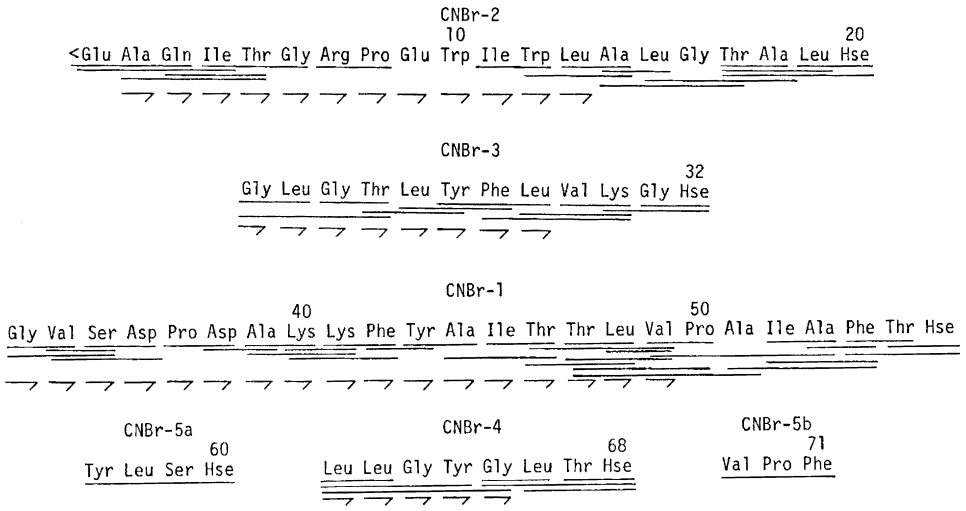

Fig. 7. Amino acid sequences of the peptides obtained by cyanogen bromide cleavage of the N-terminal segment of bacteriorhodopsin. Underlining: GCMS data; half-arrows: Edman data (adapted from Ref. [16a]).

for protein structure determination opened and DNA sequencing became an attractive alternative. However, it was still in its infancy and prone to errors: the four lanes on long one-dimensional electrophoresis gel strips had to be read manually and repeatedly, because only a short range (about 50 nucleotides) was sufficiently resolved and sharp enough in each experiment. These problems were exacerbated by the fact that a single missing or erroneously inserted nucleotide causes a phase shift and completely changes the derived amino acid sequence. Furthermore, even a correct nucleotide sequence represents three "reading frames", each one leading to a different protein sequence and the correct one had to be identified. This could be done by identification of the N- and C-terminal amino acids of the protein, but two compensating errors in the interior of the DNA strand can still result in a long stretch of incorrect amino acid sequence. In the DNA sequence, the start could be identified by the initiating ATG codon for methionine and the termination by one of the "stop" codons (ATG, TAA or TAG), but this was subject to the same phase-shift problem. In addition, the "coding strand" has to be identified, because the complementary "non-coding strand" gives rise to yet another three different amino acid sequences. All these problems become more and more severe the longer the protein chain and, therefore, the coding DNA strand is.

4.1. Combination of nucleotide sequencing with mass spectrometry

Paul Schimmel, at that time a professor of biology at MIT, was interested in the structure and mechanism of action of aminoacyl-tRNA synthetases. These are large (up to 1000 amino acids long), multifunctional enzymes which recognize and attach a specific amino acid to the corresponding transfer-

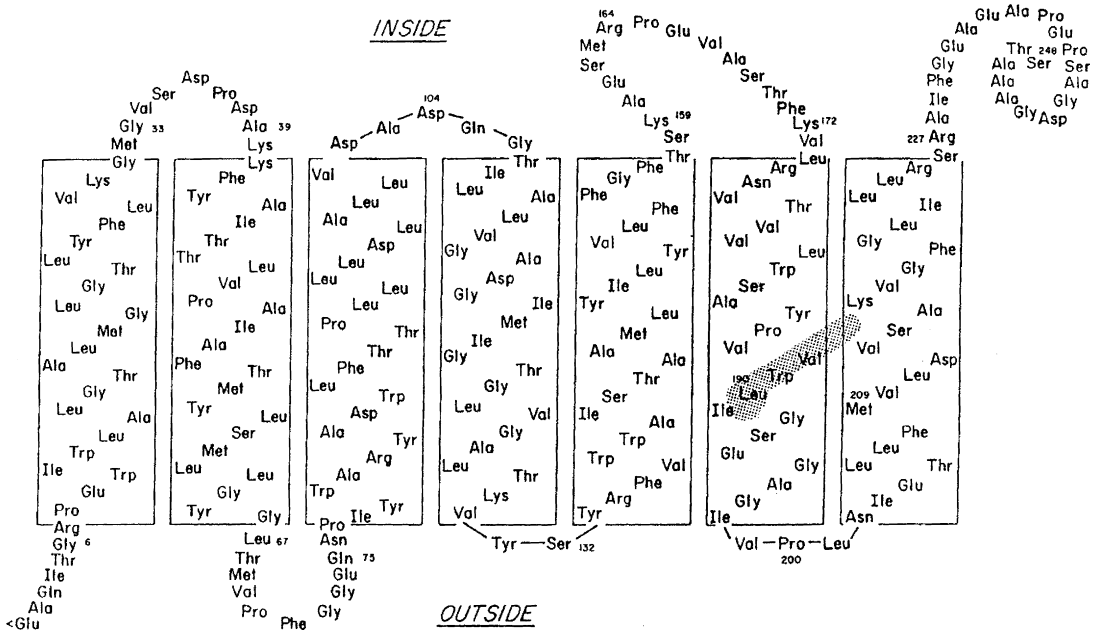

Fig. 8. Primary structure of bacteriorhodopsin from *H. halobium*. Boxes represent the regions where the protein loops though the cell.

RNA for the elongation of a growing polypeptide chain. Rather than embarking on the tedious Edman degradation of such large proteins, he decided to determine the about 3000 nucleotide long sequence of the structural gene coding for alanyl-tRNA synthetase (ARS) from *E. coli*. When discussing the difficulties mentioned above, I realized that all of these could easily and efficiently be overcome by our GCMS method. The identification (by sequence) of a relatively small number of peptides scattered over the entire protein would reveal and help to correct all these potential errors. Matching these peptides to the three protein sequences corresponding to the three reading frames not only allows bracketing the region where an error occurred, but also whether it is a deletion or an insertion of a nucleotide. Re-examination of that particular sequencing gel identifies the error and eliminates the unnecessary proofreading of all the others. As a consequence, we developed a strategy for multiple phase checks by mass spectrometry, using partial enzymatic digests of the corresponding protein [19].

In this collaboration Schimmel's group began to sequence the gene using the Maxam–Gilbert method, while we digested the about 400 amino acids long N-terminal segment (termed T-1) with thermolysin and pepsin, respectively. These enzymes were chosen to minimize the production of free amino acids and dipeptides, which would be useless. The resulting very complex digests were then derivatized and the GCMS data set processed using computer programs written for this purpose [5,7,20]. The sequences of tri- to pentapeptides so identified were then fed into our DEC PDP-11/45 computer along with the gradually accumulating nucleotide sequences to match them to the three reading frames.

The results for the first 89 codons are schematically depicted in Fig. 9. The N-terminus had been identified by a few Edman steps, thus defining reading frame 1. Three overlapping tri- and tetrapeptides matched amino acids 11 through 15 (nucleotides 31–45), but others fit amino acids 48–51 (nucleotides 142–153) and 81–84 (nucleotides 241–252) only in reading frame 3. This indicated that one nucleotide had been missed in the region between 46 and 141, but the remainder of the sequence up to nucleotide 252 was correct. Re-inspection of the gels covering this stretch of less than hundred nucleotides revealed and corrected the error. This process was continued until the entire sequence of T-1 and then also of the C-terminal section (T-2) had been defined, resulting in the complete primary structure of ARS, which turned out to be 875 amino acids long. The work

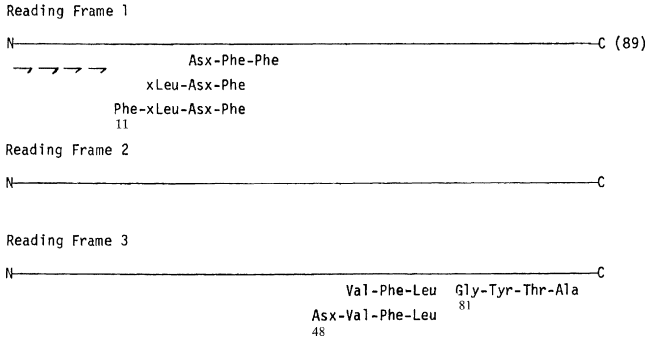

Fig. 9. Detection of an error in the DNA sequence (see text).

was published in 1981 in *Science* [21], which used the amino acid and DNA sequences to illustrate the cover of that issue. The significance of this complementary strategy was immediately recognized by others working on the determination of the structures of these and other large proteins. Even before the sequence of ARS was completed, we began a collaboration with Söll at Yale on glutaminyl-tRNA synthetase from *E. coli* [22].

A remarkable turning point in the application of mass spectrometry to peptide sequencing was reached in 1981 with Barber's (Manchester, UK) invention of "fast atom bombardment" (FAB) ionization. This novel technique made it possible to ionize an intact, underivatized peptide, such as Met-Lys-bradykinin (mw 1318) [23]. As a so-called "soft ionization" technique, it produced very stable $(M+H)^+$ ions, which have little tendency to fragment. This had the disadvantage of a lack of sequence specific fragment ions (unless a very large or very pure peptide sample was used), but the great advantage that it was now possible to measure the molecular weights of large peptides directly, even in mixtures. Furthermore, it was easy to use this method by simply fitting a commercially available argon atom gun to the ion source of almost any existing mass spectrometer. Now we could use trypsin to specifically cleave a protein at the C-terminal side of all arginines and lysines and produce a mixture of peptides of the size just right for FAB–MS. It was only necessary to separate the digest by liquid chromatography into a few simpler mixtures to obtain the molecular weight of most or all the peptides produced from the protein by cleavage with trypsin (Fig. 10).

This development, which almost over night rendered obsolete the derivatization chemistry we had developed and so successfully applied over more than two decades, greatly expanded the use of mass spectrometric peptide and protein sequencing. At that time, we were working on the sequencing of Gln-tRNA synthetase as mentioned above, but completed it using FAB–MS [22]. The sequences of Gly-, Met-, His- and Glu-tRNA synthetases were determined by the same collaborative DNA sequencing (some using the Sanger method [24])/FAB–MS strategy (Table 1). The basic difference between this and the earlier GCMS approach is demonstrated in Fig. 11. Rather than matching many short sequences, we calculated the molecular weights of all tryptic peptides predicted for the amino acid sequences corresponding to each of the three reading frames,

Table 1
Sequences of aminoacyl tRNA synthetases deduced by a combination of DNA sequencing and mass spectrometry

	Number of amino acids	References
Ala-tRNA synthetase[a,c]	875	[21]
Gln-tRNA synthetase[a–c]	550	[22]
Gly-tRNA synthetase[b,c]	990	[25]
Met-tRNA synthetase[b,d]	751	[26]
His-tRNA synthetase[b,c]	324	[27]
Glu-tRNA synthetase[b,c]	471	[28]

[a] By GCMS.
[b] By FAB–MS.
[c] From *E. coli*.
[d] From yeast.

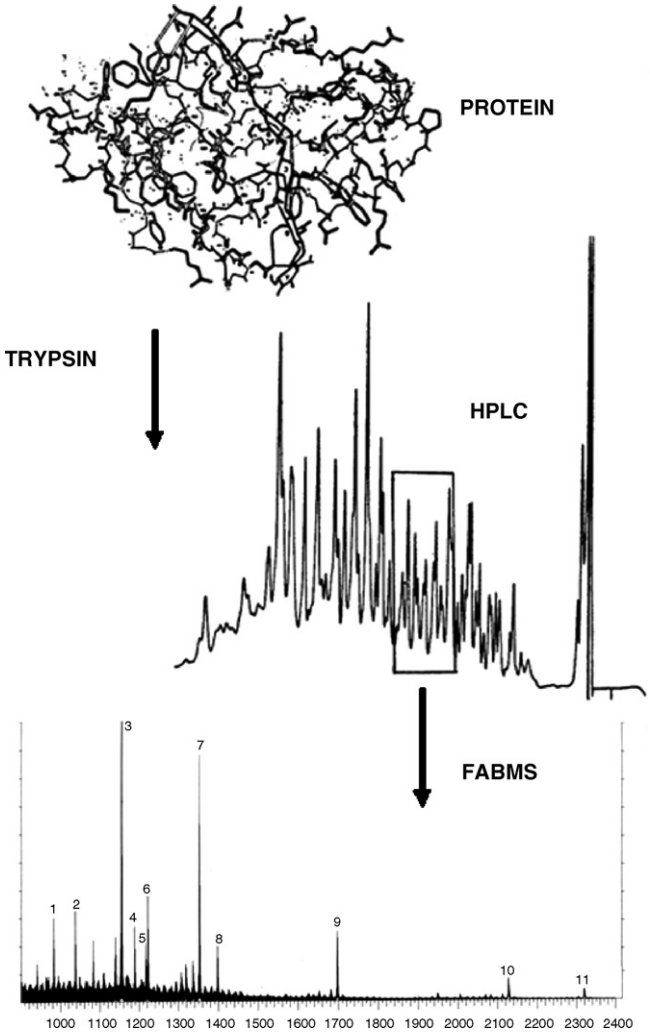

Fig. 10. Schematic illustration of the determination of the molecular weights of the components of a tryptic digest. The mass spectrum (bottom) indicates the molecular weights of 11 peptides in the range from 900 to 2400 in the HPLC fraction (box) collected.

and matched them with those experimentally determined from the tryptic digest of the protein.

In the example shown (Fig. 11), the values for two peptides matched those predicted for tryptic peptides from reading frame 1, but others matched reading frame 3. This indicated that one nucleotide had been missed in the region of the 41 in between. Insertion of one of the four nucleosides at each consecutive position would create 168 possibilities to be tested. However, the most frequent omission of a nucleotide occurs when it is preceded by the same one, which may cause the two consecutive bands in the same lane of the gel to be too close to be visually resolved. Thus, "doubling up" each one of the 41 nucleotides one at a time generated only 42 new potentially correct sequences from which 8 would produce one or more tryptic peptides. For only two of these (7 and 17) did the predicted molecular weight correspond to one (1444) actually present in the tryptic digest of the protein. These could be distinguished by three steps of "subtractive" Edman degradations, which involve measuring the molecular weight changes after each step. The N-terminus turned out to be Leu-Ala-Asp, indicating that sequence seven is the correct one and the error was due to the omission of an additional guanidylic acid between G and A [26].

A different problem arose in the case of Met-tRNA synthetase from yeast. When sequencing the gene coding for this large protein, three Met codons were found close together, but an attempt to identify the initiating codon by determination of the N-terminus of the protein using the Edman method failed. A FAB–MS experiment on a tryptic digest revealed the presence of three peptides, the $[M+H]^+$ ions of which fit only between the first and the second Met, indicating that the first one initiates transcription but is then removed and the serine that follows acetylated, a common feature of post-translational processes [26].

Such transformations of the nascent polypeptide chain to the various biologically active forms of a protein cannot be deduced or predicted from the nucleotide sequence of the gene. To pinpoint and identify these modifications, sequence specific fragmentation of the tryptic (or other enzymatically or chemically produced) peptides was necessary. This was achieved by tandem mass spectrometry (MS/MS) which involves the collision of an $(M+H)^+$ ion produced in the first mass spectrometer with a noble gas at low (a few eV) kinetic energy in a triple quadrupole MS [29] or at high (kV) energy in a four-sector magnetic MS [30]. These spectra, particularly those involving high energy collisions, provided clear, complete sequence information, including the differentiation of leucine and isoleucine. This approach not only revealed the type and position of post-translational modifications [31] but also permitted the sequencing of proteins entirely by FAB-MS/MS, as first demonstrated on a number of thioredoxins [32] and later glutaredoxins.

5. Mass spectrometry of proteins "going public"

The work described above had made the biochemistry community acutely aware of the value and unique significance of mass spectrometry to the field of gene/protein structure correlation. The potential of matching the molecular weights of tryptic (or other specific cleavage) peptides to known or predicted protein sequences was quickly recognized. In 1989, at the third Symposium of The Protein Society, Henzel from Stults' group at Genentech presented an algorithm for using this principle to identify proteins by matching such mass spectral data to then already available digital databases. Four years later four papers, three of them from other laboratories, demonstrated the utility of this approach and elaborated on it [33–36].

The strategies described above soon became the basis of today's proteomics, a new field further stimulated by two major advances in instrumentation. In 1988 matrix-assisted laser desorption ionization (MALDI) was developed by Hillenkamp at the University of Münster (Germany) [37] and soon thereafter electrospray ionization (ESI) by Fenn at Yale [38]. These techniques, particularly ESI, which lent itself well to interfacing with liquid chromatography, essentially replaced FAB–MS. Finally, "nanospray" ESI developed by Wilm and Mann in Denmark made it possible to obtain the molecular weights of the components of tryptic or other enzyme digests of a protein extracted from a spot on a two-dimensional electrophoresis gel [39]. Hunt

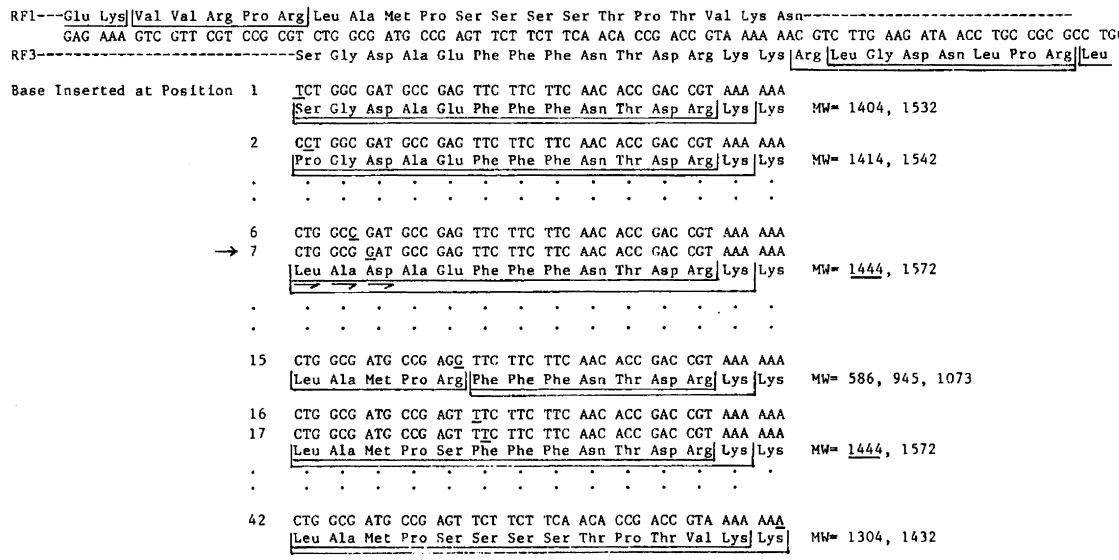

Fig. 11. Detection and correction of an error in a DNA sequence by FAB–MS. Underlining: tryptic peptides detected and predicted; half-arrows: subtractive Edman data. For details see text (Reprinted from Ref. [26] with permission from NAS).

and co-workers achieved automated high-throughput analysis of enzymatic digests of mixtures of proteins [40]. In Yates' laboratory, the database matching algorithms mentioned above were expanded to include the use of predicted MS/MS spectra [41], and later to analyze entire protein complexes by multi-dimensional liquid chromatography and ESI–MS/MS methodology using the yeast genome sequence [42]. The completion of the sequencing of the entire human genome in the year 2000 finally eliminated the need to painstakingly determine the amino acid sequence of each of the about 20,000 proteins coded by the nucleotide sequence. But to follow the cascades of post-translational conversions to the biologically active structures, mass spectrometry is still the principal methodology available for this important, but daunting task [43].

References

[1] F. Sanger, A. Thompson, Biochem. J. 53 (1953) 353.
[2] K. Biemann, F. Gapp, J. Seibl, J. Am. Chem. Soc. 81 (1959) 2274.
[3] K. Biemann, W. Vetter, Biochem. Biophys. Res. Commun. 3 (1960) 578.
[4] J.A. Kelley, H. Nau, H.-J. Forster, K. Biemann, Biomed. Mass Spectrom. 2 (1975) 313, 326.
[5] H. Nau, K. Biemann, Anal. Biochem. 73 (1976) 139, 154.
[6] J.T. Watson, K. Biemann, Anal. Chem. 36 (1964) 1135.
[7] R.A. Hites, K. Biemann, Anal. Chem. 42 (1970) 855.
[8] H.R. Morris, D.H. Williams, R.P. Ambler, Biochem. J. 125 (1971) 189.
[9] D.F. Hunt, A.M. Buko, J.M. Ballard, J. Shabanowitz, A.B. Giordani, Biomed. Mass Spectrom. 8 (1981) 387.
[10] G. Hudson, K. Biemann, Biochem. Biophys. Res. Commun. 71 (1976) 212.
[11] P. Edman, Acta Chem. Scand. 10 (1956) 761.
[12] P. Edman, G. Begg, Eur. J. Biochem. 1 (1967) 80.
[13] S.A. Carr, K. Biemann, S. Shoji, D.C. Parmelee, K. Titani, Proc. Natl. Acad. Sci. U.S.A. 79 (1982) 2362.
[14] Z. Bohak, S.-L. Li, Biochem. Biophys. Acta 427 (1976) 153.
[15] S.A. Carr, P.V. Hauschka, K. Biemann, J. Biol. Chem. 256 (1981) 9944.
[16] (a) G.E. Gerber, R.J. Anderegg, W.C. Herlihy, C.P. Gray, K. Biemann, H.G. Khprana, Proc. Natl. Acad. U.S.A. 76 (1979) 227;
(b) H.G. Khorana, G.E. Gerber, W.C. Herlihy, C. Gray, R.J. Anderegg, K. Nihei, K. Biemann, Proc. Natl. Acad. U.S.A. 76 (1979) 5046.
[17] A.V. Fowler, I. Zabin, J. Biol. Chem. 253 (1978) 5521.
[18] A.M. Maxam, W. Gilbert, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 560.
[19] W.C. Herlihy, N.J. Royal, K. Biemann, S.D. Putney, P.R. Schimmel, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 6531.
[20] W.C. Herlihy, K. Biemann, Biomed. Mass Spectrom. 8 (1981) 70.
[21] S.D. Putney, N.J. Royal, H. Neuman de Vegvar, W.C. Herlihy, K. Biemann, P.R. Schimmel, Science 213 (1981) 1497.
[22] P. Hoben, N. Royal, A. Cheung, F. Yamao, K. Biemann, D. Söll, J. Biol. Chem. 257 (1982) 11644.
[23] M. Barber, R.S. Bordoli, R.D. Sedwick, A.N. Tyler, J. Chem. Soc. Chem. Commun. (1981) 325.
[24] F. Sanger, S. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 5463.
[25] T.R. Webster, B.W. Gibson, T. Keng, K. Biemann, P.R. Scimmel, J. Biol. Chem. 258 (1983) 10637.
[26] B.W. Gibson, K. Biemann, Proc. Natl. Acad. Sci. U.S.A. 81 (1984) 1956.
[27] R. Freedman, B.W. Gibson, G. Donovan, K. Biemann, S. Eisenbeis, J. Parker, P.R. Schimmel, J. Biol. Chem. 260 (1985) 10063.
[28] R. Breton, H.H. Sanfacon, I. Papayannopoulos, K. Biemann, J. Lapointe, J. Biol. Chem. 261 (1986) 10610.
[29] D.F. Hunt, J.R. Yates III, J. Shabanowitz, S. Winston, C.R. Hauer, Proc. Natl. Acad. Sci. U.S.A. 83 (1986) 6233.
[30] H.A. Scoble, S.A. Martin, K. Biemann, Biochem. J. 245 (1987) 621.
[31] K. Biemann, H.A. Scoble, Science 237 (1987) 992.
[32] R.S. Johnson, W.R. Mathews, K. Biemann, S. Hopper, J. Biol. Chem. 263 (1988) 9589, and refs. [19–21] therein.
[33] W. Henzel, T. Billeci, J. Stults, S. Wond, C. Grinley, C. Watanabe, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 5011.
[34] J.R. Yates III, P.R. Griffin, S. Speicher, T. Hunkapiller, Anal. Biochem. 214 (1993) 397.
[35] P. James, M. Quadroni, E. Carafoli, G. Gonnet, Biochem. Biophys. Res. Commun. 195 (1993) 58.
[36] D. Pappin, P. Hojrup, A. Bleasby, Curr. Biol. 3 (1993) 327.
[37] M. Karas, F. Hillenkamp, Anal. Chem. 60 (1988) 299.
[38] J.B. Fenn, M. Mann, C.K. Meng, S.F. Wong, C.M. Whitehouse, Science 246 (1989) 64.
[39] M. Wilm, M. Mann, Anal. Chem. 68 (1996) 1.
[40] S.E. Martin, J. Shabanowitz, D.F. Hunt, J.A. Marto, Anal. Chem. 72 (2000) 4266.
[41] J. Eng, A.L. McCormack, J.R. Yates III, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[42] A.J. Link, J. Eng, D.M. Schieltz, E. Carmack, G.J. Mieze, D.R. Morris, B.M. Garvik, J.R. Yates III, Nat. Biotech. 17 (1999) 676.
[43] D.F. Hunt, J. Proteome Res. 1 (2002) 15.