Chapter 3

THE ANALYSIS OF SYNTHETIC PEPTIDES AND PROTEINS BY ^{252}CF- PLASMA DESORPTION MASS SPECTROMETRY*

Brian T. Chait

TABLE OF CONTENTS

* The present paper has appeared in large part previously in Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, Chichester, 1988, 21. Reproduced by permission of John Wiley & Sons, Ltd.

I. INTRODUCTION

There is a large and rapidly growing need among members of the biological community for high-purity synthetic peptides and small proteins. The most widely used methods for producing these materials are based on the stepwise solid-phase synthetic procedure devised by Merrifield[1] (for reviews see References 2 to 4). As the size and complexity of the target peptide increase, the opportunity for synthetic errors, unwanted modifications, and cumulative effects arising from incomplete reactions also increases. It is therefore imperative to have available effective means for rapidly verifying the correctness of the covalent structure of these complex materials, establishing their purity, and detecting and identifying undesired peptide byproducts. The numerous methods devised for these purposes have been reviewed.[2-4] The most useful have involved subjecting the products to high resolution high pressure liquid chromatography (HPLC) -separation,[4-5] amino acid analysis,[6] sequence analysis,[7] spectrometric analysis,[8] nuclear magnetic resonance (NMR) analysis,[9] and mass spectrometry.[10-15] Each of these methods has its own particular strengths and weaknesses. Over the past several years, members of the Rockefeller University Mass Spectrometric Research Resource have used ^{252}Cf-plasma desorption mass spectrometry (PDMS)[16] to analyze in detail more than 1200 synthetic peptides and proteins submitted by members of 16 different laboratories in the U.S. The results of these analyses demonstrated that PDMS provides highly useful information concerning the integrity and purity of these compounds and is a powerful complement to the more established methods. The importance of the use of the PDMS technique can be gauged from our rather staggering finding: almost half of the 1200 purified synthetic peptides and proteins that we examined were found to have either a molecular weight different from that calculated for the desired target material, or to contain significant amounts of unwanted peptide side-products. This chapter describes the utility of ^{252}Cf-PDMS for the analysis of synthetic peptides and proteins.

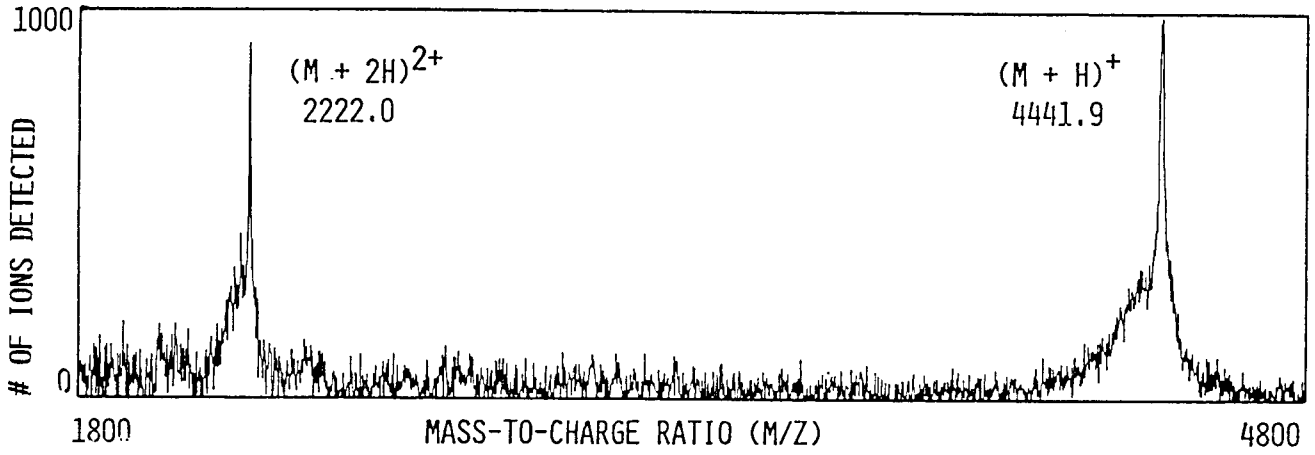
II. VERIFICATION OF THE CORRECTNESS OF THE COVALENT STRUCTURE

The most useful and easily obtained single piece of information from the ^{252}Cf-plasma desorption mass spectrometry is the molecular weight (mol wt) of the compound, as determined from the peaks corresponding to the singly and multiply protonated intact molecule. In a stepwise peptide synthesis, the identity of each added amino acid is known, and therefore a measured mol wt, which is found to agree with the mol wt calculated for the desired synthetic product, provides a relatively dependable verification that the material has been assembled correctly. Conversely, any disagreement observed between the measured and calculated mol wt of the synthetic product indicates immediately that an error or undesired modification has occurred.

Figure 1 shows that the fission fragment time-of-flight (TOF) mass spectrum of a 37-amino acid residue analog of the egg-laying hormone from the mollusk *Aplysia californica* produced in 86% yield with automated stepwise synthesis by Kent and Schiller at the California Institute of Technology. The sample is prepared for mass spectrometry by absorbing approximately 1 nmol of the peptide onto a specially prepared thin film of nitrocellulose as described previously.[17] The measured mol wt of 4441.4 U*[18] agrees well with the mol wt of 4441.2 U calculated from the amino acid sequence shown in Figure 1. This agreement provides a valuable initial verification of the correctness of the synthesized structure. Because the analysis shown in Figure 1 was completed within less than 2 h of receipt of the sample, it can be seen that this mass spectrometric procedure is relatively rapid and straightforward. It should be emphasized that this

* The measured molecular weight is taken as the simple average of the values deduced from the singly and doubly protonated molecule peaks.

```
1                                                   10                                                  19
Ile - Ser - Ile - Asn - Gln - Asp - Leu - Lys - Ala - Ile - Thr - Asp - Met - Leu - Leu - Thr - Glu - Gln - Ile -
20                                                  30                                                  37
Arg - Glu - Arg - Gln - Arg - Tyr - Leu - Ala - Asp - Leu - Arg - Gln - Arg - Leu - Leu - Glu - Lys - Gly.amide
```



MEASURED MW = 4441.4 (AVERAGE OF THE VALUES DEDUCED FROM $(M+H)^+$ & $(M+2H)^{2+}$ PEAKS)

CALCULATED MW = 4441.2

DIFFERENCE = +0.2

FIGURE 1. Partial ^{252}Cf-PDMS of a synthetic analog of the egg-laying hormone from *Aplysia californica*. M designates the intact molecule. (From Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, New York, 1988, 21. By permission of John Wiley & Sons.)

sample mol wt determination provides a necessary but not a sufficient condition for confirming the correctness of sequence. More stringent confirmation of the proposed sequence can be obtained, for example, by classical Edman sequence analysis[7] or by mass spectrometric sequence analysis.[11-14] Such detailed mass spectrometric sequence information is sometimes available directly from the normal ^{252}Cf-PDMS. Thus, for example, we were able to use PDMS to verify the detailed identity between natural and synthetic alamethicin I, a 20-residue, pore-forming peptide antibiotic.[11] The mass spectrometric verification of sequence was of special value in this case, because alamethicin has a blocked amino terminus. Unfortunately, in many cases, the amount and the nature of the fission fragment bombardment-induced fragmentation produces only weak and incomplete sequence information. In these instances, tandem mass spectrometric sequence determination is of great value.[12-14]

III. VERIFICATION OF THE HOMOGENEITY OF THE SYNTHETIC PRODUCT — IDENTIFICATION OF UNWANTED SIDE-PRODUCTS

^{252}Cf fission fragment ionization mass spectrometry is also a powerful tool for determining the homogeneity of the desired synthetic product. The strength of the method resides in its high resolving power, and it is usually straightforward in the discernment of the mass spectrum side-products that differ from the desired target material by as little as a fraction of a percent. In addition, if such side-products are observed, their masses can be determined accurately to provide an important clue to their identity. Thus, for example, the quasi-molecule ion region of the fission fragment mass spectrum obtained from a synthetic sample of the 35-residue antibacterial peptide cecropin A[19] (Figure 2) showed the presence of a small amount of undesired impurity with a mol wt 28 U higher than the desired material. The relative peak heights indicate that this unwanted side-product is present in approximately 20% abundance. The mass difference of 28 U suggested to the synthetic chemists that the error involved a failure to eliminate fully from the molecule the formyl group, which originally protected the tryptophan residue (shown circled in Figure 2). Once the presence of such an impurity is clearly recognized and its origin established, steps can be taken to eliminate its formation. Figure 3 shows another example of a relatively subtle inhomogeneity in a small methionine-containing synthetic peptide. The impurity peak has a mol wt 15.9 U higher than the desired material, suggesting the occurrence of partial oxidation of the sample, which contains a methionyl sulfur.

It is unfortunate that presently, the most popular technique for assessing the homogeneity of a purified peptide product, i. e., reversed-phase HPLC (RP HPLC), is also the technique that is often used for purifying the desired compound from the crude synthetic peptide product. It is therefore not surprising that undesired materials that co-purify with the compound of interest are also frequently not resolved in the analytical HPLC analysis. Clearly, it is desirable to check for homogeneity using an analytical technique that separates compounds by a different principle from that used in the purification. Mass spectrometry serves this function well. Figure 4A shows the results of an HPLC analysis of a purified 39-residue synthetic cytochrome c fragment where the analysis was made using a C4 RP column. The dominant chromatographic peak looks quite symmetrical and sharp, and one might predict that the synthetic product is homogeneous. Inspection of the mass spectrum obtained from this material (Figure 4B) shows that this conclusion is not correct, however. Although the desired material is present and its experimentally determined mol wt agrees with the predicted mol wt to within 0.1 U, so is a substantial amount of a second compound having mol wt 114.0 U lower. The mass difference between the latter and desired compounds suggests that a deletion(s) involving asparagine occurred during the synthesis. Subsequent HPLC analysis using a high resolution C18 RP column confirmed the presence of the second unwanted compound. The cytochrome c fragment was thus resynthesized carefully, and the mass spectrum from this second synthesis is shown in Figure 4C. The peak corresponding to the deletion peptide is now no longer present in the spectrum, demonstrating that this second synthesis yielded a much purer product.

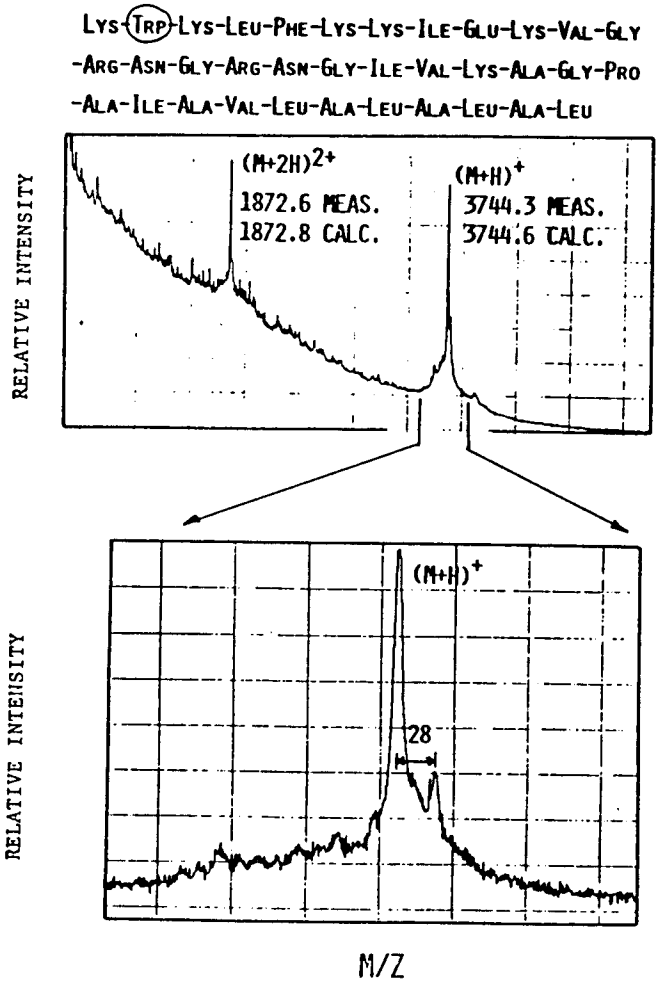


FIGURE 2. ^{252}Cf-PDMS of synthetic cecropin A. The bottom panel shows a detailed plot of the $(M+H)^+$ ion region. The measured and calculated m/z values for the $(M+H)^+$ and $(M+2H)^{2+}$ ions are given in the top panel. (From Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, New York, 1988, 21. By permission of John Wiley & Sons.)

Frequently, the peptide chemist does have an indication that an error may have occurred during the synthesis, but has little or no information concerning the detailed nature of the error. The C18 RP HPLC analysis of a purified synthetic sample of the 26-residue bee venom peptide, melittin, is shown in Figure 5A. Close inspection of the top trace obtained with 214 nm absorbance shows that the intense peak consists of three unresolved components, indicating the presence of at least three distinct compounds in the sample. The fission fragment mass spectrum of this same sample (Figure 5B) confirmed the presence of three main components, designated M, X, and Y, with abundances of 69, 21, and 10%, respectively. These mass spectrometrically inferred abundances are consistent with the abundances of the three chromatographic components. Component M has a measured mol wt of 2845.9 U, which corresponds (to within 0.5 U) to the calculated mol wt of 2846.4 U of melittin. Component X has a measured mol wt 128.0 U lower, and component Y an mol wt 28.0 U higher than the melittin peak. These mol wt differences provide valuable clues to the identities and origins of X and Y. The mass difference of 128 U suggests that X was produced by deletion of a lysine (Lys) or a glutamine (Gln) residue

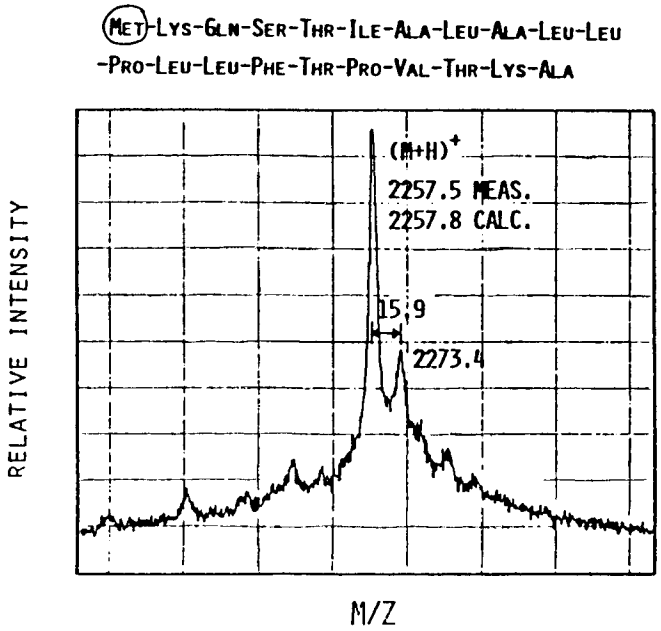


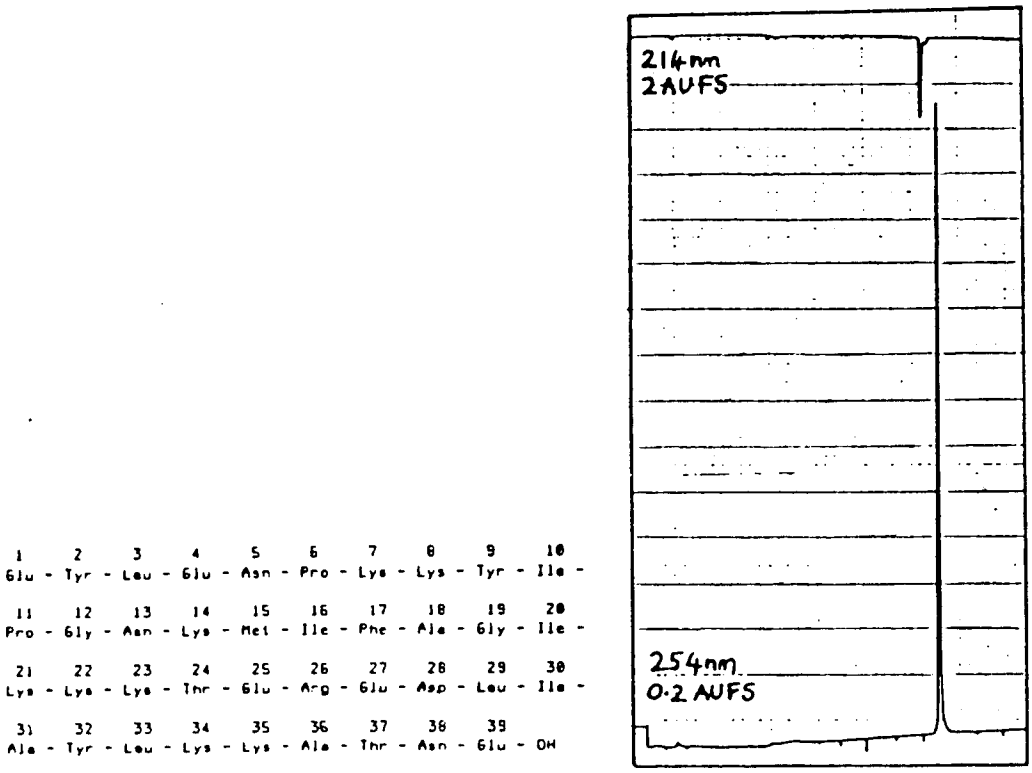
FIGURE 3. Detail of the ^{252}Cf-PDMS of a 21-residue synthetic peptide showing the quasi-molecular ion region. (From Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, New York, 1988, 21. By permission of John Wiley & Sons.)

during the stepwise synthesis. The simple mol wt measurement does not, however, provide any definitive information on the actual position(s) of the deletion(s). The mass increment of 28 U suggests that Y was produced by incomplete removal of the formyl group from the tryptophan (Trp) residue at position 19 during the final deprotection step, although again direct confirmation of this possibility cannot be obtained from the simple mol wt measurement. Although this mol wt information is valuable, it is also desirable to have available techniques that focus more tightly onto the precise site in the molecule where the synthetic error or modification has occurred. Tandem mass spectrometry appears to be a good technique for this job, because it provides a powerful means of directly pinpointing sites of error or modification. Indeed, Biemann and Scoble[14] have recently utilized tandem mass spectrometry to positively identify an internal cyclization of an aspartic acid side chain in a short synthetic peptide, and Carr and co-workers[20] have used the technique to ascertain the site of attachment of a formyl group in an undesired side-product of synthetic melittin.

We have developed a complementary approach to tandem mass spectrometry for rapidly extracting information concerning the sites of synthetic errors. The approach, which has been described previously in detail,[17,21,22] involves three sequential steps:

1. Practically nondestructive ^{252}Cf-PDMS of monolayer amounts of the peptide(s) of interest bound to a thin layer of nitrocellulose
2. Enzyme-catalyzed microscale chemical reaction of the surface-bound peptide(s) to produce structurally informative hydrolysis products
3. PDMS of these hydrolysis products

The first step determines the presence and the mol wt of unwanted byproducts, and the subsequent two steps provide information on the location in the peptides where errors have

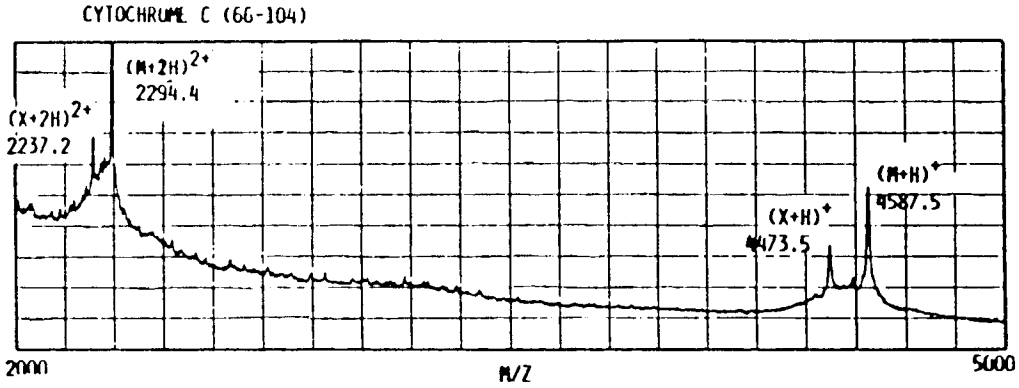


A

FIGURE 4. (A) RP HPLC of synthetic cytochrome *c* (66 to 104). Top trace obtained with 214 nm absorbance. Bottom trace obtained with 254 nm absorbance. (B) Partial ^{252}Cf-PDMS of synthetic cytochrome *c* (66 to 104). First synthesis. (C) Partial ^{252}Cf-PDMS of synthetic cytochrome *c* (66 to 104). Second synthesis. (From Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, New York, 1988, 21. By permission of John Wiley & Sons.)

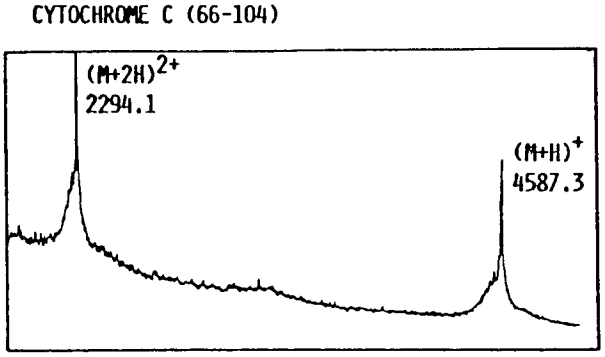
occurred. Thus, for example, inspection of the mass spectra of the previously discussed synthetic melittin sample taken prior to (Figure 5B) and after (Figure 5C) incubation with trypsin provides information concerning the regions of the molecule where the errors have occurred. Prior to any reaction, the mass spectrum of 10^{-9} mol of the sample (Figure 5B) exhibited the three peaks discussed previously. After reaction with 10^{-11} mol of trypsin, the mass spectrum (Figure 5C) exhibited a series of additional peaks, which corresponds to the protonated reaction products arising from partial hydrolysis on the carboxy-terminal side of residues 7, 21, 22, 23, and 24. Thus, for example, the peak labeled 1-22 corresponds to the tryptic fragment that includes residues 1 to 22. Because no significant peak is observed 128 U below the 1-22 peak, it can be deduced immediately that the deletion byproduct X does not arise by deletion of either Lys 7 or Lys 21. The error must then have occurred by deletion of Lys 23, Gln 25, or Gln 26. This information was given to the peptide chemists, who found upon close inspection of their records that there was indeed cause for concern during the first four synthetic cycles, which were involved in the incorporation of residues 23 to 26. The compound was therefore resynthesized, and the new preparation was found to give a mass spectrum (Figure 5D) that no longer showed the presence of any significant deletion peptides. The side-product, Y, which was hypothesized to arise from a failure to fully eliminate from the molecule the formyl group that originally protected the Trp residue at position 19, was, however, still present in the sample.

In a separate series of experiments, we have used this technique to investigate the detailed structures of such formyl group-containing byproducts, which are observed frequently when



ION SPECIES	MEASURED MW	CALCULATED MW	Δ
$(M+H)^+$	4586.5	4586.4	+0.1
$(X+H)^+$	4472.5	4586.4	-113.9
$(M+2H)^{2+}$	4586.8	4586.4	+0.4
$(X+2H)^{2+}$	4472.3	4586.4	-114.1

FIGURE 4B.



ION SPECIES	MEASURED MW	CALCULATED MW	Δ
$(M+H)^+$	4586.2	4586.4	-0.2
$(M+2H)^{2+}$	4586.3	4586.4	-0.1

FIGURE 4C.

formyl-Trp is used in the synthesis of Trp-containing peptides.[22] Figure 6 shows the RP HPLC analysis of a melittin sample prepared and purified separately than that discussed previously. In this case, a single symmetrical peak was observed, suggesting that the sample was pure. However, inspection of the PDMS obtained from 10^{-9} mol of this melittin sample bound to nitrocellulose (NC) (Figure 7a) again indicated the presence of a formyl group-containing impurity M_i^i with a mol wt 28 U higher than that of melittin. Because formylated Trp was used in the synthesis, the observed impurity was likely the result of either an incomplete deprotection

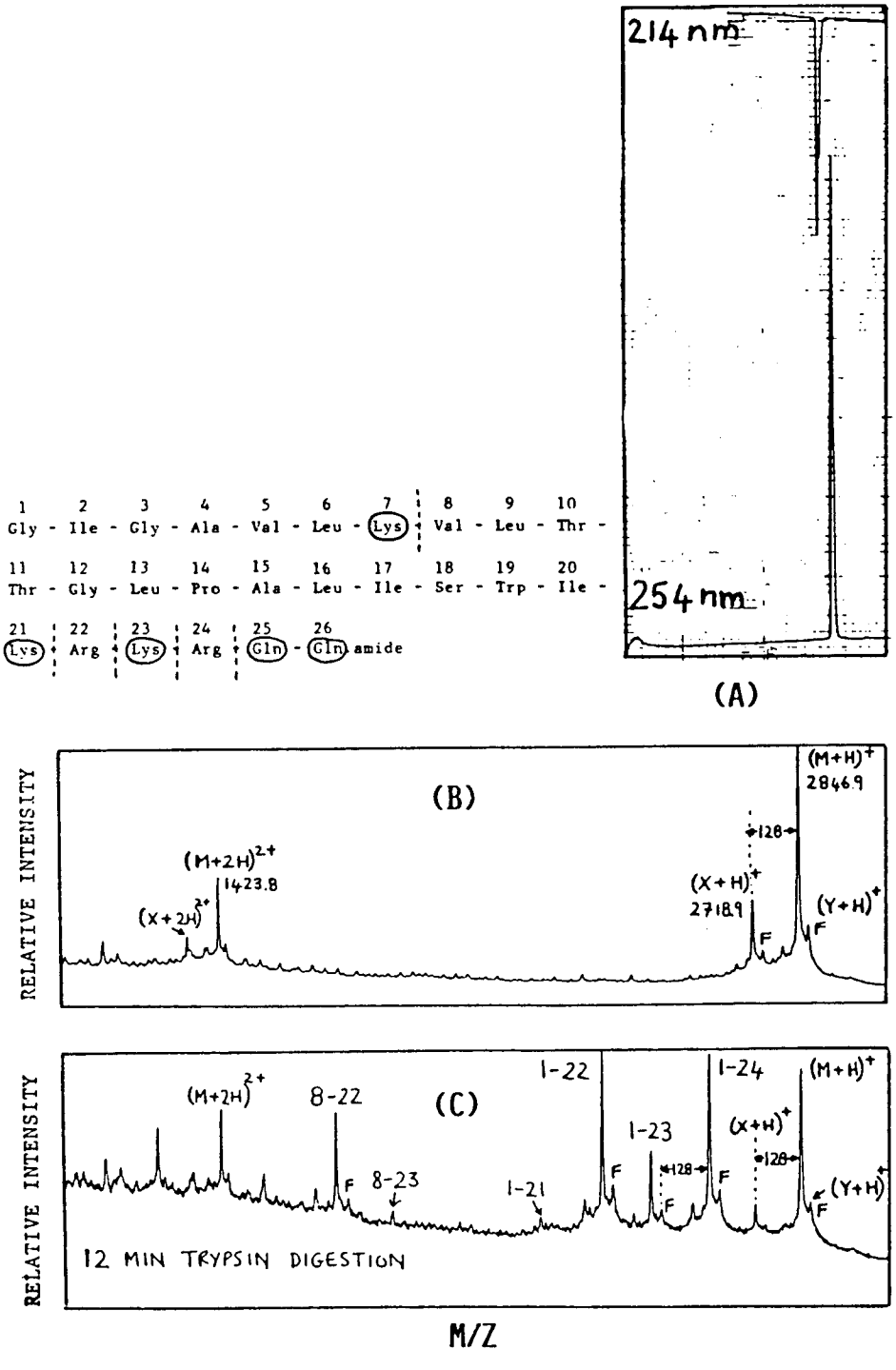


FIGURE 5. (A) C18 RP HPLC of synthetic melittin. Top trace obtained with 214 nm absorbance. Bottom trace obtained with 254 nm absorbance. (B) Partial mass spectrum of melittin. First synthesis. (C) Partial mass spectrum of the same sample of melittin after 12 min trypsin digestion of the nitrocellulose-bound peptide. The small peaks labeled F are each 28 U higher than the large adjacent peak. (D) Partial mass spectrum of melittin. Second synthesis. (From Chait, B. T., The use of ^{252}Cf plasma desorption mass spectrometry for the analysis of synthetic peptides and proteins, in *The Analysis of Peptides and Proteins by Mass Spectrometry*, McNeal, C. J., Ed., John Wiley & Sons, New York, 1988, 21. By permission of John Wiley & Sons.)

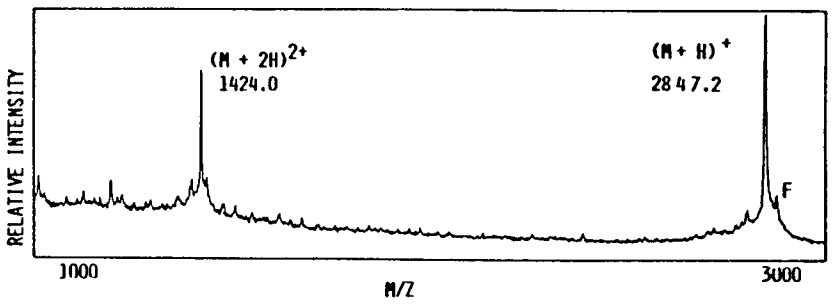


FIGURE 5 (continued).

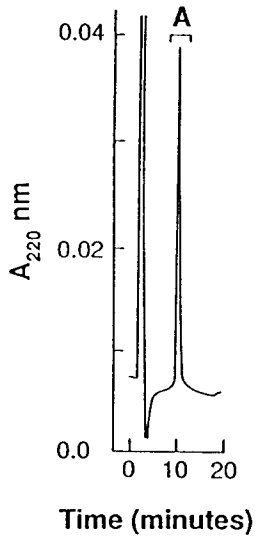


FIGURE 6. HPLC trace of the melittin products using a C-18 RP column. (From Chait, B. T., *Anal. Chem.*, 180, 387, 1989. With permission.)

of Trp (For) or a transformylation modification reaction occurring during deprotection.[3,23,24] The deprotection scheme was the low-high two-step hydrogen fluoride (HF) cleavage procedure of Tam et al.[24,25] In order to extract information about the location of the formyl group in the impurity, a series of enzymatic hydrolysis reactions was carried out on the sample mixture containing both the authentic melittin and the impurity. A description of the procedure follows.

The sample used to obtain the spectrum shown in Figure 7a was removed from the mass spectrometer and treated with trypsin. The series of observed tryptic fragments is summarized diagrammatically in Figure 8, together with the results from other enzymatic treatments (discussed later in this chapter). The tryptic fragments observed in Figure 7b occur as paired peaks, 28 U apart. The lower mass component of each pair of peaks (e.g., [1 to 22]$^+$ representing protonated melittin [1 to 22]) originates from the authentic melittin, (M_1), whereas the upper component (e. g., [1 to 22]$_i^+$) originates from the formylated melittin impurity (M_1^i). Because all the fragments are observed as doublets, we can deduce that the formyl group in the impurity, M_1^i, is located between residues 8 and 21, inclusive.

The position of the formyl group in the impurity melittin (M_1^i) was established more accurately by treatment with proteinase K of a freshly deposited sample of the mixture on NC. The mass spectrum of the resulting hydrolysis products is shown in Figure 7c. All of the enzyme-generated fragment ions in the mass spectrum occur as pairs 28 U apart, with the exception of

Gly-Ile-Gly-Ala-Val-Leu-Lys-Val-Leu-Thr-Thr-Gly-Leu-Pro-Ala-Leu-Ile
Ser-Trp-Ile-Lys-Arg-Lys-Arg-Gln-Gln-NH_2.

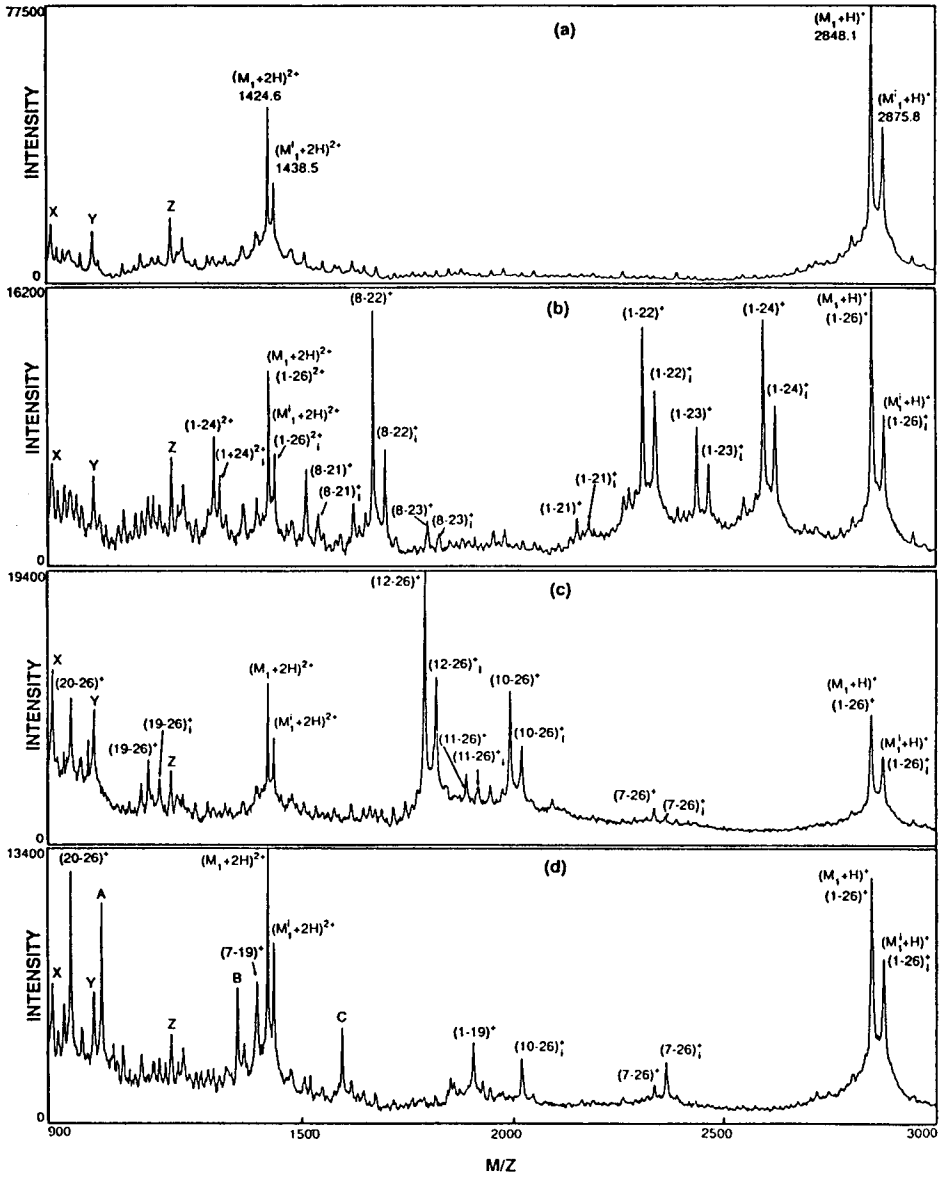


FIGURE 7. ^{252}Cf plasma TOF mass spectra of the purified synthetic melittin product (10^{-9} mol) bound to a thin film of nitrocellulose. (a) Prior to any enzymatic treatment. $(M_1+H)^+$ and $(M_1+2H)^{2+}$ are singly and doubly protonated ions of authentic melittin. $(M_1'+H)^+$ and $(M_1'+2H)^{+2}$ are similar ions of a byproduct with mass approximately 28 U higher than that of melittin. X, Y, and Z result from unidentified impurities. (b) Following a 10-min digestion with trypsin (10^{-11} mol, pH 8.0) at 37°C. The residues contained in each hydrolysis product ion are indicated on top of the corresponding peak. For example, $(1 \text{ to } 22)^+$ represents the protonated tryptic fragment of melittin, M_1, which contains residues 1 to 22, and $(1 \text{ to } 22)^+_i$ represents the protonated ion of the corresponding tryptic fragment from the melittin impurity, M_1'. The mass separation between the members of each such pair of ions is 28 U. (c) Following a 10-min reaction with proteinase K (10^{-10} mol, pH 8.5) at 37°C. Ions A, B, and C were not identified. (d) Following a 10-min reaction with α-chymotrypsin (10^{-10} mol, pH 8.5) at 37°C. (From Chait, B. T., *Anal. Biochem.*, 180, 387, 1989. With permission.)

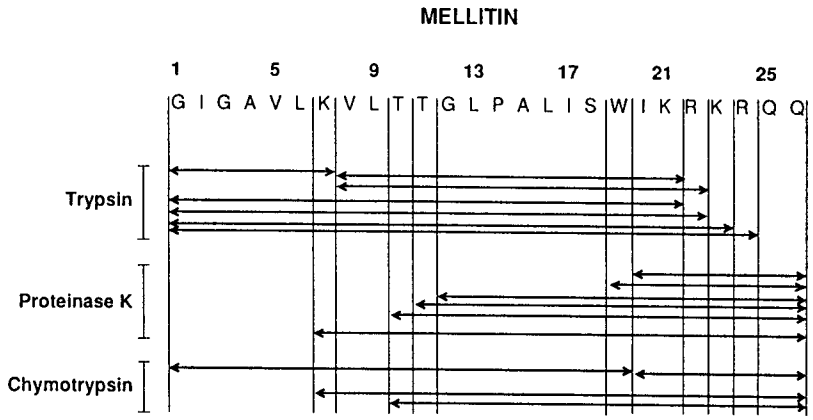


FIGURE 8. A summary of al the mass spectrometrically observed enzymatically generated fragments of melittin. these results are obtained from the data shown in Figure 7 for the digestion of melittin by trypsin, proteinase K, and α-chymotrypsin, respectively.

the ion at m/z 975.3. We attribute this ion to the fragment (20 to 26)$^+$, having a calculated m/z 956.1. The absence of a companion ion peak 28 U higher indicates that the formyl group in the impurity melittin is not located between residues 20 and 26, inclusive. On the other hand, the presence of the pair (19 to 26)$^+$ (observed m/z 1143.2, calculated m/z 1143.3) and (19 to 26)$_i^+$ (observed m/z 1170.8, calculated m/z 1171.3), which are approximately 28 U apart, indicates that the formyl group is present between residues 19 and 26, inclusive, of the impurity melittin. Therefore, the formyl group is located on Trp-19 of the impurity melittin species.

Supporting evidence for the location of the formyl group on Trp-19 was deduced from the mass spectrum (Figure 7d) obtained after α-chymotrypsin treatment (see Figure 8) of the melittin sample bound to NC. Here too, a single peak (20 to 26)$^+$ with m/z 956.6 (calculated m/z 956.1) was observed with no accompanying (20 to 26)$_i^+$ peak, again indicating the absence of the formyl group in residues 1 to 19. However, this conclusion is incorrect because the presence of the formyl group on the indole nitrogen of the Trp reduces greatly the rate of hydrolysis at Trp by α-chymotrypsin. This result was confirmed in a separate experiment comparing the rates of hydrolysis by α-chymotrypsin of two model substrates, which were identical, except that one contained formylated Trp in place of Trp.[22] We found (data not shown) that the rate of hydrolysis at Trp(For) was more than an order of magnitude slower than that at unprotected Trp. Thus, the nonobservation of (1 to 19)$_i^+$ provides additional evidence that the formyl group is present on Trp-19 in the impurity. Another manifestation of the reduced rate of hydrolysis at Trp(For)-19 is the enhanced intensities observed in Figure 7d of (10 to 26)$_i^+$ and (7 to 26)$_i^+$, compared to those of (10 to 26)$^+$ and (7 to 26)$^+$.

A similar strategy was employed to locate the positions of residual formyl groups in an abundant side-product observed in a [*p*-benzoylphenylalanine-8] dynorphin A (1 to 17) preparation in which the removal of the formyl group from Trp-14 was carried out using piperidine. In this case, the impurity was shown to result from migration of the formyl group from Trp-14 to the adjacent Lys-11 or Lys-13 residues.[22]

In addition to the errors describe above, a great many other errors occur during solid phase synthesis.[1-4] These errors range from operator- or machine-related errors to relatively subtle chemical modifications of the peptide during and after synthesis. We have found that the majority of these errors can be detected by PDMS. Thus, we have, for example, been able to detect and identify the elimination of water from amino terminal glutamine as well as from other amino acid residues; S-alkylation of methionyl residues by the tertiary butyl cation; many other examples of incomplete deprotection (see also Reference 26); the production of insertion