MASS SPECTROMETRIC SEQUENCING OF PROTEINS.

THE STRUCTURE OF SUBUNIT I OF MONELLIN.

Gail Hudson and K. Biemann
Department of Chemistry, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139

Received May 10,1976

Summary

The amino acid sequence of subunit I of the sweet protein monellin has been completed by gas chromatographic mass spectrometry. A partial acid hydrolyzate was converted to the corresponding mixture of N_ω-trifluoroethyl-polyamino alcohol-trimethylsilyl ethers and a total of 55 di- to hexapeptides were identified which could be assembled to a polypeptide chain of 44 amino acids.

Introduction

The sweet protein monellin isolated from "serendipity berries", the fruit of *Dioscoreophyllum comminsii* appeared from published data (1) to be of a size (92 amino acid residues) which would lend itself well to the gas chromatographic-mass spectrometric amino acid sequencing technique which we have developed over the years (2). Its utility had been demonstrated on a few protein fragments, for example an undecapeptide (2e) and an eicosapeptide (3) of actin as well as the carboxypeptidase inhibitor from potatoes (2g) which is 39 amino acids long. It was hoped that monellin could be sequenced by this technique using partial acid hydrolysis and enzymatic hydrolysis, even though the resulting peptide mixtures would be extremely complex. Alternatively, the presence of one single methionine residue (1) would open the possibility to cleave the protein into two sizeable polypeptide chains which could, after separation, be sequenced separately.

We were fortunate to obtain a sample of monellin from Dr. R. Cagan of the Monell Institute, Philadelphia, who also informed us that contrary to the

---

Abbreviations used: HOAc, acetic acid; TFA, trifluoroacetyl; TMSDEA, trimethylsilyl diethylamine.

earlier reports, monellin consists of two subunits of about equal size; a similar observation had also been referred to in the literature at about that time (4). Thus, it appeared merely necessary to separate the two subunits without any chemical or enzymatic cleavage, and to sequence the two subunits separately.

Although we first carried out a preliminary experiment on the partial acid hydrolysis of monellin which yielded some very useful results (see the later discussion) the major effort was devoted to the two subunits. After denaturation of monellin and carboxymethylation of the cysteine residue separation was indeed achieved (Fig. 1) and the fractions labeled A, B and C were pooled as shown. The amino acid analyses of A and C demonstrated that these fractions indeed represented two separate parts of the molecule as judged from the absence of methionine and cysteine in part A and of serine in part C. Determination of the N-terminus by Edman degradation did not lead to an extractable product from part A, suggesting that the N-terminus is arginine (there is no histidine in monellin) while part C showed the presence of glycine as the N-terminal amino acid.

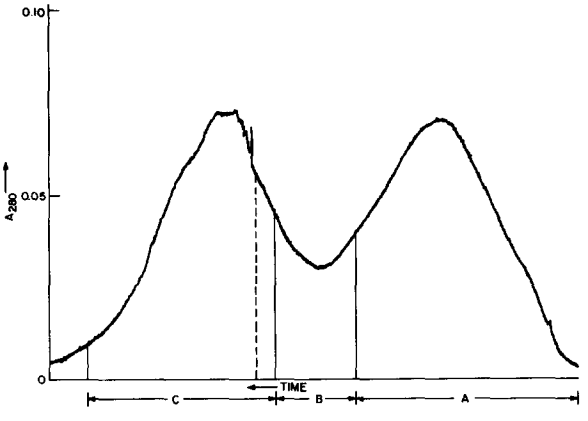

Figure 1. Separation of monellin into subunit I (Fraction A) and subunit II (Fraction C) on Biorex-70. Broken line refers to change of flow rate from 30 ml to 40 ml/hr.

As the first step, both fraction A and C were subjected to a partial acid hydrolysis and the resulting two peptide mixtures converted to the corresponding mixtures of N-trifluoroethylpolyaminoalcohol-trimethylsilyl ethers. Each of the two mixtures was then injected into a gas chromatograph coupled to a mass spectrometer and interfaced to a computer for fast data acquisition and processing (5).

During the process of interpretation of the resulting data there appeared a paper by Bohak and Li on the amino acid sequence of monellin (6). These authors employed a stepwise Edman degradation utilizing a sequenator, and reported a complete sequence for subunit II with the N-terminal glycine (thus corresponding to our fraction C), but were unable to arrive at a complete sequence for subunit I (our fraction A) because of a rapid decline of the yield of PhNCS-amino acids after step 29, resulting in an uncertain sequence 28 through 37 and no information about the remaining C-terminal region. Since it was precisely that part of the molecule whose amino acid sequence was immediately obvious from our data, we are prompted to report these results because they not only complete the sequence of subunit I but, more importantly, represent an excellent demonstration of the complementarity rather than competitiveness of the automated Edman method and our mass spectrometric sequencing technique.

Methods and Procedures

Monellin (25 mg) was carboxymethylated by dissolving the protein in 4 ml of .05 M Tris buffer containing 6 M guanidine hydrochloride (pH 8.4) and adding a solution of 6 mg sodium iodoacetate in 0.2 ml Tris buffer. The pH was maintained at 8.4 by automatic titration with 1.5 M NaOH, the reaction stopped by adding 0.4 ml HOAc, and the solution desalted using Bio Gel P-2 and 10% HOAc for the elution. The protein fraction was collected and the solvent removed on a rotary evaporator.

The two subunits were separated on a Biorex 70 column which had been washed and equilibrated with 2% HOAc. Five milligram portions of protein were loaded in 2% HOAc, and gradient eluted with 10 to 60% HOAc. The effluent was monitored at 280 nm and fractions were collected and pooled as shown in Fig. 1. Pooled fractions from two such experiments were combined and rechromatographed.

Small portions of each peak were hydrolyzed in 6N HCl at 110°C for 28 and 48 hours, respectively, for amino acid analysis on a Durrum amino acid analyzer. The results indicated 95+% purity.

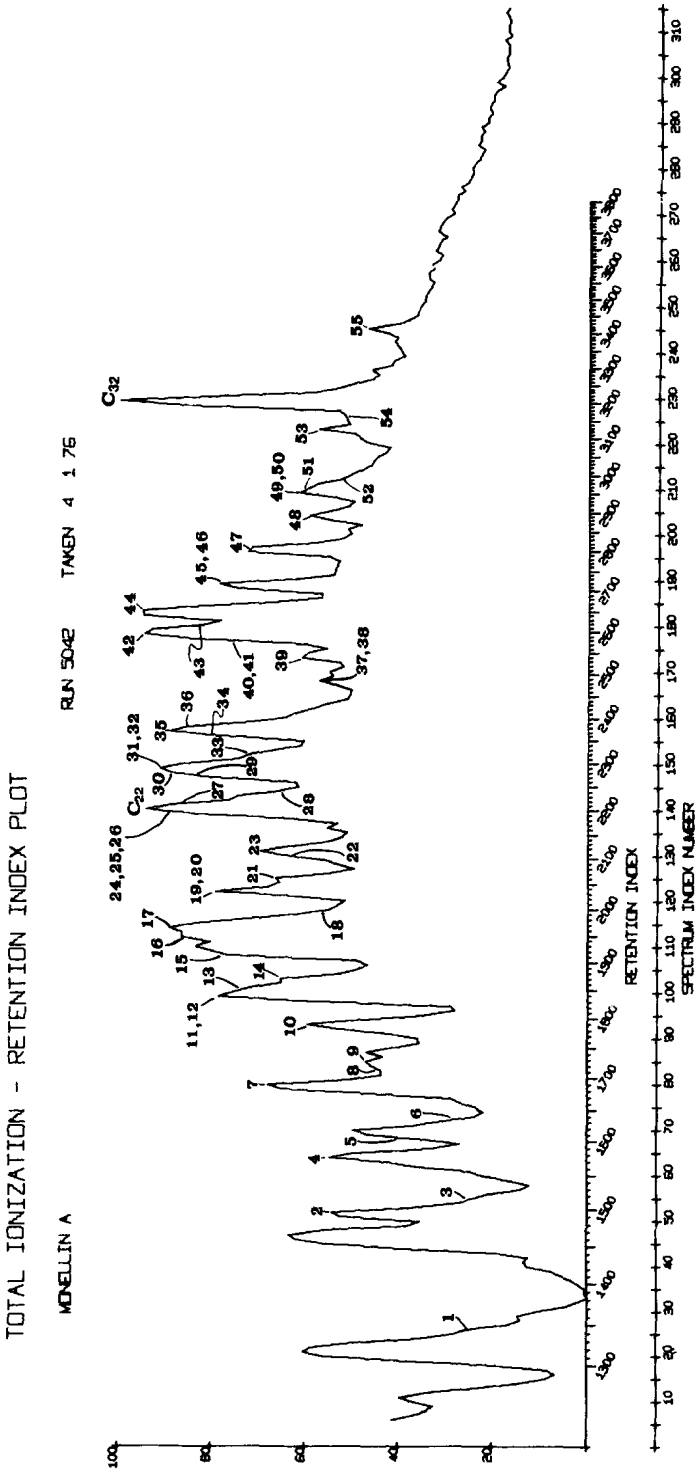

Figure 2. Total ionization plot (gas chromatogram) of derivatized partial acid hydrolyzate of subunit I of monellin (Fraction A in Fig. 1). Arabic numbers refer to peptide derivatives listed in Table 1. C_{22} and C_{32} refers to retention index standards. Unlabeled peaks at the beginning of the gas chromatogram are due to derivatives of free amino acids.

Partial acid hydrolysis was carried out by heating 3.4 mg of fraction A for 40 minutes in 6N HCl in an evacuated, sealed tube. The hydrolyzate was transferred to a flask and solvents removed on a rotary evaporator. The resulting mixture of small peptides was then hydrazinolyzed with aqueous hydrazine (1:1), esterified with 3N methanol/HCl, trifluoroacetylated with TFA-OMe in methanol, reduced with lithium aluminum deuteride, and O-trimethylsilylated with TMSDEA (Pierce). For details of these procedures see refs. 2d-g. About 20% of each of the resulting mixtures was injected together with three standard hydrocarbons into the GC-MS-computer system described elsewhere (5).

Results and Discussion

In the gas chromatogram (Fig. 2) of the derivatized partial acid hydrolyzate of subunit I, a total of 55 peptides (33 di-, 15 tri-, 4 tetra-, 2 penta-, and one hexapeptide(s)) were identified (Table 1) from the 315 consecutively

Table 1

Peptides identified in partial acid hydrolyzate of subunit I

No. in Fig. 2	Derivative[a] of	Sequence in Fig. 3	No. in Fig. 2	Derivative[a] of	Sequence in Fig. 3
1	Gly-Pro	39-40	29	Val-Tyr	12-13
2	Ala-Ser	14-15	30	Asp-(Leu)-Ser	22-24
3	Val-Pro	41-42	31	Phe-Arg	19-20
4	(Leu)-(Leu)	34-35	32	Arg-Phe	36-37
5	Ala-Asp	21-22	33	(Leu)-Tyr	10-11
6	Asp-Gly	38-39	34	Gly-Pro-Val-Pro	39-42
7	(Leu)-Ser	23-24	35	Lys-(Leu)-(Leu)	33-35
8	Gly-Arg	31-32	36	(Leu)-Ser-Glu	23-25
9	Arg-Ala	20-21	37	Ser-Glu-Asp	24-26
10	Asp-(Leu)	22-23	38	Asp-Tyr	26-27
11	Ser-Asp	15-16	39	Arg-Glu-(Leu)	1-3
12	Gly-Pro-Val	39-41	40	Tyr-Glu	6-7, 8-9
13	(Leu)-Arg	35-36	41	Glu-Tyr	7-8
14	Glu-(Leu)	2-3, 9-10	42	Val-Pro-Pro-Pro	41-44
15	Thr-Arg	29-30	43	Glu-(Leu)-Lys	2-4
16	Ser-Glu	24-25	44	Tyr-Lys	27-28
17	Lys-(Leu)	17-18,33-34	45	(Leu)-Phe-Arg	18-20
18	Lys-Thr	28-29	46	(Leu)-Tyr-Val	10-12
19	(Leu)-Phe	18-19	47	Lys-(Leu)-Phe	17-19
20	Glu-Asp	25-26	48	Gly-Pro-Val-Pro-Pro	39-43
21	Val-Pro-Pro	41-43	49	Lys-(Leu)-(Leu)-Arg	33-36
22	Asp-Lys	16-17	50	Arg-Lys-(Leu)-(Leu)	32-35
23	Arg-Glu	1-2	51	Glu-(Leu)-Tyr	9-11
24	Arg-Lys	32-33	52	Tyr-Glu-(Leu)	8-10
25	Tyr-Ala	13-14	53	Pro-Val-Pro-Pro-Pro	40-44
26	Gly-Tyr	5-6	54	Asp-Tyr-Lys	26-28
27	Phe-Asp	37-38	55	Gly-Pro-Val-Pro-Pro-Pro	39-44
28	Thr-Arg-Gly	29-31			

[a] (Leu) refers to Leu or Ile; Arg converted to Orn during hydrazinolysis; Gln and Asn converted to Glu and Asp during acid hydrolysis.

recorded mass spectra and their gas chromatographic behavior. The interpretation of the mass spectral data as well as the computer techniques for the further resolution of multicomponent gas chromatographic peaks were outlined in the earlier papers (particularly refs. 2e-g).

The very last peak (No. 55) in the chromatogram immediately reveals the sequence (Gly-Pro-Val-Pro-Pro-Pro) unaccounted for in the incomplete structure of Bohak and Li (6). Indeed, the same hexapeptide derivative was clearly borne out by the data obtained from the much more complex partial acid hydrolyzate of intact monellin.

In addition to this segment, the data summarized in Fig. 2 and Table 1 permit us now to complete the structure for subunit I of monellin as shown in Fig. 3. Comparison with the proposal of Bohak and Li reveals not only the above mentioned missing part (sequence 39-44), the connecting residue 38 (Asx), one of the unidentified residues (No. 30, Arg), but also confirms most of the uncertain portion (sequence 28-37). The startling discrepancy lies in the sequence 32-33 which are shown as ?-Arg in ref. 6. Our data (Fig. 3 and Table 1) show that the "unidentified" amino acid does not exist but that a previously unreported lysine residue is interposed between arginine and leucine.

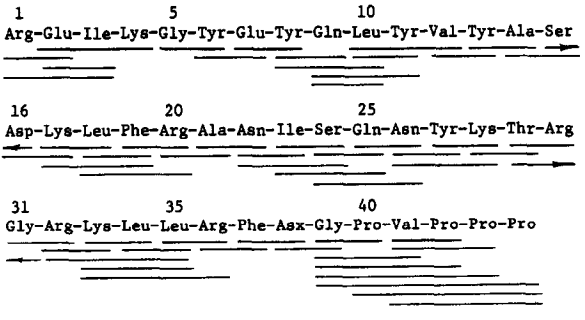

Figure 3. Structure of subunit I of monellin. Underlining refers to peptides obtained by partial acid hydrolysis (see Table 1). Differentiation of Leu and Ile, Asp and Asn, Glu and Gln based on the partial structure of Bohak and Li (6).

This discrepancy is difficult to rationalize and while we first thought that the unidentified amino acid No. 32 in ref. 6 is the same as 31, namely glycine, we found the evidence for the Arg-Lys-Leu sequence overwhelming (peptides 24, 35, 49 and 50). Further evidence for these features of the structure shown in Fig. 3 is the amino acid composition of subunit I (Table 2),

Table 2.

Amino acid composition of fraction A (subunit I)

Amino Acid	28 hour Hydrolysis	48 hour Hydrolysis	According to Sequence (Fig. 3)
Aspartic acid*	4	4	4
Threonine	1.0	1.0	1
Serine	1.9	1.6	2
Glutamic acid	4.3	4.3	4
Proline	3.9	4.0	4
Glycine	3.1	3.1	3
Alanine	2.0	1.9	2
Valine	1.9	2.1	2
Isoleucine	1.8	2.0	2
Leucine	3.7	3.8	4
Tyrosine	4.9	4.5	5
Phenylalanine	1.9	1.9	2
Lysine	3.8	3.6	4
Arginine	4.3	4.0	5

*The amino acid analyses were normalized to aspartic acid equalling 4 moles/mole of subunit I.

which demands a lysine (in addition to one each of Asp, Gly, Val and four Pro) to be placed within or beyond the uncertain portion of the reported partial structure (6). It should be noted that the more detailed interpretation of the mass spectra necessary to distinguish leucine from isoleucine and the enzymatic degradations required to distinguish Asp from Asn and Glu from Gln have not yet been carried out. However, all of these, with the exception of position 38, were unambiguously identified by Bohak and Li and we are presently completing the interpretation of these and other enzymatic digests which will be discussed in a future more detailed paper.

As mentioned earlier the coincidence of these two investigations under-

scores our conviction that the mass spectrometric approach to peptide sequencing and the automated Edman degradation provide mutually supporting data and should not be viewed as mutually exclusive or merely competitive. This is, for example, particularly well demonstrated by the redundancy of sequence information residing in the multitude of overlaps of mass spectrometrically identified peptides (underlining in Fig. 3) precisely in the regions where the Edman degradation has led to doubtful information or none at all (sequence 28-44). There is no possibility of an unidentified amino acid in a mass spectrum of a peptide derivative, but there is the greater difficulty in differentiating leucine and isoleucine. On the other hand, in the acid hydrolyzate one needed to identify the dipeptide Glu-Tyr (No. 41) which has, of course, almost the same retention index as Tyr-Glu to recognize that there are two consecutive Tyr-Glx moieties. Finally, the evidence of the Lys-Gly (sequence 4-5) is indirect: Placing the sequence 1-4 at the end (after Pro-44) would lead to N-terminal Gly which is inconsistent with the data.

Acknowledgment

The authors are indebted to Dr. Robert H. Cagan, Monell Chemical Senses Center, University of Pennsylvania, for the sample of monellin and suggestions for the separation of the subunits, Dr. Lorenz Cueni and Dr. James Riordan, Biophysical Research Laboratories, Harvard Medical School, for the amino acid analyses and help with the separation, and Robert Sauer, Biological Laboratory, Harvard University, for the determination of the N-termini. This work was supported by a Research Grant from the National Institutes of Health (GM 05472).

References

1. Morris, J.A., Martenson, R., Deibler, G., and Cagan, R.H. (1973) J. Biol. Chem. 248, 534-539.
2. (a) Biemann, K., Gapp, F., and Seibl, J. (1959) J. Am. Chem. Soc. 81, 2274-2275.
   (b) Biemann, K. and Vetter, W. (1960) Biochem. Biophys. Res. Commun. 2, 93-96.

(c) Förster, H.-J., Kelley, J.A., Nau, H., and Biemann, K. (1972) in Chemistry and Biology of Peptides (J. Meienhofer, ed.), pp. 679-686, Ann Arbor Science Publishers, Ann Arbor.
(d) Nau, H. (1974) Biochem. Biophys. Res. Commun. 59, 1088-1096.
(e) Kelley, J.A., Nau, H., Förster, H.-J., and Biemann, K. (1975) Biomed. Mass Spectrom. 2, 313-325.
(f) Nau, H., Förster, H.-J., Kelley, J.A., and Biemann, K. (1975) Biomed. Mass Spectrom. 2, 326-339.
(g) Nau, H. and Biemann, K. (1976) Analyt. Biochem. 72, in press.
3. Nau, H., Kelley, J.A., and Biemann, K. (1973) J. Am. Chem. Soc. 95, 7162-7164.
4. Wlodawer, A. and Hodgson, K.O. (1975) Proc. Natl. Acad. Sci. USA, 72, 398-399.
5. Hites, R.A. and Biemann, K. (1968) Analyt. Chem. 40, 1217-1221.
6. Bohak, Z., and Li, S.-L. (1976) Biochim. Biophys. Acta 427, 153-170.