BULLETIN OF THE CHEMICAL SOCIETY OF JAPAN, VOL. 51 (6), 1909—1910 (1978)

Application of Structure Construction Logic in CHEMICS for Determination of Amino Acid Sequence of Polypeptide by CI Mass Spectrometry

Yasuhiko YOTSUI, Hisashi NOMURA, Mitsuji SANO, Yoshihiro KUDO,* and Shin-ichi SASAKI*

Research Institute, Daiichi Seiyaku Co. Ltd., Edogawa-ku, Tokyo 132
**Miyagi University of Education, Aoba, Sendai 980*

(Received December 5, 1977)

Synopsis. The amino acid composition for each oligopeptide fragment in the hydrolysis mixture of polypeptides was determined using mass spectrometer with a CI source. The amino acid sequence was elucidated by structure construction logic in CHEMICS.

There have been several investigations[1-4] using mass spectrometry to determine the sequence of amino acids in polypeptide. Biemann *et al.*[4] developed a method in which di- or tripeptide fragments generated by the partial hydrolysis of the original peptide were determined by gas chromatography-mass spectrometry and reconstructed with the aid of a computer. This paper describes a new technique for the determination of the amino acid sequence in polypeptides using CI mass spectrometry and structure construction logic. The latter technique, established by two of the authors (S.S. and Y.K.) enumerates the isomers based on the molecular formula (*composition*) and probable partial structures (*components*) of an organic compound in the study of system CHEMICS.[5] The enumeration logic was thought applicable to the determination of the amino acid sequence of polypeptides, where the *composition* was the number and the kind of amino acids, and the *components* the nature of oligopeptides originated by the partial hydrolysis of the sample of an unknown polypeptide. In practice the *composition* may be determined by ordinary amino acid analysis and the *components* elucidated by CI mass spectrometry of the oligopeptide fragments generated by the partial acid hydrolysis of polypeptides.

The new determination method consists of six steps. Step 1: The composition of the unknown polypeptide was elucidated by amino acid analysis. Step 2: The sample was partially hydrolyzed with acid into oligopeptide fragments (mostly di- and tripeptides). The carboxylic and amino groups of all the fragments except the ε-amino group of lysine were converted into methoxycarbonyl and acetoamido groups, respectively, by the successive treatment of the mixture with CH_3OH–HCl and an equivalent mixture of acetic anhydride and acetic anhydride-d_6. Step 3: The CI mass spectrum for the esterified and acetylated mixture at step 2 was measured. This mixture was introduced into the spectrometer, *without any previous separation.* Each oligopeptide fragment may be detected as a doublet due to the acetylated molecule (M) and deuterioacetylated molecule (M+3). The execution of steps 4 through to 6 were performed by computer. Step 4: The constitution of amino acids in each oligopeptide was determined by comparing the mass with the composition of the sample determined at step 1. Step 5: The number of possible oligopeptides present in the sample were computed. The probable polypeptide structures (candidates) were constructed from the set of oligopeptides as executed in CHEMICS. Step 6: This routine involves the elimination of candidates which did not fulfil the necessary conditions. Compared with other methods, this method has the advantages of simplicity and high sensitivity brought about by the effective use of CI mass spectrometry and the computer logic of structure construction. Furthermore the logic can treat any type of peptide, either cyclic and branched.

PGL-LEU-TYR-GLU-ASP-LYS-PRO-ARG-ARG-PRO-TYR-ILE-LEU

Fig. 1. Neurotensin.

The present method was implemented for commercially available neurotensin[6] (Fig. 1). Ordinary amino acid analysis demonstrated that the polypeptide consisted of thirteen amino acids of nine types. One of two glutamic acids in the sample molecule was shown to be a pyroglutamic acid (PGL),** (there was no reaction of the sample with ninhydrin reagent) and this data was fed to the computer.

After hydrolyzing neurotensin with 6M-HCl (80 °C, 4 h), the hydrolysis mixture was treated with CH_3OH–HCl followed by a mixture of acetic anhydride and acetic anhydride-d_6. This was conducted at room temperature and overnight. The reaction mixture was introduced into a mass spectrometer with a CI source *without previous separative treatment.* The CI spectrum of the mixture suggested the presence of 14 oligopeptides (Nos. 1 through 10′) and the amino acid composition was computed by comparing the mass with the amino acid composition of the original polypeptide neurotensin, as shown in Table 1. In addition to these oligopeptides, FD mass spectrum demonstrated the presence of a dipeptide, ARG–ARG.

In order to construct the structure of neurotensin it is not necessary to use all the kinds of oligopeptides shown in Table 1. The reasons for this are as follows: No. 11 seems to be a fragment of No. 7, therefore No. 7 appears to be present in the form of either LYS–ASP–PRO or ASP–LYS–PRO. Paying attention to ASP, No. 7 may be connected to No. 4 to yield GLU–ASP–

** PGL was used for pyro-glutamic acid instead of P-GLU for the convenience of computer input and output.

Table 1. Peptide fragments of neurotensin analyzed by CI mass spectra

No.	*m/e*		Components
1	299	QM^+	LEU, PGL
2	301	QM^+	LEU, LEU
2′		QM^+	PRO, ASP
3	335	QM^+	PRO, TYR
4	347	QM^+	ASP, GLU
5	351	QM^+	LEU, TYR
6	381	QM^+	GLU, TYR
7	429	QM^+	PRO, LYS, ASP
7′		QM^+	LEU, LEU, LYS
8	464	QM^+	LEU, LEU, TYR
8′		QM^+	PRO, ASP, TYR
9	494	QM^+	LEU, GLU, TYR
10	561	QM^+	LEU, LEU, PRO, TYR
10′		QM^+	PRO, PRO, ASP, TYR
11	300	F	LYS, ASP
12	416	F	LEU, PRO, TYR
13	387	QM^+	ARG, ARG

QM^+: quasi-molecular ion of oligopeptide $(M+1)^+$
F: fragment ion ARG-ARG (No. 13): detected by FD mass spectrometry

LYS–PRO (No. 14). Nos. 2′, 7′, 8′, and 10′, which are not in agreement with No. 14, may be eliminated. Furthermore, two possible structures, ASP–GLU–TYR and TYR–GLU–ASP, are derived from Nos. 4 and 6. This structure is successively connected to No. 14 and 9 to give LEU–TYR–GLU–ASP–LYS–PRO (No. 15) as a longer partial structure of neurotensin.

Thus all the possible structures for neurotensin may be constructed from Nos. 1, 2, 3, 5, 8, 10, 12, 13, and 15. The number of structures (candidates) amounts to eight

```
001 P PGL-LEU-LEU-TYR-PRO-LEU-TYR-GLU-ASP-LYS-PRO-ARG-ARG
002 P PGL-LEU-PRO-TYR-LEU-LEU-TYR-GLU-ASP-LYS-PRO-ARG-ARG
003 P PGL-LEU-ARG-ARG-PRO-TYR-LEU-LEU-TYR-GLU-ASP-LYS-PRO
004 P PGL-LEU-TYR-PRO-LEU-LEU-TYR-GLU-ASP-LYS-PRO-ARG-ARG
005 P PGL-LEU-TYR-GLU-ASP-LYS-PRO-PRO-TYR-LEU-LEU-ARG-ARG
006 P PGL-LEU-TYR-GLU-ASP-LYS-PRO-ARG-ARG-PRO-TYR-LEU-LEU
007 P PGL-LEU-TYR-GLU-ASP-LYS-PRO-TYR-LEU-LEU-PRO-ARG-ARG
008 P PGL-LEU-TYR-GLU-ASP-LYS-PRO-TYR-LEU-LEU-ARG-ARG-PRO

                                                      END
```

Fig. 2. Candidate polypeptides afforded by computer-aided CI mass spectrometry for neurotensin.

as shown in Fig. 2. In the present method, LEU cannot be distinguished from ILE, since both possess the same molecular weight. Consequently all the possible candidates amount to 24.

Experimental

Isobutane CI mass spectra were determined on a HITACHI RMU-6MG mass spectrometer with a combined EI-CI source under conditions of 100 eV ionizing potential and 80 μA ionizing current. FD mass spectra were obtained with a JEOL JMS-D100 mass spectrometer equipped with a combined FD-FI-EI source using an emitter current of 10—15 mA and a cathode potential of −5kV.

Partial hydrolysis of neurotensin: About 0.1 μmol of neurotensin was heated with 6M hydrochloric acid at 80 °C for 4 h in a sealed tube. Subsequently the hydrochloric acid was removed *in vacuo* and the residue used without purification.

Esterification: A solution of the partial hydrolysis product of neurotensin in 0.5 ml of methanol was allowed to stand overnight at room temperature.

Acetylation: To a solution of the methyl esters of the oligopeptides, 0.25 ml of acetic acid (d_0: d_3=1: 1) and 0.25 ml of acetic anhydride (d_0: d_6=1: 1) were added, then the mixture was allowed to stand overnight. Excess reagent was removed *in vacuo* and the residue used directly for the determination of CI spectrum.

Under these conditions, the ε-amino group of lysine appears resistant to acetylation. No mass peaks corresponding to oligopeptides including diacetamidolysine methyl ester were detected.

References

1) H. K. Wipf, P. Irving, M. McCamish, R. Venkataraghavan, and F. W. McLafferty, *J. Am. Chem. Soc.*, **95**, 3369 (1973).

2) G. M. Schier, B. Halpern, and G. W. A. Milne, *Biomed. Mass Spectrom.*, **1**, 212 (1974).

3) S. A. Poku, G. W. Wood and D. E. Schmidt Jr., *Biomed. Mass Spectrom.*, **2**, 121 (1975).

4) H. Nau and K. Biemann, *Anal. Biochem.*, **73**, 175 (1976).

5) S. Sasaki, H. Abe, Y. Kudo, and T. Yamasaki, 1977 ACS Symposium Series on Computer-assisted Structure Elucidation of Organic Compounds, No. 54, 108 (1977); Y. Kudo and S. Sasaki, *J. Chem. Inf. Comp. Sci.*, **16**, 43 (1976).