DETERMINATION OF THE AMINO ACID SEQUENCE OF PEPTIDES USING DIPEPTIDYL AMINOPEPTIDASES

Yu.A. OVCHINNIKOV and A.A. KIRYUSHKIN
Shemyakin Institute for Chemistry of Natural Compounds, USSR Academy of Sciences, Moscow V-312, USSR

1. Introduction

We have earlier shown that the possibilities of the mass spectrometric method of amino acid sequence determination can be considerably extended by its combination with chemical and enzymatic methods [1, 2]. In the present paper another method is described involving: i) digestion of the starting peptide with dipeptidyl aminopeptidase I (DAP I) to give a mixture of dipeptides; ii) esterification and trifluoroacetylation of the dipeptides to afford trifluoroacetylpeptide methyl esters; iii) separation of the mixture by gas-liquid chromatography (GLC); iv) mass spectrometric determination of the amino acid sequence in peptide derivatives.

2. Methods

DAP I (catepsin C) treated with diisopropyl fluorophosphonate was a gift of Dr. J. Ken McDonald. Peptide digestion was carried out at 37°, pH 5.0, as described in [3, 4]. The resulting peptide mixture was brought to pH 3, evaporated, dried in vacuo, esterified with 0.1 N HCl/MeOH (20°, 24 hr) and then acylated by methyl trifluoroacetate in dioxan–dimethyl formamide solution (1:1) in the presence of triethylamine. The separation of the peptide derivative mixture and mass spectra measurements were carried out on a gas chromatograph – mass spectrometer combination LKB 9000 (LKB-Producter, Sweden). For gas chromatographic separation a glass column (1.5 m long by 0.2 cm, i.d.) packed with 1% OV–17 on Supelcort (80–100 mesh) was used. Other operation parameters were: the rate of helium gas carrier, 20 ml/min; injection port, ≃300°; molecular separator, 280°; ion source, 280°.

3. Results and discussion

DAP I which is able to split dipeptides off the *N*-terminus of the polypeptide chain can be utilized in protein primary structure determination. For this purpose it has been suggested that the dipeptides released in the course of digestion should be identified [3, 4]. It seems more convenient, however, to proceed as follows: i) exhaustively digest the initial peptide with DAP I; ii) split off the *N*-terminal amino acid of the initial peptide according to Edman and again subject the resulting compound to exhaustive digestion with DAP I. This would supply a series of overlapping peptides [5].

The main difficulties in both approaches lie in the separation and identification of dipeptides. This can be achieved by paper and/or column chromatography [6]; however, the procedures are somewhat time consuming and require relatively large amounts of substance. On the contrary, separation of dipeptides (in the form of trifluoroacetylpeptide methyl esters) can be achieved quite easily by GLC with subsequent structure elucidation by mass spectrometry (cf. [7]).

We here explain the advantages of the proposed technique using the example of a tetradecapeptide. The tetradecapeptide was isolated from the tryptic hydrolysate of cytoplasmic aspartate aminotransferase [8] and had the following amino acid composition: Arg_1, Gly_5, Val_2, Glu_1, Ser_1, Leu_2, Thr_1, Ala_1. The peptide (0.5 μmole) was subjected to exhaustive digestion with DAP I, and the resulting peptide mixture was esterified, acylated with CF_3COOMe and then analyzed on the gas chromatograph – mass spectrometer. The results of GLC are given in fig. 1. The mass spectra of the separated components of the mixture showed the first peak to be

North-Holland Publishing Company – Amsterdam

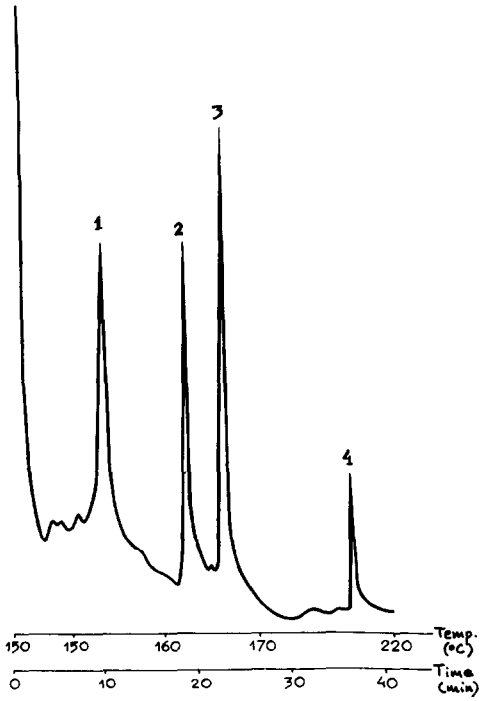

Fig. 1. Gas chromatographic separation of trifluoroacetyl-peptide methyl esters prepared from the tetradecapeptide. Isothermal at 150° for 6 min, then programmed at 1° per min to 170°, further at 4° per min.

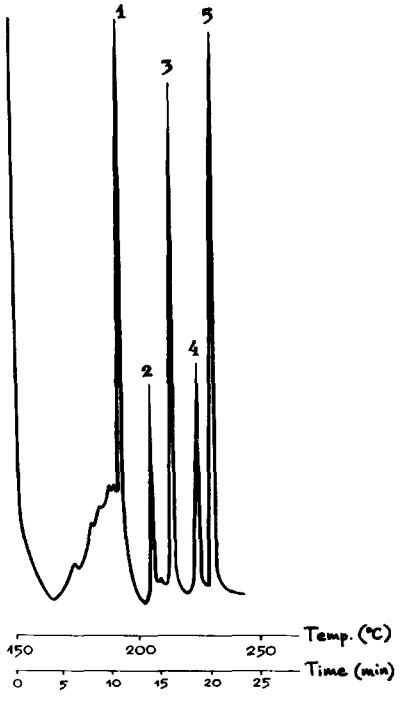

Fig. 2. Gas chromatographic separation of trifluoroacetyl-peptide methyl esters prepared from the tridecapeptide. Programmed at 4° per min from 150° (starting temp).

"Scheme 1"

Val–Gly–Gly–Val–Glu–Ser–Leu–Gly–Gly–Thr–Gly–Ala–Leu–Arg

Val–Gly Gly–Val Glu–Ser Leu–Gly Gly–Thr Gly–Ala
Gly–Gly Val–Glu Ser–Leu Gly–Gly Thr–Gly Ala–Leu

TFA–Gly–Thr–OMe, the second and the third are mixtures of TFA–Val–Gly–OMe + TFA–Gly–Ala–OMe and TFA–Gly–Val–OMe + TFA–Leu–Gly–OMe, respectively, and peak 4 contains TFA–Glu-(OMe)–Ser–OMe.

Another sample of the tetradecapeptide (0.5 μmole) was treated according to Edman and an amino acid was split off, which was found to be valine. The resulting tridecapeptide was subsequently digested with DAP I and treated as described above. The results of the GLC are given in fig. 2. Peaks 1, 2, 3, 4 and 5 are TFA–Ala–Leu–OMe, TFA–Gly–Gly–OMe, TFA–Thr–Gly–OMe, TFA–Ser–Leu–OMe and TFA–Val–Glu$(OMe)_2$, respectively.

This peptide is an exceptionally rare case in which, due to the specificity in the amino acid composition (there are 5 glycine residues), a knowledge of all the possible depeptides only gives us the partial sequence: Val–Gly–Gly–(Ala; Thr–Gly; Leu–Gly–Gly; Val–Glu–Ser)–Leu–Arg. However, even here it is suffi-

cient to determine the fourth amino acid residue (which can be easily done by the Edman method) to obtain the total sequence shown in scheme 1. In most cases a knowledge of the dipeptide sequences is sufficient to establish the primary structure of the starting polypeptide (cf. [9]).

The advantages of the technique described here lie in its simplicity and high sensitivity: easily interpreted mass spectra can be obtained from 0.03–0.01 μmole of the starting peptide. The use of capillary columns and other ionization methods (e.g. chemical ionization) can further increase the sensitivity of the technique. The deciphering of the mass spectra of trifluoroacetylpeptide methyl esters presents no difficulties even if they are not completely separated by GLC. Moreover, since the number of all possible dipeptides in protein chemistry is limited, their mass spectra can be recorded beforehand, thus making possible automatic deciphering of the spectra with the aid of computers.

The main shortcoming of the technique is due to specificity of DAP I which does not cleave proline peptide bonds as well as peptides with *N*-terminal lysine or arginine residues. The latter difficulty is possibly not very serious, since usually in protein chemistry the complete sequences of tryptic peptides only are determined in which these amino acids are located on *C*-termini. On the contrary, the amino acid sequence in proline-containing peptides can be determined up to but not including the amino acid next to the proline residue. The feasibility of this approach to the structural investigation of proline-containing peptides can be estimated by their previous treatment with leucine aminopeptidase. On the other hand, depeptidyl aminopeptidases are known which are able to hydrolyse peptides with *N*-terminal basic amino acids and proline peptide bonds [10]. The use of a mixture of different dipeptidyl aminopeptidases can extend the scope of the technique.

However, in spite of its limitations, even DAP I alone can be successfully utilized in sequence studies using the technique described in this paper.

Acknowlegement

The authors are greatly indebted to Dr. J. Ken McDonald for supply of a DAP I sample.

References

[1] M.M. Shemyakin, Yu.A. Ovchinnikov, E.I. Vinogradova, A.A. Kiryushkin, M.Yu. Feigina, N.A. Aldanova, Yu.B. Alakhov, V.M. Lipkin, A.I. Miroshnikov, B.V. Rosinov and S.A. Kazaryan, FEBS Letters 7 (1970) 8.

[2] N.A. Aldanova, E.I. Vinogradova, S.A. Kazaryan, B.V. Rosinov and M.M. Shemyakin, Biokhimiya 35 (1970) 854.

[3] J.K. McDonald, P.X. Callahan, B.B. Zeitman and S. Ellis, J. Biol. Chem. 224 (1969) 6199.

[4] J.K. McDonald, B.B. Zeitman, T.J. Reilly and S. Ellis, J. Biol. Chem. 244 (1969) 2693.

[5] R.-A.A. Valiulis and V.M. Stepanov, Biokhimiya 36 (1971) 866.

[6] R.X. Callahan, J.A. Shepard, T.J. Reilly, J.K. McDonald and S. Ellis, Anal. Biochem. 38 (1970) 330.

[7] F. Weygand, A. Prox, H.H. Fessel and K.K. Sun, Z. Naturforschung 20b (1965) 1169.

[8] Yu.A. Ovchinnikov, A.A. Kiryushkin, Ts.A. Egorov, N.G. Abdulaev, A.P. Kiselev, N.N. Modyanov, E.V. Grishin, A.P. Sukhikh, E.I. Vinogradova, M.Yu. Feigina, N.A. Aldanova, V.M. Lipkin, A.E. Braunstein, O.L. Polyanovsky and V.V. Nosikov, FEBS Letters 17 (1971) 133.

[9] V.Yu. Gavrilov, A.D. Frank-Kamenetsky and M.D. Frank-Kamenetsky, Biokimiya 31 (1966) 799.

[10] J.K. McDonald, P.X. Callahan, S. Ellis and R. Smith, in: Tissue Proteinases, eds. A.J. Barrett and J.T. Dingle (North-Holland Publishing Co., Amsterdam, 1970) p. 69.