Journal of Chromatography, 647 (1993) 301–309
Elsevier Science Publishers B.V., Amsterdam

CHROMSYMP. 2806

Rapid analysis of enzymatic digests of a bacterial protease of the subtilisin type and a "bio-engineered" variant by high-performance liquid chromatography–frit fast atom bombardment mass spectrometry

W.D. van Dongen* and C. Versluis

Bijvoet Centre for Biomolecular Research, Department of Mass Spectrometry, Utrecht University, Utrecht (Netherlands)

P.D. van Wassenaar

Unilever Research Laboratory, P.O. Box 114, 3130 Vlaardingen (Netherlands)

C.G. de Koster, W. Heerma and J. Haverkamp

Bijvoet Centre for Biomolecular Research, Department of Mass Spectrometry, Utrecht University, Utrecht (Netherlands)

ABSTRACT

Amino acid sequencing of a subtilisin-type bacterial protease and a bio-engineered variant was carried out by investigating various enzymatic digests using HPLC–frit fast atom bombardment MS methods. The fast atom bombardment mass spectral data allowed rapid identification of the enzymatically generated peptides and differentiation between both proteins. The feasibility of determining the positions and nature of mutations in the amino acid sequence depends mainly on the size of the peptides containing the modifications.

INTRODUCTION

Elucidating the amino acid sequence of proteins can be done either at the DNA level or at the protein level. DNA sequencing can be carried out relatively easily with considerable speed, but there are many situations in which this approach cannot be followed and sequencing must be done directly at the protein level. A commonly applied method for elucidating the primary structure of large proteins often includes proteolysis followed by chromatographic separation and structural analysis of the resulting fragments. For the latter purpose classical Edman degradation is more and more being replaced by faster mass spectrometric methods.

Fast atom bombardment (FAB) is a very suitable ionization method for combined high-performance liquid chromatography–mass spectrometry (HPLC–MS) [1–5].

A frit FAB interface makes it possible to administer a flow of 2–10 μl/min of column effluent to the ion source of the mass spectrometer under stable conditions. The bombarding xenon beam is directed to a fritted disc covering the exit of a thin fused-silica capillary that connects the HPLC system with the ion source [6].

In this study we used a conventional HPLC

* Corresponding author.

0021-9673/93/$06.00 © 1993 Elsevier Science Publishers B.V. All rights reserved

system coupled to a frit FAB interface and a sector mass spectrometer for the analysis of various protein digests. An advantage of FAB is that the mass spectra of peptides show not only the protonated molecules but also ions that are characteristic of the amino acid sequence in the peptide. Moreover, the behaviour of peptides under FAB conditions is well documented in the literature [7–10], so that not only is confirmation of the presence of expected peptides in mixtures possible, but also structural determination of unknowns is feasible.

The aim of this study was to identify specific mutations in the sequence of an enzyme variant (a subtilisin type of bacterial protease) produced by protein engineering. The strategy followed was to first subject the enzyme to specific cleavages by cyanogen bromide and trypsin and/or the less specific proteolysis procedures of pepsin, chymotrypsin and autolysis, followed by the analysis of the resulting peptides by HPLC–frit FAB-MS.

EXPERIMENTAL

Digestion of the proteases

Cyanogen bromide digestion. The bacterial protein and its bio-engineered variant (pretreated with 0.1 *M* 2-mercaptoethanol at 37°C for 2 h) were dissolved to 10 mg/ml in 70% formic acid. A 50-fold molar excess over methionine of solid cyanogen bromide (Pierce) was added. After flushing with nitrogen the digest was stored in a dark room for 18 h, followed by addition of 10 ml of water and lyophilization.

Tryptic digestion. The resulting cyanogen bromide digests were further digested with 1-tosylamido-2-phenylethylchloromethyl (TPCK) treated trypsin (Wornington) for 16–20 h at 37°C with a substrate-to-enzyme ratio of 100:1 (w/w) in 0.1 *M* ammonium bicarbonate buffer solution at pH 8.5. The digest was then lyophilized and redissolved in the starting HPLC solvent A (see below) with an approximate sample concentration of 2 mg/ml.

Pepsin digestion. The bacterial protein and its bio-engineered variant were dissolved to 20 mg/ml in 99% formic acid. Pepsin dissolved in 1 *M* hydrochloric acid (0.05 mg/ml) was added in a substrate-to-enzyme ratio of 50:1 (w/w). After flushing with nitrogen the digest was stored in a dark room for 18 h, followed by addition of 10 ml of water and lyophilization. The digest was then redissolved in the starting HPLC solvent A with an approximate sample concentration of 2 mg/ml.

Chymotrypsin digestion. The digestion procedure with chymotrypsin (Boehringer) was the same as for trypsin. After lyophilization the digestion solution was redissolved in the starting HPLC solvent A with a sample concentration of about 1–3 mg/ml.

Autolysis. The proteins were incubated at a protein concentration of 1 mg/ml at 60°C for 25 min in 0.1 *M* sodium hydrogencarbonate. Further autolysis was then stopped by addition of trifluoroacetic acid (TFA) (0.1 ml per ml of sample) and the samples were further stored at −20°C.

Liquid chromatography

Equipment. Binary gradient separation of the digests was carried out using two HPLC pumps (Shimadzu Model LC9A, Kyoto, Japan). Separated components were monitored at 214 nm with a Shimadzu SPD-6AV UV detector fitted with a standard 8-μl flow cell. The injection valve (Model 7010) with a 50-μl injection loop was obtained from Rheodyne (Berkeley, CA, USA).

LC separation. Solvent A: 0.06% TFA (v/v) (gradient grade for chromatography, Merck, Darmstadt, Germany) and 1% glycerol (v/v) (Jansen Chimica, Beerse, Belgium) to provide a FAB matrix in bidistilled water. Solvent B: 0.06% TFA (v/v) in acetonitrile–water (80:20, v/v) solution and 1% glycerol (v/v). The gradient used was 100% A, isocratic for 5 min, followed by a linear gradient to 70% B over the next 45 min, and finished with a linear gradient of 70–100% B for a further 5 min. A C_{18} column, type 218TP54 (250 mm × 4.0 mm) (Vydac, CA, USA) with a flow-rate of 1 ml/min, was used. A Jeol pneumatic splitter device (MS-PNS) maintained a constant flow-rate to the mass spectrometer of about 10 μl/min, effectively using a split ratio of 100:1.

Mass spectrometry

Positive FAB mass spectra, first field-free region (1st FFR) and 3rd FFR B/E-linked scan spectra were obtained with a Jeol JMS-SX102/102A four-sector instrument of a B_1E_1–B_2E_2 geometry. HPLC–MS mass spectra were obtained with MS-1 operating at 3000 resolution (10% valley definition). The acceleration voltage was maintained at 6 kV and the magnet was scanned from 10 to 4000 u in 5 s. The most stable conditions were obtained with a source temperature of 85°C. The estimated probe tip temperature was 5–10°C less than that of the ion source block; the probe temperature cannot be controlled separately. On-line HPLC–B/E-linked scan spectra were obtained with MS-1 only. MS–MS mass spectra were acquired by selecting the desired precursor ion with MS-1, and colliding the ion of 6 keV translational energy in a collision cell at ground potential located in the 3rd FFR of the instrument [11,12]. The resulting fragment ions were determined by scanning MS-2. Linked scan spectra were recorded from a 50% attenuated main beam using air (1st FFR) and helium (3rd FFR) as collision gas. Xenon was used as the FAB gas; the gun was operated at 6 kV and a 5 mA discharge current.

RESULTS AND DISCUSSION

The amino acid sequence of the subtilisin-type (B) bacterial protease (269 amino acids, molecular mass 26 683, monoisotopically) is presented in Fig. 1 [13]. The variant protease (V) obtained by protein engineering has similar dimensions; the difference is caused by differences in a few amino acids.

For the HPLC–FAB-MS analysis it is necessary to hydrolyse the proteases into manageable fragments (preferably <3000 Da). As each protein has a unique character, defined by its sequence and higher order structure, the hydrolysis strategies used must be tailored to these proteases. Optimization procedures for hydrolysis are focused on obtaining maximum coverage of the sequence of the proteases, to be certain that all substitutions are traced. Thus, four complementary hydrolysis strategies were chosen: cyanogen bromide degradation followed

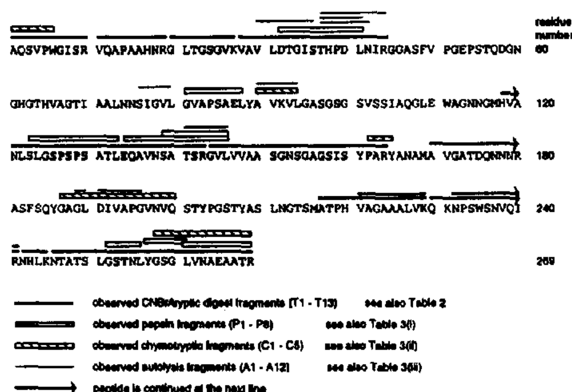

Fig. 1. Amino acid sequence of the bacterial protease (B) and the observed fragments from various hydrolysis strategies.

by tryptic cleavage, pepsin digestion, chymotrypsin digestion and autolysis. Table I gives the sites of cleavage and the specificity of the used processes. Cyanogen bromide tryptic treatment (Table II) cleaves highly specifically at sites well distributed in these proteases. Cyanogen bromide digestion preceding tryptic action is necessary for denaturation purposes; without cyanogen bromide treatment the HPLC–UV chromatograms of the digests showed no significant breakdown of the proteins. This procedure results in an easily predictable and limited number of peptides in the digest. The peptides of protease B (with known sequence) can be identified unequivocally on the basis of their MH^+ ions. The MH^+ ion chromatograms can be easily used for identification purposes. Important additional sequence information could be obtained by the less specific proteolysis strategies, using the characteristic series of sequence fragment ions of the various peptides (Table III).

In Tables II and III the identified and expected fragments (Table II only) are compiled together with their pseudomolecular masses and retention times for both proteins studied.

Fig. 2 shows the total ion current (TIC) chromatograms of the cyanogen bromide/tryptic digest peptide mixtures. The unique occurrence of the MH^+ ions in the chromatographic time domain is demonstrated in the selected mass chromatograms (Fig. 3) and allowed an unambiguous identity assignment of the peaks in the

TABLE I

CLEAVAGE SITES OF THE HYDROLYSIS METHODS

Enzymatic digestion	Cleavage sites	Specificity
(i) Cyanogen bromide/trypsin	Cyanogen bromide: carboxyl site of Met (M) trypsin[a]: carboxyl site of Arg (R), Lys (K)	Relatively high Relatively high
(ii) Pepsin digestion[b,c]	On either site of Phe (F), Leu (L), Tyr (Y)	Relatively low
(iii) Chymotrypsin digestion[a,d]	Carboxy site of Phe (F), Trp (W), Tyr (Y), Leu (L), Met (M)[e], His (H)[e]	Relatively low
(iv) Autolysis	Difficult to predict	Low

[a] Peptide bonds involving Pro (P) relatively resistant.
[b] Cleavage sites vary from one protein to another.
[c] Peptide bonds involving Arg (R), Lys (K), Pro (P) and Ile (I) are not hydrolysed.
[d] In order of decreasing susceptibility.
[e] Peptides bonds sometimes involved in cleavage.

TABLE II

EXPECTED PEPTIDE FRAGMENTS AFTER CYANOGEN BROMIDE/TRYPTIC DIGESTION

t_R = Retention time.

Peptide code	Residues	Peptide expected	MH^+	t_R (min)	Found B/V [a,b]
T1	242–245	NHLK	511.3	14	+/+
T2	170–180	AVGATDQNNNR	1159.6	15	+/+
T3	11–19	VQAPAAHNR	963.5	16	+/+
T4	20–27	GLTGSGVK	718.4	19	+/+
T5	221–229	VAGAALVK	799.5	22.2	+/+
$T6_B$	217–229	ATPHVAGAAALVK	1205.7	23.5	+/−
T7	230–241	QKNPSWSNVQIR	1456.8	24.6	+/+
T8	232–241	NPSWSNVQIR	1200.6	25	+/+
T9	1–10	AQSVPWGISR	1100.6	26.1	+/+
T10	144–164	GVLVVAASGNSGAGSISYPAR	1933.0	26.6	+/+
T11	246–269	NTATSLGSTNLYGSGLVNAEAATR	2368.2	28	+/+
T12	28–44	VAVLDTGISTHPDLNIR	1821.0	29	+/+
T13	118–143	HVANLSLGSPSPSATLEQAVNSATSR	2594.3	30.5	+/+
Tu_V	?–?	unidentified	>4000?	32.5	−/+
–	230–231	QK	275.2	–	−/−
–	165–169	YANAm[c]	539/521	–	−/−
–	93–117	VLGASGSGSVSSIAQGLEWAGNNGm[c]	2319/2301	–	−/−
–	181–216	AFSQYGAGLDIVAPGVNVQSTYPGST YASLNGTSm[c]	3580/3562	–	−/−
–	45–92	GGASFVPGEPSTQDGNGHGTHVAGTI AALNNSIGVLGVAPNAELYAVK	4588.3	–	−/−

[a] B = Traced in bacterial protein; V = traced in variant.
[b] + = Found; − = not found.
[c] m = Homoserine or homoserine lactone.

TABLE III

IDENTIFIED PEPTIDE FRAGMENTS RESULTING FROM VARIOUS DIGESTS

Peptide code	Residues	Identity	MH^+	t_R (min)	B/V [a,b]
(i) Pepsin digestion					
P1	252–256	GSTNL	491.2	17.8	+/+
P2	257–261	YGSGL	496.2	21.1	+/+
P3	139–146	SATSRGVL	79.4 [c]	22.3	+/+
$P4_V$	81–88	GVAPNAEL	770.4	23	−/+
$P4_B$	81–88	GVAPSAEL	743.4	24	+/−
P5	134–146	EQAVNSATSRGVL	1331.7	24	+/+
P6	32–41	DTGISTHPDL	1055.5	25	+/+
P7	123–133	SLGSPSPSATL	1016.5	28.5	+/+
P8	261–269	LVNAEAATR	944.5	28.5	+/+
(ii) Chymotrypsin digestion					
C1	162–165	PARY	506.3 [c]	17	+/+
C2	258–269	GSGLVNAEAATR	1145.6	22.2	+/+
C3	90–94	AVKVL	529.3 [c]	24	+/+
C4	1–6	AQSVPW	687.3	27.2	+/+
C5	187–200	GAGLDIVAPGVNVQ	1309.6 [c]	28.8	+/+
(iii) Autolysis					
A1	189–190	GL	189.1	16.5	+/+
A2	37–42	THPDLN	696.3 [c]	18.1	+/+
A3	191–196	DIVAPG	571.3	21	+/+
A4	37–41	THPDL	582.3	21.5	+/+
A5	141–146	TSRGVL	632.4	22.4	+/+
A6	30–36	VLDTGIS	704.4 [c]	22.5	+/+
A7	90–94	AVKVL	529.3 [c]	24	+/+
A8	37–44	THPDLNIR	965.5	23.5	+/+
A9	76–80	SIGVL	488.3	27.5	+/+

[a] B = Traced in bacterial protein; V = traced in variant.
[b] + = Found; − = not found.
[c] Identified from the unseparated mixture using 3rd FFR MS–MS spectra.

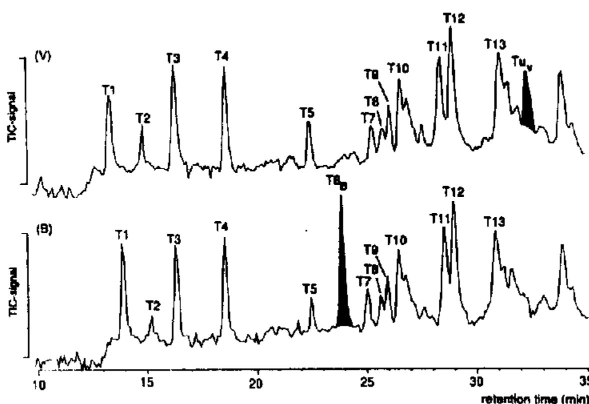

Fig. 2. Total ion current chromatograms of the bacterial protease (B) and the bio-engineered variant (V).

TIC chromatogram. The sequences of the various peptides were confirmed by their HPLC–FAB mass spectra.

Mass spectra of a reasonable quality even for the larger peptides could be obtained after separation by HPLC. As a typical example, the mass spectrum of the peptide T12 with assignment of sequence ions [14] is given in Fig. 4. In spite of the presence of a relatively strong background the spectrum allows confirmation of the expected peptide sequence.

From the obtained HPLC–FAB-MS data amino acid substitutions in variant V can be traced by carefully inspecting and comparing the TIC profiles of the digests and the MH^+ mass

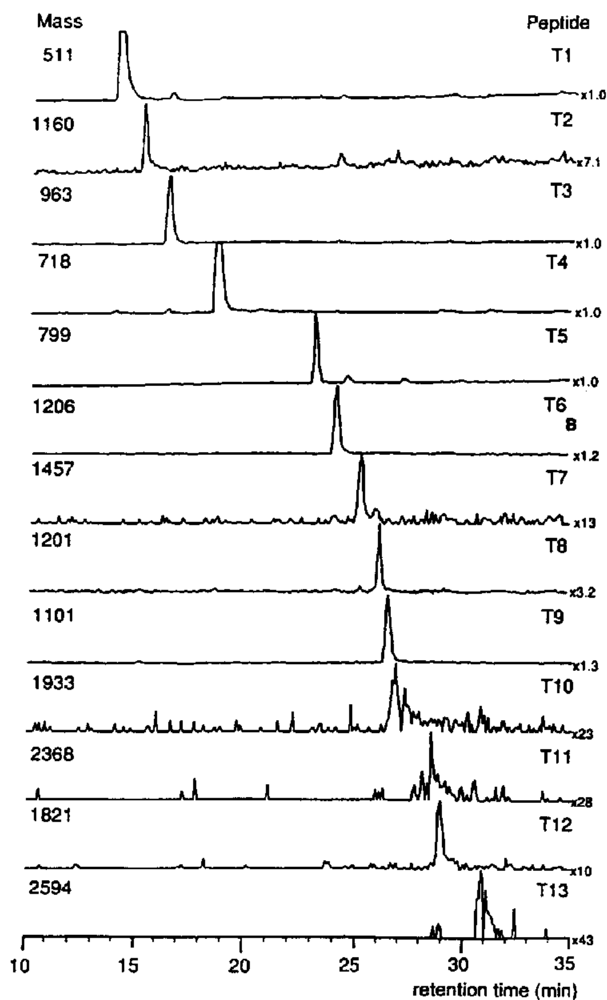

Fig. 3. Mass chromatograms of MH^+ ions in cyanogen bromide/tryptic digest peptide mixtures.

chromatograms of the individual peptides from the two proteases.

Some marked differences in the TIC chromatograms of the cyanogen bromide/tryptic digests were found. The marked peak in Fig. 2B could be traced only in the digest of protease B and is identified as peptide $T6_B$ (positions 217–229, Table II). This peptide is a product of cleavage after a methionine and a C-terminal lysine residue. The marked peak in Fig. 2V is only present in the digest of protease V and is assigned as peptide Tu_V (Table II). The molecular mass of peptide Tu_V is apparently beyond the scanning mass range at the chosen conditions, so that only a partial mass spectrum could be recorded. Based on some fragment ion peaks in the spectrum that can be assigned to A-, B- and a few Y″- type ions, the peptide probably represents the sequence fragment 181–229. Apparently cyanogen bromide scission in V is not occurring in this sequence part, which can be explained by assuming that the methionine residue at 216 present in protease B is replaced by another amino acid. Complete sequence analysis of Tu_V is in progress.

Comparison of the MH^+ mass chromatograms of the various components in both pepsin digests revealed another difference between the two proteases. The pepsin digests (Table III) contain different peptides, $P4_B$ and $P4_V$, respectively. Owing to the low specificity of pepsin the peptide mixtures are highly complex and so are the mass spectra. Usually only the MH^+ ions outgrow the abundant chemical background, which obstructs full sequence analysis. If the normal FAB mass spectra (*e.g.* $P4_V$ $MH^+ = m/z$ 770, Fig. 5A) show many interfering ions, 3rd FFR B/E-linked scanning of the MH^+ ion is the most obvious choice to eliminate background ions and to improve the quality of the resulting mass spectra. However, in a continuous-flow system, it is very difficult to select the desired precursor ion with MS-1 and tune MS-2, since the precursor ions are only present during about 1 min of the HPLC run. Moreover, its intensity is continuously changing during the elution of the component. Therefore, instead of using the "real" precursor ions, background ions have to be used for tuning procedures. Unfortunately, the intensity of the background ions at the mass of the precursor ion was not constant under continuous-flow conditions so that only poor-quality 3rd FFR spectra could be obtained. As an alternative, experimentally simpler 1st FFR B/E-linked scans of the MH^+ were performed. Thus better spectra with a relatively low chemical background, which could be easily interpreted, were obtained. A disadvantage of B/E-linked scan spectra compared with the normal mass spectra is the relatively strong ion intensity discrimination in the low-mass region as a result of the small translational energy of the low-mass

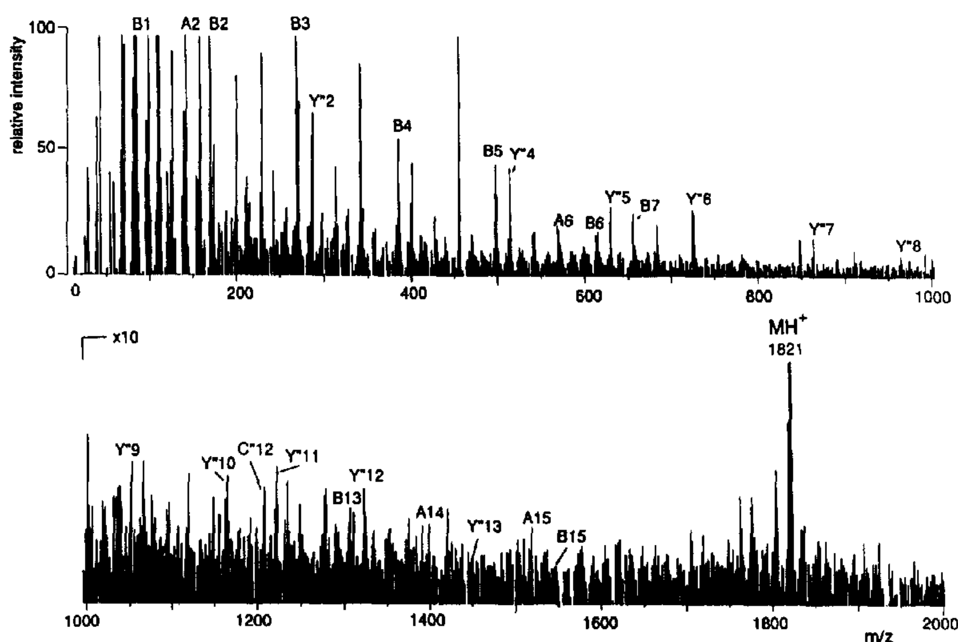

Fig. 4. Positive-ion FAB mass spectrum of a typical digest peptide VAVLDTGISTHPDLNIR (peptide T12); sequence ions are indicated.

fragment ions. Immonium ions are less than 20% of the precursor ion mass and will thus effectively be suppressed in B/E-linked scans (Fig. 5B and V). The 1st FFR B/E spectra of the protonated molecules of the peptides $P4_B$ and $P4_V$ show unequivocally the substitution of serine (S) at position 85 in B by asparagine (N) in V (Fig. 5B and V). Note that for both proteases B and V the Y_5'' ions (at *m/z* 516 and 543, respectively) are very pronounced in the B/E-linked scan spectra. This relative abundance of the Y_5'' ion is caused by a favourable fragmentation of the peptide bond between proline and its preceding amino acid leaving the charge at the C-terminal ion [15].

Chymotryptic and autolytic digestion also lead to a complex peptide fragment mixtures, resulting in the generation of several mass spectra in which only the MH^+ can be identified. These spectra are not of high enough quality for sequencing purposes. The FAB mass spectra of the unseparated digests (B and V) exhibit a number of MH^+ ions representing the more hydrophobic peptides. These ions occur at intensity sufficient for 3rd FFR MS–MS analysis. The peptides of which the sequence could be confirmed from static FAB-MS–MS spectra are indicated in Table III.

The identified peptide fragments obtained by applying the four hydrolysis strategies have been matched to the relevant positions in the total protease sequences. As shown in Table I and Fig. 1 there is a considerable amount of redundancy in the data, as many of the peptides originate from overlapping sequences. Nevertheless approximately 72% of the sequence of the proteases could be assigned from the various identified peptides within a few days.

Work is in progress to develop an improved fragmentation strategy using a combination of proteases and advanced chromatographic methods resulting in easy-to-analyse peptide mixtures effectively covering the total sequence of this type of proteases. This HPLC–MS procedure should allow easy and rapid check of the sequence of novel subtilisin variants produced by site-directed or localized random protein engineering approaches.

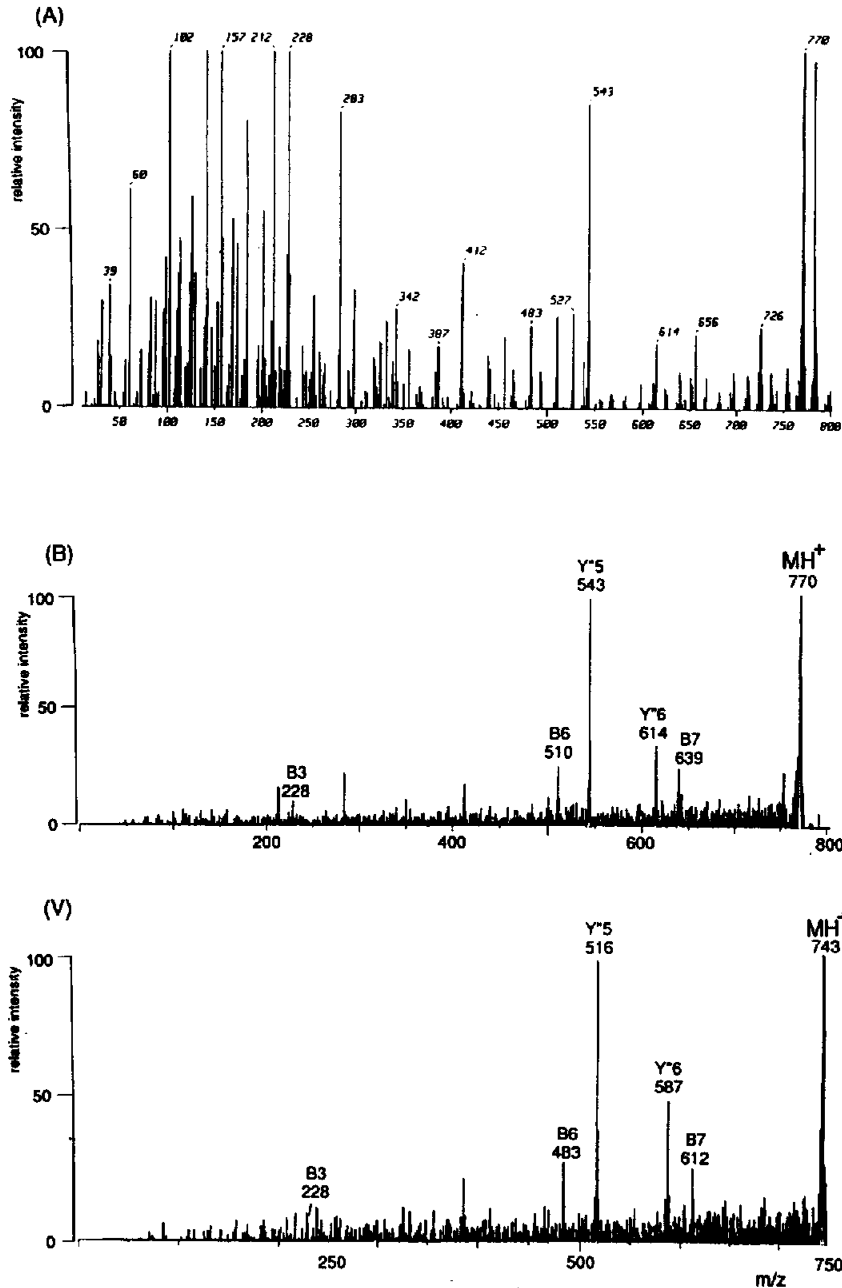

Fig. 5. (A) FAB–mass spectrum of the pepsin digest of protease V at $t_R = 23$ min. (V) B/E-linked scan of $m/z = 770$ of the pepsin digest of protease V at $t_R = 23$ min. (B) B/E-linked scan of $m/z = 743$ of the pepsin digest of protease B at $t_R = 24$ min.

REFERENCES

1 R.M. Caprioli, W.T. Moore, B. DaGue and M. Martin, *J. Chromatogr.*, 443 (1988) 355.

2 D.J. Bell, M.D. Brightwell, W.A. Neville and A. West, *Rapid Commun. Mass Spectrom.*, 4 (1990) 88.

3 D.J. Bell, M.D. Brightwell, W.A. Neville and A. West, *Org. Mass Spectrom.*, 4 (1991) 454.

4 W.J. Henzel, J.H. Bourell and J.T. Stults, *Anal. Biochem.*, 187 (1990) 228.

5 D.S. Jones, W. Heerma, P.D. van Wassenaar and J. Haverkamp, *Rapid Commun. Mass Spectrom.*, 5 (1991) 192.

6 Y. Ito, T. Takeuchi, D. Ishi and M. Goto, *J. Chromatogr.*, 346 (1985) 161.

7 S.A. Carr, G.D. Roberts and M.E. Hemling, in C.N. McEwen and B. Larsen (Editors), *Mass Spectrometry of Biological Materials*, Practical Spectroscopy Series, Vol. 8, Marcel Dekker, New York, 1990, p. 87.

8 K. Biemann, *Biomed. Mass Spectrom.*, 16 (1988) 9.

9 K. Biemann and S.A. Martin, *Mass Spectrom. Rev.*, 6 (1987) 1.

10 K. Biemann, *Annu. Rev. Biochem.*, 61 (1992) 977.

11 S.A. Martin, R.S. Johnson, C.E. Costello and K. Biemann, in C.J. McNeal (Editor), *The Analysis of Peptides and Proteins by Mass Spectrometry*, Wiley, New York, 1988, p. 135.

12 R.K. Boyd, *Int. J. Mass Spectrom. Ion Processes*, 75 (1987) 243.

13 C. Betzel, S. Klupsch, G. Papendorf, S. Hastrup, S. Branner and K.S. Wilson, *J. Mol. Biol.*, 223 (1992) 427.

14 P. Roepstorff and J. Fohlman, *Biomed. Mass Spectrom.*, 11 (1984) 601.

15 W. Heerma, C. Versluis, H. Lankhof, R.C.H.M. Oudejans, F.P. Kooiman and A.M.Th. Beenakkers, *Anal. Chim. Acta*, 248 (1991) 553.