
LC-Mass Spectrometry Analysis of N- and C-Terminal Boundary Sequences of Polypeptide Fragments by Limited Proteolysis

Justin G. Stroh, Pat Loulakis, Anthony J. Lanzetti, and Julie Xie*

PGRD-Groton Laboratories, Pfizer Inc., Groton, Connecticut, USA

Limited proteolysis is an important and widely used method for analyzing the tertiary structure and determining the domain boundaries of proteins. Here we describe a novel method for determining the N- and C-terminal boundary amino acid sequences of products derived from limited proteolysis using semi-specific and/or non-specific enzymes, with mass spectrometry as the only analytical tool. The core of this method is founded on the recognition that cleavage of proteins with non-specific proteases is not random, but patterned. Based on this recognition, we have the ability to determine the sequence of each proteolytic fragment by extracting a common association between data sets containing multiple potential sequences derived from two or more different mass spectral molecular weight measurements. Proteolytic product sequences derived from specific and non-specific enzymes can be accurately determined without resorting to the conventional time-consuming and laborious methods of SDS-PAGE and N-terminal sequencing analysis. Because of the sensitivity of mass spectrometry, multiple transient proteolysis intermediates can also be identified and analyzed by this method, which allows the ability to monitor the progression of proteolysis and thereby gain insight into protein structures. (J Am Soc Mass Spectrom 2005, 16, 38–45) © 2004 American Society for Mass Spectrometry

Proteins are dynamic systems containing structured domains linked by flexible loop/hinge regions. Flexibility is important for protein conformational changes that carry out specific functions, such as protein-ligand and protein-protein interactions [1–3]. Limited proteolysis is a widely used method to probe conformational changes, gain insights into protein tertiary structure, and to define domain boundaries and dynamics of proteins [4–6]. This approach, which utilizes proteases to cleave one or a limited number of peptide bonds in a protein, relies on the fact that mobile and flexible regions within a protein are more sensitive to protease digestion than organized protein structures, and thus, the structural and dynamic features of proteins can be defined [7, 8].

Various methods have been used to analyze the products of limited proteolysis. Routinely, after the target protein has been digested with sequence-specific or non-specific proteases, the resulting products are normally analyzed by gel electrophoresis to examine

the fragment profile or to identify any band pattern changes of the protein derived from pretreatment of the protein under different conditions [9–11]. To obtain further information on the N- and C-terminal boundaries of proteolytic products, the bands can be analyzed by Western blotting [4] or by N-terminal sequencing [5] to acquire amino acid sequences of the products. The C-terminus of the product is usually derived when the definitive N-terminal sequence of the digested sequence is known and is combined with the molecular weight of the product estimated by gel analysis or measured by mass spectrometry. In a few cases, the C-terminus of the products have been determined directly by laborious and complicated C-terminal sequencing [12–14].

Mass spectrometry is a powerful tool for gaining molecular weight information on biological samples having molecular weights greater than 30 kDa and is frequently used in conjunction with limited proteolysis to determine the molecular weight of the proteolyzed products [11, 15–24]. However, mass measurement alone does not produce a determination of the sequence of the proteolyzed products, since multiple subsequences usually will be retrieved fitting a given experimental mass, within experimental error. Thus, it is very difficult to determine the sequence of the product by mass spectrometry alone.

Published online November 18, 2004

Address reprint requests to Dr. J. G. Stroh, PGRD-Groton Laboratories, Pfizer Inc., Groton, CT 06340, USA. E-mail: Justin_g_stroh@groton.pfizer.com

* Readers can also contact Julie (Zhi) Xie, Protein & Peptide Chemistry, EMS Discovery, Groton Laboratories, Pfizer, Inc., MS 8118W-112 Eastern Point Road, Groton, CT 06340. E-mail: zhi_xie@groton.pfizer.com.

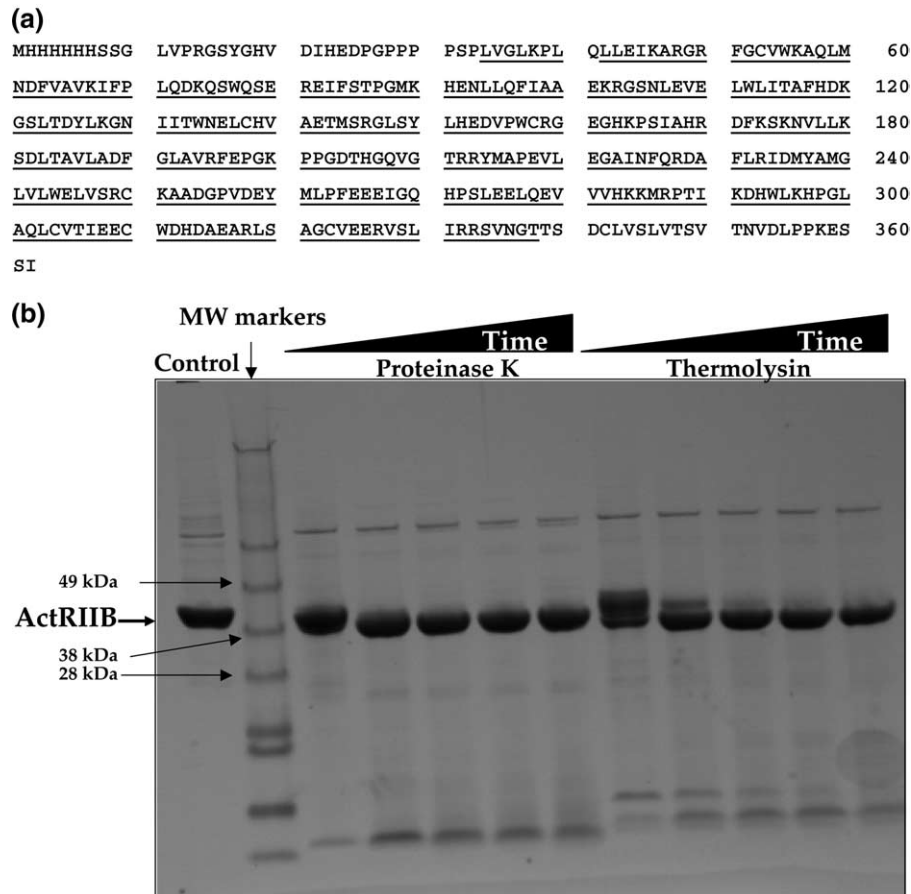


Figure 1. (a) Sequence of ActRIIBc. Underlined sequence (L34-T338) is the core structure as determined by limited proteolysis. (b) Gel of limited digestion of ActRIIBc. Lane 1 = undigested starting material. Lane 2 = molecular weight markers. Lane 3 = 5 min digestion with proteinase K. Lane 4 = 30 min digestion with proteinase K. Lane 5 = 60 min digestion with proteinase K. Lane 6 = 90 min digestion with proteinase K. Lane 7 = 120 min digestion with proteinase K. Lane 8 = 5 min digestion with thermolysin. Lane 9 = 30 min digestion with thermolysin. Lane 10 = 60 min digestion with thermolysin. Lane 11 = 90 min digestion with thermolysin. Lane 12 = 120 min digestion with thermolysin.

To counteract this limitation, specific endoproteases [15–21, 24–28] or exopeptidases [12–14], which only cut at specific amino acids or from the N- or C-terminus of the protein, are routinely used in combination with mass spectrometry to narrow the number of possibilities for a given observed molecular weight. However, the effort to define the structural domain can be compromised by the lack of the amino acid(s) required for the cleavage at the boundaries and the limited action of the exopeptidases at the N- or C-terminus. To determine the limited proteolysis products by non-specific proteases, additional sequence information derived either from multi-step traditional methods or sub-sequences defined with specific proteases treatment [22] is often required.

Therefore, the current analysis of limited proteolysis products is of restricted versatility. Even though mass spectrometry provides accurate and sensitive mass measurement, determination of the sequence of the proteolyzed products still routinely requires the tradi-

tional time-consuming analysis steps and also demands significant amounts of sample. Here, we describe a new method using mass spectrometry as the only analytical tool for determination of N- and C-terminal boundary sequences of limited proteolysis products by non-specific enzymes.

Experimental

Limited Proteolysis

A His-tagged ActRIIB construct (MW 40,772 Da) containing the catalytic domain [29] of the kinase was purified by NiNTA affinity column. The purified protein with a concentration of 1 mg/ml was incubated with a series of specific or non-specific proteases in a digestion buffer containing 20 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM DTT, and 10% glycerol at 37 °C. The enzyme to substrate ratio (wt/wt) was 1:500 for trypsin (Roche, Indianapolis, IN), 1:2000 for proteinase

Table 1. Theoretical sub-sequences of ActRIIBc that fit the observed molecular weight from a 30-min trypsin digest

Mass	Dev	From-To	Sequence
39083	0.1	6-353	HH HSSGLVPRGSYGHVDIHED....SVNGTTSDCLVSLVTSVTNV DL
39081	-2.0	3-349	MH HHHHSSGLVPRGSYGHVDI....LIRRSVNGTTSDCLVSLVTS VT
39079	-3.9	15-362	PR <GSYGHVDIHEDPGPPPPSPL....LVSLVTSVTNVDLPPKESS>

Column 1 is the theoretical mass, Column 2 is the difference between the observed and calculated mass, Column 3 is the subsequence position, and Column 4 is the actual sub-sequence. In Column 4, the first two and last two (red highlighted amino acids) are the two N-terminal and C-terminal amino acids, respectively, that are beyond the sub-sequence.

K (Roche), and 1:50 for thermolysin (Sigma, St. Louis, MO). To the thermolysin digest, 5 mM CaCl₂ was supplemented for the protease to function. 50 μ L aliquots were removed from the 500 μ L digestion samples at 5, 30, 60, 90, and 120 min. Digestion was terminated by the addition of PMSF for trypsin and proteinase K, and with 10 mM EDTA for thermolysin.

Liquid Chromatography/Mass Spectrometry Analysis

From each of the 50 μ L aliquot of digested material, a 5 μ L sample was extracted that was diluted with 25 μ L of 0.1% TFA in water to make a total of 30 μ L of solution. Each 30 μ L sample was placed in a 96 well plate that was inserted into a Famos (LC Packings, San Francisco, CA) autosampler. Samples were chilled to 5 °C until injection. Then, a 10 μ L aliquot of each sample was injected on-column to a Vydac C-18 capillary column (0.3 \times 50 mm). An Applied Biosystems (Foster City, CA) 140D solvent delivery module was used to deliver an HPLC gradient of 10-80% B over 30 min (A = 0.02% TFA in 95% water/5% acetonitrile, B = acetonitrile) at a flow rate of 10 μ L/min. Samples eluting the column flowed through an Applied Biosystems single wavelength UV detector operating at 220 nm. The UV detector was fitted with an LC Packings (San Francisco, CA) U-Z View flow cell. Samples exiting the UV detector entered a Micromass (Altrincham, UK) Ultima triple quadrupole mass spectrometer employing electrospray ionization and operating in the single quadrupole mode. The instrument was operated in the positive ion mode using a capillary voltage of 3.31 kV, a cone voltage of 110 V, a source temperature of 80 °C, a desolvation temperature of 120 °C, a cone gas flow of nitrogen equal to 123 L/h, and a desolvation flow of 710 L/h. The instrument was scanned over the *m/z*

range of 350–2400 at a rate of 5.2 s per scan. Data was collected over a 30-min period. Raw data was deconvoluted to produce molecular weight information of the protein and peptide fragments using MaxEnt, which is the software provided by Micromass for deconvoluting multiply charged data. The data was processed to approximately 30% convergence. Other parameters for MaxEnt were as suggested by the manufacturer.

Data Analysis

A table of molecular weights was prepared for each digestion time point and compared against other time points for the same digestion enzyme. For the molecular weight of each product formed in digestion, a list of possible sub-sequences of the original protein was calculated using the program GPMW, version 4.2 (Lighthouse Data, Odense, Denmark). Subsequences were calculated with a theoretical error of $\pm 0.03\%$. Each set of subsequences was compared with adjacent molecular weight subsequences in order to determine if there was an association between the two molecular weights. Comparison of the data points occurred and the common data point was determined.

N-Terminal Sequencing

Automated Edman sequencing was performed on an Applied Biosystems Procise Sequencer, Model 494. Data were collected and analyzed using the Applied Biosystems Model 610A software.

Results and Discussion

We used the catalytic domain of the activin receptor type II B (ActRIIBc), a Thr/Ser kinase, as an example to demonstrate this method. The construct lacked the first

Table 2. Theoretical sub-sequences of ActRIIBc that fit the observed molecular weight from a 30-min thermolysin digest

Mass	Dev	From-To	Sequence
35066	0.5	16-325	RG SYGHVDIHEDPGPPPPSPLV....TIEECWDHDAEARLSAGCVE ER
35066	0.6	40-348	KP LQLEIKARGRFGCVWKAQL....SLIRRSVNGTTSDCLVSLV SV
35062	-3.4	33-347	LK PLQLEIKARGRFGCVWKAQ....VSLIRRSVNGTTSDCLVSLV TS
35069	3.7	28-336	PG PPPPSPLVGLKPLQLEIKA....ARLSAGCVEERVSLIRRSVN GT
35061	-4.4	35-343	PL VGLKPLQLEIKARGRFGCV....VEERVSLIRRSVNGTTSDCLV VS
35061	-4.4	36-344	LV GLKPLQLEIKARGRFGCV....EERVSLIRRSVNGTTSDCLV SL
35061	-4.4	34-342	SP LVGLKPLQLEIKARGRFGC....CVEERVSLIRRSVNGTTSDCV LV
35070	4.5	7-316	HH HSSGLVPRGSYGHVDIHEDP....HPGLAQLCVTIEECWDHDAE AR

Explanation of the column of this table is the same as given in Table 1.

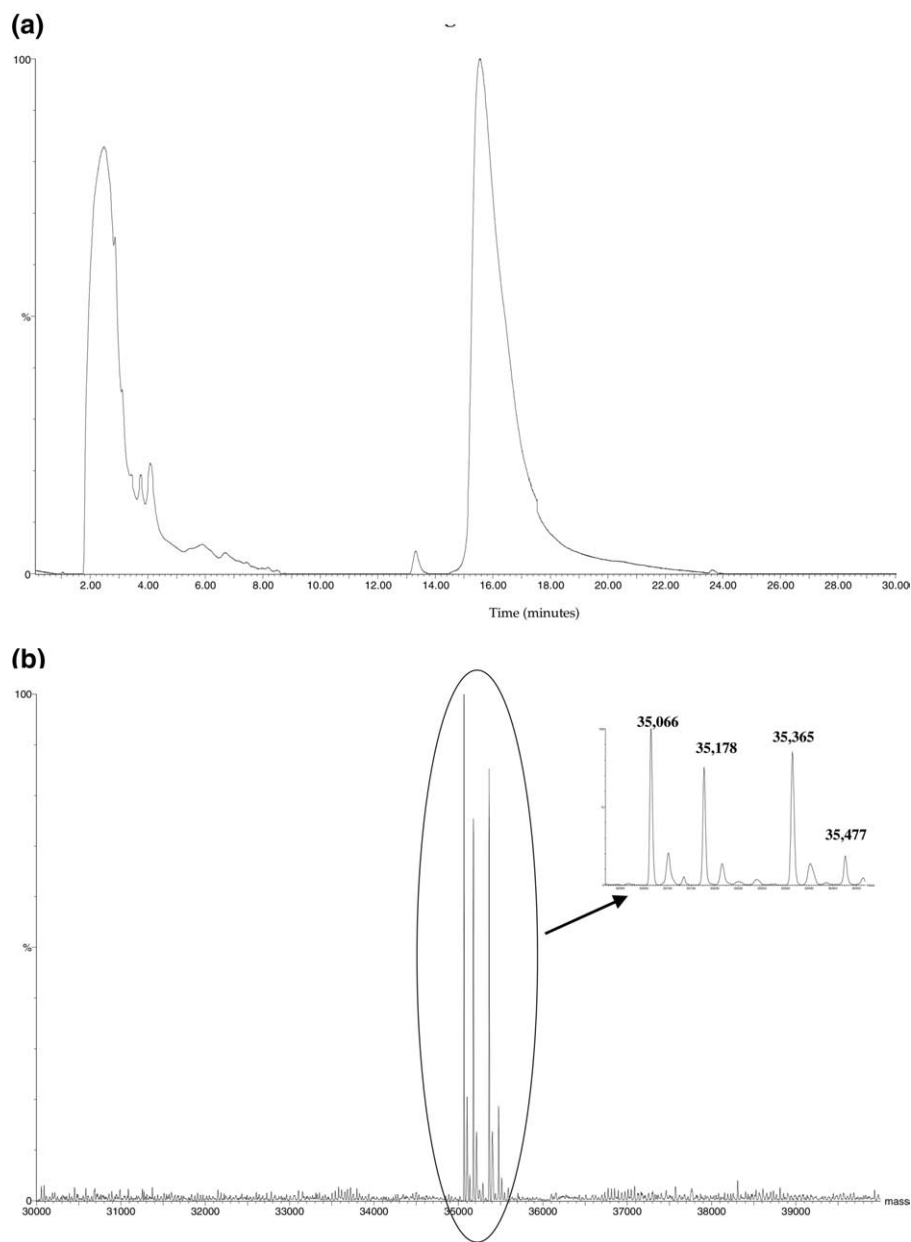


Figure 2. (a) HPLC/UV of 30-min thermolysin digest. (b) Deconvoluted electrospray mass spectrum of 30-min thermolysin digest. Taken from HPLC/UV peak in Figure 2a at a retention time of 16 min.

166 amino acids of the enzyme and was His-tagged at the N-terminus. The molecular weight of the ActRIIBc (167–512) was 40,775 Da (average) when measured by LC-MS and matched the theoretical mass (40,772 Da) of the construct with the addition of acetylation at the N-terminal amino acid. The position of the amino acids of ActRIIBc in the following text was defined as showed in Figure 1a. Limited proteolysis of the protein was performed with trypsin and two non-specific enzymes, thermolysin and proteinase K. Aliquots of the reaction mixture were extracted at different times (5 to 120 min) and the products were analyzed by SDS-PAGE (Figure 1b) and electrospray ionization LC/MS as described in the Experimental Methods. Cleavage with trypsin gave

one peak in the high mass range at $39,082.8 \pm 7.8$ Da (data not shown). Three possible subsequences (Table 1) were calculated that matched the observed mass within experimental error. Since trypsin cuts only at the C-terminus of Lys/Arg, the only subsequence to match this requirement was G15-I362, which resulted from cleavage at R14 and contains the C-terminus of the protein. Thus, selective action of specific proteases helps to determine the correct subsequence from multiple possibilities and this is one of the major reasons that trypsin has been widely used in conjunction with mass spectrometry to analyze sequences derived from limited proteolysis.

On the other hand, non-selective enzymes such as

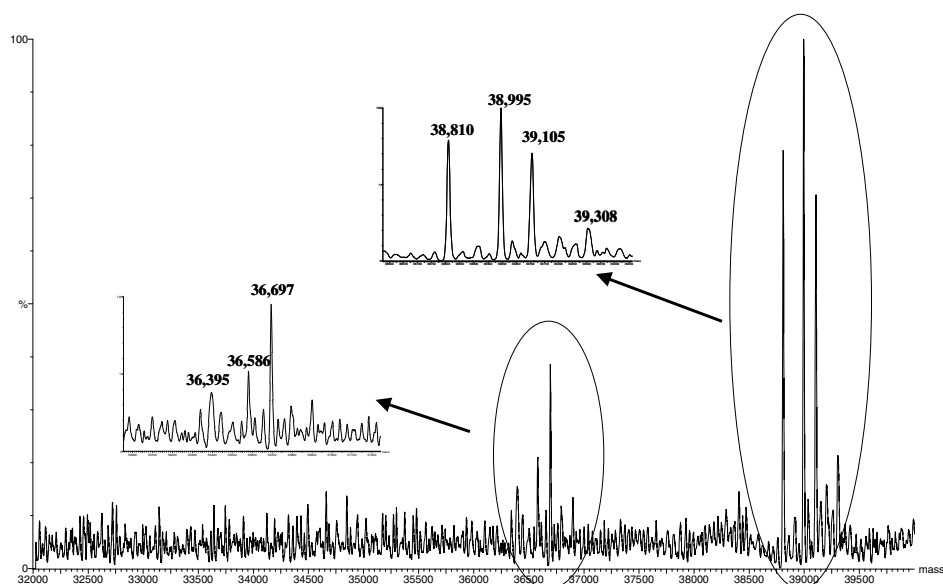


Figure 3. Deconvoluted electrospray mass spectrum of 5-min proteinase K digest of ActRIIBc.

thermolysin and proteinase K are less successful when used only in conjunction with mass spectrometry, since they can digest a much broader range of amino acids. For example, after the ActRIIBc was digested with thermolysin for 30 min, a mass of 35,066 Da was observed using LC-MS analysis of the proteolysis products. Even when the mass spectral experimental error was reduced to 0.015% (or ± 5.3 Da), a total of eight possible subsequences were retrieved based on the observed mass within the experimental error (Table 2). Furthermore, three of the eight possibilities have the same exact mass (35,061 Da), as they are isomers of each other (V35-L343, G36-V344, and L34-C342). Without additional sequence information, it is not possible to determine which of the eight possibilities is the correct one. Therefore, to determine the sequence of proteolytic products of non-specific proteases, additional steps, such as N-terminal or C-terminal sequencing, or comparison with other known subsequences, are normally required. However, we demonstrate here a method that uses mass spectrometry as the only measurement device, and from mass spectral measurements the particular subsequence can be unambiguously identified derived from proteolytic cleavage by non-specific enzymes.

The fundamental premise of this method is the recognition that the cleavage of proteins by non-specific enzymes (such as thermolysin and proteinase K) is not random, but patterned. As an example, the HPLC/UV of a fast gradient LC/MS derived from a 30-min digest of ActRIIBc with thermolysin is shown in Figure 2a. The protein peak with a retention time of 16 min gave the deconvoluted (molecular weight) electrospray mass spectrum shown in Figure 2b. Deconvoluted molecular weights were processed over the mass range 30,000–41,000 Da. It appeared that this is not a spectrum of

random molecular weights; on the contrary, all of the molecular weight peaks are congregated in a narrow mass range, suggesting that these mass species are related. This is again demonstrated in the deconvoluted mass spectrum of the 5-min proteinase K digestion of ActRIIBc (Figure 3), which shows two different congregations of peaks.

The question is how these molecular weight peaks relate to each other. As described above, it is not possible, based on a single mass measurement to determine which of the eight possibilities is the correct sequence for the observed mass of 35,066 Da (Table 2). However, when the possible subsequences matching all of the measured masses in a congregation of peaks are analyzed simultaneously, a pattern emerges showing the cleavage footprint of the protease at a sensitive site. As an example, the masses shown in Figure 2b (35,066 Da, 35,178 Da, 35,365 Da, and 35,477 Da) produce four sets of data of potential subsequences as shown in Figure 4. A common association running through the data sets was identified as the sequential cleavage using non-specific proteases (Figure 4). Polypeptides L34-C342, L34-343L, L34-345S, and L34-L346L were recognized from each of the data sets that showed the footprint of thermolysin cleavage. Through this analysis, the consensus sequence was identified from multiple data sets (underlined in Figure 4). A table of all cleavage products at various digestion time points can also be determined to demonstrate the progression of digestion by thermolysin (Table 3). A similar strategy had been successfully applied to determine the N- and C-terminal boundary sequences of products digested by proteinase K at various time-points as shown Figure 5a and b and Table 4.

In addition, the 90-min digests of ActRIIBc with thermolysin and proteinase K were also analyzed by

Observed mass: 35066		5.3 Da		Sequence	
Mass (Da)	From-	To			
35066	16-	325	RG	SYGHVDIHED....	EARLSAGCVE ER
35066	40-	348	KP	LQLLEIKARG....	TSDCLVSLVT SV
35062	39-	347	LK	PLQLLEIKAR....	TTSDCLVSLV TS
35061	34-	342	SP	LVGLKPLQLL....	RSVNGTTSDC LV
35061	35-	343	PL	VGLKPLQLE....	VNGTTSDC LV
35061	36-	344	LV	GLKPLQLE....	VNGTTSDC LV
35069	28-	336	PG	PPPPSPLVGL....	RVSLIRRSVN GT
35070	7-	316	HH	HSSGLVPRGS....	IEECWDHDAE AR

Observed mass: 35178		5.3 Da		Sequence	
Mass (Da)	From-	To			
35176	1-	310		MHHHHHSSG....	AQLCVTIEEC WD
35181	49-	359	AR	GRFGCVWKAQ....	VTNVDLPPKE SS
35175	34-	343	SP	LVGLKPLQLL....	SVNGTTSDC LV
35182	46-	356	EI	KARGRFGCVW....	VTSVTNVDLP PK

Observed mass: 35365		5.3 Da		Sequence	
Mass (Da)	From-	To			
35364	17-	328	GS	YGHVDIHEDP....	LSAGCVEERV SL
35362	1-	311		MHHHHHSSG....	QLCVTIEECW DH
35361	35-	346	PL	VGLKPLQLE....	GTTSDCLVSL VT
35361	36-	347	LV	GLKPLQLE....	TTSDCLVSLV TS
35361	34-	345	SP	LVGLKPLQLL....	NGTTSDC LV
35361	11-	323	SG	LVPRGSYGHV....	DAEARLSAGC VE
35361	8-	320	HH	SSGLVPRGSY....	WDHDAEARLS AG

Observed mass: 35477		5.3 Da		Sequence	
Mass (Da)	From-	To			
35477	1-	312		MHHHHHSSG....	LCVTIEECWD HD
35476	20-	332	GH	VDIHEDPGPP....	CVEERVSLIR RS
35476	12-	325	GL	VPRGSYGHVD....	EARLSAGCVE ER
35478	38-	350	GL	KPLQLEIKA....	DCLVSLVTSV TN
35474	34-	346	SP	LVGLKPLQLL....	GTTSDCLVSL VT
35473	27-	340	DP	GPPPPSPLVGL....	IRRSVNGTTS DC
35482	4-	316	HH	HHHSSGLVPR....	IEECWDHDAE AR

Figure 4. Calculation of all theoretical sub-sequences from ActRIIBc for each individual mass shown in Figure 2b, within experimental error.

conventional SDS-PAGE followed by transferring to a PVDF membrane and N-terminal sequencing. The amino acid sequence LVGLK was identified at the N-terminal of the proteolytic fragment that resulted from the thermolysin digestion, while amino acid sequences IHEDP and EDPGP were identified as the N-terminal amino acids for the proteinase K digestion products (data not shown). These findings were consistent with the results obtained using the method described in this paper that proteolytic fragments containing L34 as the N-terminus were present in the 90-min thermolysin digests (Table 3) and fragments containing I22 and E24 as the N-terminal amino acids were present in the 90-min proteinase K digests (Table 4). Thus, without traditional gel electrophoresis and N-terminal sequencing, we are able to determine the sequence of

Table 3. Thermolysin digest data

5 min	30 min	60 min	90 min	120 min	Range
35,063 (4)	35,066 (1)	35,066 (1)	35,067 (1)	35,068 (1)	34–342
35,176 (2)	35,178 (3)	35,179 (2)	35,178 (2)	35,178 (2)	34–343
35,363 (1)	35,365 (2)	35,365 (3)	35,365 (3)	35,368 (3)	34–345
35,474	35,477 (4)	35,477 (4)			34–346
37,577					11–343
37,764					11–345
38,812					2–345
38,994 (3)					1–345
39,107					1–346

Column 1 = observed digest molecular weights after 5-minutes of digestion. Column 2 = as in 1 after 30-minutes of digestion. Column 3 = 60-min, 4 = 90 min, and 5 = 120 min. Column 6 is the sequences range identified for each observed molecular weight in the first five columns. Numbers in parentheses correspond to relative peak height. (1) = highest peak, (2) = second highest peak, etc.

(a)

Observed mass: 36395		± 5.5 Da		Sequence	
Mass (Da)	From-	To			
36394	20-	341	GH	VDIHEDPGPP....	RRSVNGTTSDC CL
36396	22-	343	VD	IHEDPGPPP....	SVNGTTSDC LV
36398	21-	342	HV	DIHEDPGPPP....	RSVNGTTSDC LV
36391	27-	349	DP	GPPPPSPLVGL....	SDCLVSLVTS VT

Observed mass: 36586		± 5.5 Da		Sequence	
Mass (Da)	From-	To			
36587	9-	332	HS	SGLVPRGSYG....	CVEERVSLIR RS
36587	32-	355	PP	SPLVGLKPLQ....	LVTSVTNVDL PP
36587	26-	350	ED	PGPPPPSPLV....	DCLVSLVTSV TN
36585	37-	359	VG	LKPLQLEIK....	VTNVDLPPKE SS
36588	1-	323		MHHHHHSSG....	DAEARLSAGC VE
36588	18-	341	SY	GHVDIHEDP....	RRSVNGTTSDC CL
36590	14-	336	VP	RGSYGHVDIH....	RVSLIRRSVN GT
36582	22-	345	VD	IHEDPGPPP....	NGTTSDC LV
36582	23-	346	DI	HEDPGPPPP....	GTTSDCLVSL VT
36591	27-	351	DP	GPPPPSPLVGL....	CLVSLVTSV NV

Observed mass: 36697		± 5.5 Da		Sequence	
Mass (Da)	From-	To			
36697	21-	345	HV	DIHEDPGPPP....	NGTTSDC LV
36695	22-	346	VD	IHEDPGPPP....	GTTSDCLVSL VT
36694	33-	357	PS	PLVGLKPLQ....	TSVTNVDLPP KE
36693	15-	339	PR	GSYGHVDIHE....	LIRRSVNGTT SD
36702	25-	350	HE	DPGPPPPSPL....	DCLVSLVTSV TN

(b)

Observed mass: 38810		± 6.2 Da		Sequence	
Mass (Da)	From-	To			
38809	8-	353	HH	SSGLVPRGSY....	VSLVTSVTNV DL
38807	5-	349	HH	HHHSSGLVPR....	SDCLVSLVTS VT
38806	11-	355	SG	LVPRGSYGHV....	LVTSVTNVDL PP
38805	1-	343		MHHHHHSSG....	SVNGTTSDC LV

Observed mass: 38995		± 6.2 Da		Sequence	
Mass (Da)	From-	To			
38994	3-	348	MH	HHHHSSGLV....	TSDCLVSLVT SV
38991	1-	345		MHHHHHSSG....	NGTTSDC LV
39000	11-	357	SG	LVPRGSYGHV....	TSVTNVDLPP KE

Observed mass: 39105		± 6.3 Da		Sequence	
Mass (Da)	From-	To			
39104	1-	346		MHHHHHSSG....	GTTSDCLVSL VT

Observed mass: 39308		± 6.3 Da		Sequence	
Mass (Da)	From-	To			
39311	6-	355	HH	HHSSGLVPRG....	LVTSVTNVDL PP
39304	1-	348		MHHHHHSSG....	TSDCLVSLVT SV
39314	10-	359	SS	GLVPRGSYGH....	VTNVDLPPKE SS

Figure 5. (a) and (b) Calculation of all theoretical sub-sequences from ActRIIBc for each individual mass shown in Figure 3, within experimental error.

products of limited proteolysis by non-specific enzymes. Since non-specific proteases do not require specific amino acids for digestion, they are likely to be

Table 4. Proteinase K digest data

5 min	30 min	60 min	90 min	120 min	Range
		35,634 (4)	35,632 (4)	35,632 (3)	24–338
	35,879 (4)	35,881 (3)	35,882 (1)	35,881 (1)	22–338
	36,069 (3)	36,070 (2)	36,070 (2)	36,071 (2)	22–340
36,395	36,400 (1)	36,401 (1)	36,401 (3)	36,401 (4)	22–343
36,586	36,586 (2)				22–345
36,697 (4)					22–346
38,810 (2)					1–343
38,995 (1)					1–345
39,105 (3)					1–346
39,308					1–348

Explanation of columns identical to that of Table 3.

more continuous and efficient in cleaving flexible regions of a protein compared with specific proteases that may be hindered by the absence of the specific amino acid in an accessible region of a protein. Thus, this rapid LC-MS analysis of limited proteolysis products by non-specific proteases provides a more effective tool to probe and determine the domain boundaries of proteins.

The accuracy of the mass determination is critical, since the number of potential polypeptide sequences that match a given measured mass grows proportionally with experimental error. A large experimental error makes the identification of the common association between data sets difficult, because it compromises our ability to determine the single correct sequence from a list of many polypeptides matching the measured mass.

Because this method employs LC-MS, it is a sensitive tool for the analysis of minor transient proteolytic intermediates and is useful for monitoring the progression of proteolysis. Table 3 shows the proteolytic products that are identified during the time course of digestion of ActRIIBc with thermolysin. Multiple N- and C-terminally truncated fragments are identified after 5-min of incubation of ActRIIBc with thermolysin. Some products are transient and are further cleaved (e.g., H2-S345 and L11-S345), while others are relatively stable even when treated for up to 2 h (e.g., L34-C342 and L34-L343). Combining the digestion data of thermolysin and proteinase K (Figures 4 and 5), we determine that the polypeptide containing L34-T338 of ActRIIBc is relatively resistant to proteases digestion and composes a compact core structure of the kinase. The ability to analyze and identify transient intermediates and relatively stable polypeptides during the course of digestion allows us to track the footprints of non-specific proteases, define domain boundaries, and gain more insight to the dynamic features of the protein.

Another important feature of this method is robustness. With traditional methods, the probability of obtaining sequence information of the proteolytic products is often compromised by the resolution and sensitivity of SDS-PAGE and N-terminal sequencing. It often requires an optimization process to either resolve

digestion products or identify a set of digestion conditions that result in only a very few cleavage products accumulating for successful N-terminal sequencing. As shown in Figure 4, at least four mass species were detected using LC-MS, though the digestion products were barely separated by traditional gel electrophoresis, which made downstream N-terminal sequencing analysis challenging. However, with the sensitivity and accuracy of the current method, sequences of the proteolytic products can be successfully determined during the course of digestion (Figures 4, 5 and Tables 3, 4) even if the products are transient and within a narrow mass range. Although an LC/MS run time of 30 min/sample was used in the current set of experiments, we have found that ballistic gradients employing a 10-min LC/MS run time (data not shown) are quite sufficient for this analysis. With optimization, the speed can be further improved.

Thus, this method provides a rapid, simple, accurate, and robust analysis for determining the sequences of limited proteolysis products using non-specific proteases. Because of the sensitivity of LC-MS, only a very small quantity of protein is required for analysis. The process could be automated in a high throughput format and be applied to define domain boundaries of proteins for construct design for structure studies and to probe detailed conformational changes of proteins when exposed to various treatments, such as activation or suppression.

Conclusions

We developed a rapid and accurate LC-MS method for the analysis and determination of product sequences of limited proteolysis using non-specific enzymes. The sequences of the proteolysis products at a given time point can be determined based solely on mass measurement and the progressive action of a non-specific protease. This finding eliminates the dependence on the traditional methods of gel electrophoresis and N-terminal sequencing, which often compromise the accuracy and sensitivity empowered by mass spectrometry.

References

- Todd, A. E.; Orenge, C. A.; Thornton, J. M. Evolution of Protein Function, from a Structural Perspective. *Curr. Opin. Chem. Biol.* **1999**, 3(5), 548–556.
- Kumar, S.; Ma, B.; Tsai, C. J.; Sinha, N.; Nussinov, R. Folding and Binding Cascades: Dynamic Landscapes and Population Shifts. *Prot. Sci.* **2000**, 9(1), 10–19.
- Bennett, W. S.; Huber, R. Structural and Functional Aspects of Domain Motions in Proteins. *CRC Crit. Rev. Biochem.* **1984**, 15(4), 291–384.
- Oren, D. A.; Li, Y.; Volovik, Y.; Morris, T. S.; Dharia, C.; Das, K.; Galperina, O.; Gentz, R.; Arnold, E. Structural Basis of BLYS Receptor Recognition. *Nat. Struct. Biol.* **2002**, 9(4), 288–292.
- Monti, M.; Principe, S.; Giorgetti, S.; Mangione, P.; Merlini, G.; Clark, A.; Bellotti, V.; Amoresano, A.; Pucci, P. Topological Investigation of Amyloid Fibrils Obtained from β 2-Microglobulin. *Prot. Sci.* **2002**, 11(10), 2362–2369.
- Yang, S. A.; Klee, C. Study of Calcineurin Structure by Limited Proteolysis. *Methods Mol. Biol.* **2002**, 172, 317–334.
- Polverino de Laureto, P.; De Filippis, V.; Di Bello, M.; Zambonin, M.; Fontana, A. Probing the Molten Globule State of α -Lactalbumin by Limited Proteolysis. *Biochemistry* **1995**, 34(39), 12596–12604.
- Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zama, M.; Zambonin, M. Correlation Between Sites of Limited Proteolysis and Segmental Mobility in Thermolysin. *Biochemistry* **1986**, 25(8), 1847–1851.
- Zhang, B.; Peng, Z. Y. Structural Consequences of Tumor-Derived Mutations in P16INK4a Probed by Limited Proteolysis. *Biochemistry* **2002**, 41(20), 6293–6302.
- Henn, A.; Shi, S. P.; Zarivach, R.; Ben-Zeev, E.; Sagi, I. The RNA Helicase DbpA Exhibits a Markedly Different Conformation in the ADP-Bound State When Compared with the ATP- or RNA-Bound States. *J. Biol. Chem.* **2002**, 277(48), 46559–4665.
- Chen, G. Q.; Sun, Y.; Jin, R.; Gouaux, E. Probing the Ligand Binding Domain of the GluR2 Receptor by Proteolysis and Deletion Mutagenesis Defines Domain Boundaries and Yields a Crystalizable Construct. *Prot. Sci.* **1998**, 7(12), 2623–2630.
- Borowski, P.; Kuehl, R.; Mueller, O.; Hwang, L. H.; Schulze zur Wiesch, J.; Schmitz, H. Biochemical Properties of a Minimal Functional Domain with ATP-Binding Activity of the NTPase/Helicase of Hepatitis C Virus. *Eur. J. Biochem.* **1999**, 266(3), 715–723.
- Stoffer, B.; Frandsen, T. P.; Busk, P. K.; Schneider, P.; Svendsen, I.; Svensson, B. Production, Purification, and Characterization of the Catalytic Domain of Glucoamylase from *Aspergillus niger*. Part 1. *Biochem. J.* **1993**, 292, 197–202.
- Villanueva, J.; Villegas, V.; Querol, E.; Aviles, F. X.; Serrano, L. Protein Secondary Structure and Stability Determined by Combining Exoproteolysis and Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **2002**, 37(9), 974–984.
- Leite, J. F.; Amoscato, A. A.; Cascio, M. Coupled Proteolytic and Mass Spectrometry Studies Indicate a Novel Topology for the Glycine Receptor. *J. Biol. Chem.* **2000**, 275(18), 13683–13689.
- Calabresi, L.; Tedeschi, G.; Treu, C.; Ronchi, S.; Galbiati, D.; Airolidi, S.; Sirtori, C. R.; Marcel, Y.; Franceschini, G. Limited Proteolysis of a Disulfide-Linked ApoA-I Dimer in Reconstituted HDL. *J. Lipid Res.* **2001**, 42(6), 935–942.
- Bothner, B.; Dong, X. F.; Bibbs, L.; Johnson, J. E.; Siuzdak, G. Evidence of Viral Capsid Dynamics Using Limited Proteolysis and Mass Spectrometry. *J. Biol. Chem.* **1998**, 273(2), 673–676.
- Aghajanian, S.; Hovsepian, M.; Geoghegan, K. F.; Chrunyk, B. A.; Engel, P. C. A Thermally Sensitive Loop in Clostridial Glutamate Dehydrogenase Detected by Limited Proteolysis. *J. Biol. Chem.* **2003**, 278(2), 1067–1074.
- Shields, S. J.; Oyeyemi, O.; Lightstone, F. C.; Balhorn, R. Mass Spectrometry and Non-Covalent Protein-Ligand Complexes: Confirmation of Binding Sites and Changes in Tertiary Structure. *J. Am. Soc. Mass Spectrom.* **2003**, 14(5), 460–470.
- Schlosser, A.; Klockow, B.; Manstein, D. J.; Lehmann, W. D. Analysis of Post-Translational Modification and Characterization of the Domain Structure of Dynamin a from Dictyostelium Discoideum. *J. Mass Spectrom.* **2003**, 38(3), 277–282.
- D'Ambrosio, C.; Talamo, F.; Vitale, R. M.; Amodeo, P.; Tell, G.; Ferrara, L.; Scaloni, A. Probing the Dimeric Structure of Porcine Aminoacylase 1 by Mass Spectrometric and Modeling Procedures. *Biochemistry* **2003**, 42(15), 4430–4443.
- Seielstad, D. A.; Carlson, K. E.; Kushner, P. J.; Greene, G. L.; Katzenellenbogen, J. A. Analysis of the Structural Core of the Human Estrogen Receptor Ligand Binding Domain by Selective Proteolysis/Mass Spectrometric Analysis. *Biochemistry* **1995**, 34(39), 12605–12615.
- Rudiger, A.; Rudiger, M.; Weber, K.; Schomburg, D. Characterization of Post-Translational Modifications of Brain Tubulin by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry: Direct One-Step Analysis of a Limited Subtilisin Digest. *Anal. Biochem.* **1995**, 224(2), 532–537.
- Hamuro, Y.; Burns, L.; Canaves, J.; Hoffman, R.; Taylor, S.; Woods, V. Domain Organization of D-AKAP-2 Revealed by Enhanced Deuterium Exchange-Mass Spectrometry (DXMS). *J. Mol. Biol.* **2002**, 321(4), 703–714.
- Gervasoni, P.; Staudenmann, W.; James, P.; Pluckthun, A. Identification of the Binding Surface on β -Lactamase for GroEL by Limited Proteolysis and MALDI-Mass Spectrometry. *Biochemistry* **1998**, 37(33), 11660–11669.
- Bronsoms, S.; Villanueva, J.; Canals, F.; Querol, E.; Aviles, F. X. Analysis of the Effect of Potato Carboxypeptidase Inhibitor Pro-Sequence on the Folding of the Mature Protein. *Eur. J. Biochem.* **2003**, 270(17), 3641–3650.
- Cohen, S. L.; Ferre-D'Amare, A. R.; Burley, S. K.; Chait, B. T. Probing the Solution Structure of the DNA-Binding Protein Max by a Combination of Proteolysis and Mass Spectrometry. *Prot. Sci.* **1995**, 4(6), 1088–1099.
- Vinci, F.; Couprie, J.; Pucci, P.; Quemeneur, E.; Moutiez, M. Description of the Topographical Changes Associated to the Different Stages of the DSBA Catalytic Cycle. *Prot. Sci.* **2002**, 11(7), 1600–1612.
- Persson, U.; Souchelnytskyi, S.; Franzen, P.; Miyazono, K.; ten Dijke, P.; Heldin, C. H. Transforming Growth Factor (TGF- β)-Specific Signaling by Chimeric TGF- β Type II Receptor with Intracellular Domain of Activin Type IIb Receptor. *J. Biol. Chem.* **1997**, 272(34), 21187–21194.