Chemical Cleavage at Aspartyl Residues for Protein Identification

ARTICLE *in* ANALYTICAL CHEMISTRY · DECEMBER 2001

Impact Factor: 5.64 · DOI: 10.1021/ac010619z · Source: PubMed

6 AUTHORS, INCLUDING:

Raymond C Sowder
U.S. Department of Health and Human Services
84 PUBLICATIONS **5,469** CITATIONS
SEE PROFILE

David J. Garfinkel
University of Georgia
86 PUBLICATIONS **6,917** CITATIONS
SEE PROFILE

Robert J Fisher
National Institutes of Health
133 PUBLICATIONS **4,697** CITATIONS
SEE PROFILE

Anal. Chem. **2001,** *73,* 5395−5402

Chemical Cleavage at Aspartyl Residues for Protein Identification

Aiqun Li,*,§ Raymond C. Sowder, II,‡ Louis E. Henderson,‡ Sharon P. Moore,† David J. Garfinkel,† and Robert J. Fisher§

Protein Chemistry Laboratory, AIDS Vaccine Program, SAIC Frederick, and Gene Regulation and Chromosome Biology Laboratory, SAIC Frederick, NCI at Frederick, P.O. Box B, Frederick, Maryland 21702-1201

An alternative method to enzymatic digestion for protein identification by mass spectrometry has been developed that is based on chemical cleavage by formic acid. This method was tested on gel-purified apomyoglobin and BSA, as well as unknown proteins that cofractionate with Ty1-virus-like particles from *Saccharomyces cerevisiae*. Cleavage at aspartyl residues was found to be efficient and specific, and this specificity of cleavage lent itself easily to database searches. Parallel digestions using trypsin were also performed. The formic acid cleavage method generated comparable or better results than tryptic digestion for protein identification.

Peptide mass mapping of proteins separated by polyacrylamide gel electrophoresis with enzymatic degradation has become a routine technique for protein characterization.[1] The most commonly used protease is trypsin because of its well-defined specificity and the appropriate size of tryptic peptides for mass spectrometric analysis. A few other commercially available proteases such as Lys-C, Glu-C, and Asp-N are also occasionally applied; however, enzymatic digestions are time-consuming and discriminate against certain substrates. Trypsin, for example, is not efficient for hydrophobic or very basic proteins. Although one-fourth of sequence coverage may be sufficient for identification of a medium-sized protein, higher coverage is necessary for characterization of larger proteins and posttranslationally modified proteins.[2] In addition, protease autolysis products sometimes interfere with spectrum interpretation, although well-documented signature peptides from trypsin may serve as internal calibrants. Furthermore, some buffers required for efficient and specific proteolysis may generate chemical noise, thus requiring additional purification before mass spectrometric analysis.

Recently, several different approaches have been reported that pertain to efficient cleavage of proteins. Bark et al. developed a high-temperature proteolytic digestion method using thermolysin.[3] High temperature increases the flexibility of the protein structure, which results in better proteolytic susceptibility; however, the lack of specificity of thermolysin is potentially problematic. Gobom et al. suggested vapor-phase acid hydrolysis using pentafluoropropionic acid (PFPA),[4] but three different types of cleavages, including sequence ladders, were observed, increasing spectrum complexity. Additionally, some special devices are required to carry out this method.

We have developed a strategy based on chemical cleavage at aspartyl residues using dilute formic acid. The principle behind this strategy was described in 1967 by Schultz, and the procedure was optimized in the late 1970s by Inglis.[5,6] Aspartyl peptide bonds are found to be preferentially cleaved under the pH condition just sufficient to maintain the undissociated β-carboxyl group. The mechanism for hydrolysis at aspartyl residues is proposed to involve an anhydride[5] or a cyclic imide[7] intermediate formed between the β-carboxyl group and the amide group at either terminal through elimination of one water molecule (Figure 1). Although the Asp-X bond is most readily cleaved, the X-Asp may also be cleaved**.** Therefore, the resulting peptides may keep at least one aspartyl residue at the point of cleavage, or they may lose the aspartyl residue by a double cleavage event. Because the β-carboxyl group of aspartic acid acts as a proton donor, it is necessary to maintain the pH of the hydrolyzing medium below the pK_a of the β-carboxyl group. Inglis reported that pH 2.0, 108 °C, and 2 h of incubation are optimal conditions for in-solution hydrolysis.[6] Either dilute HCl or formic acid can be used to maintain the correct pH of the medium.[8]

Interestingly, a search of the human genome protein file from the NCBI database revealed that the abundance of aspartic acid is 4.9%, and lysine and arginine have an abundance of 5.7% and 5.6%, respectively. The abundance of aspartic acid compared with that of lysine and arginine suggests that cleavage specificity for aspartyl bonds may be useful for protein sequencing. Indeed, our results demonstrate that chemical cleavage at aspartyl residues by formic acid is an efficient alternative to enzymatic digestion for peptide mass mapping.

* To whom correspondence should be addressed. SAIC Frederick, NCI at Frederick, Protein Chemistry Laboratory, Bldg. 469, Rm. 240, Frederick, MD 21702-1201. Phone: 301-846-6745. Fax: 301-846-7269. E-mail: liai@ncifcrf.gov.

§ Protein Chemistry Laboratory.

‡ AIDS Vaccine Program.

† Gene Regulation and Chromosome Biology Laboratory.

(1) Patterson, S. D. *Electrophoresis* **1998**, *19*, 883−1054.

(2) Celis, J. E. *Electrophoresis* **1995**, *12*, 2177−2240.

(3) Bark, S. J.; Muster, N.; Yates, J. R.; Siuzdak, G. *J. Am. Chem. Soc.* **2001**, *123*, 1774−1775.

(4) Gobom, J.; Mirgorodskaya, E.; Nordhoff, E.; Hojrup, P.; Roepstorff, P. *Anal. Chem*. **1999**, *71*, 919−927.

(5) Schiltz, J. *Methods Enzymol.* **1967**, *11*, 255−263.

(6) Ingris, A. S. *Methods Enzymol.* **1983**, *91*, 324−332.

(7) Xie, M.; Velde, D. V.; Morton, M.; Borchardt, R. T.; Schowen, R. L. *J. Am. Chem. Soc.* **1996**, *118*, 8955−8956.

(8) Smith, B. R. In *The Protein Protocols Handbook*; Walker, J. M., Ed.; Humana Press: Totowa, NJ, 1996; 381−384.

10.1021/ac010619z CCC: $20.00
Published on Web 10/06/2001

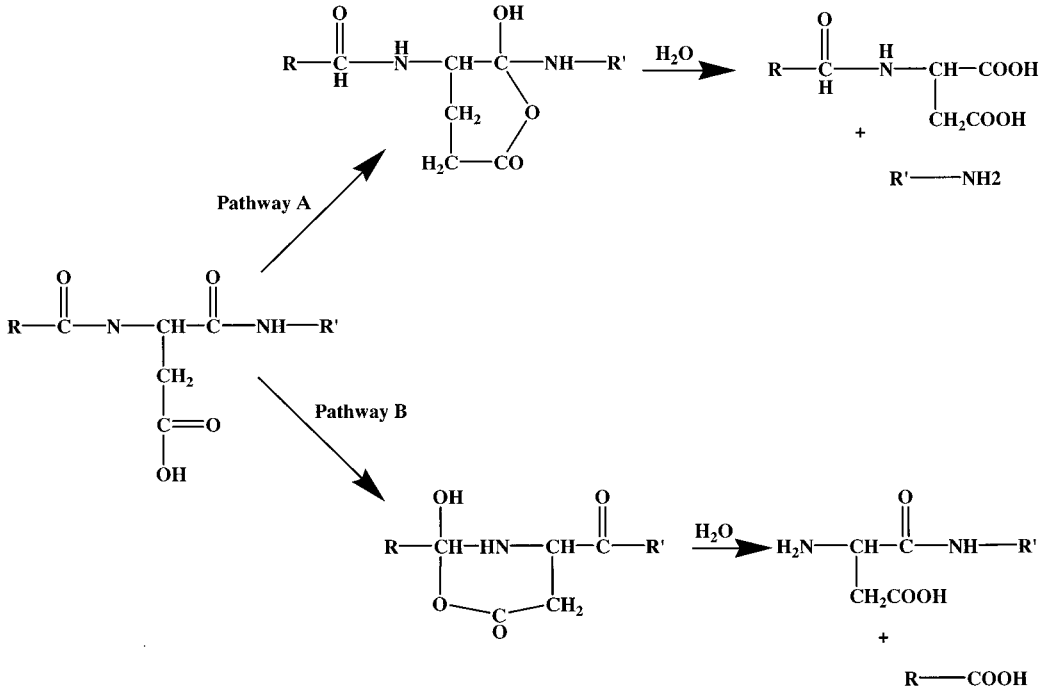


Figure 1. Mechanisms of the acid cleavage at (A) Asp-X and (B) X-Asp (adapted from ref 8).

Table 1. Assignment of Observed Peptide Fragments of Apomyoglobin after Formic Acid Cleavage at Different Incubation Time and Temperature

			obs *m*/*z*, 108 °C				
sequence assignment	calc *m*/*z*	peak	5 min	30 min	60 min	120 min	240 min
1−20 GLS**D**GEWQQVLNVWGKVEA**D**	2230.07	15		2230.09	2230.01	2230.23	
1−19 GLS**D**GEWQQVLNVWGKVEA	2115.05	14		2115.19	2115.11	2115.15	
4−19 (**D**)GEWQQVLNVWGKVEA(**D**)/5−20	1857.91	10		1857.96	1857.88	1857.87	1857.98
5−19 GEWQQVLNVWGKVEA	1742.89	7			1742.84	1742.90	1742.87
4−43 (**D**)GEWQQVLNVWGKVEA**D**IAGHGQEVLIRLFT-GHPETLEKF(**D**)/5−44	4549.1 (ave)	23				4548.8	
20−43 (**D**)IAGHGQEVLIRLFTGHPETLEKF(**D**)/21−44	2707.42	17			2707.49	2707.26	2707.52
21−43 IAGHGQEVLIRLFTGHPETLEKF	2592.39	16				2592.85	2592.55
20−59 (**D**)IAGHGQEVLIRLFTGHPETLEKF**D**KFKHLKTE-AEMKASE(**D**)/22−60	4583.2 (ave)	24		4583.1	4583.1	4582.5	4583.3
21−59 IAGHGQEVLIRLFTGHPETLEKF**D**KFKHLKTEA-EMKASE	4468.2 (ave)	22					4468.2
44−59 (**D**)KFKHLKTEAEMKASE(**D**)/45−60	1891.96	11		1891.95	1891.92	1891.93	1892.07
45−59 KFKHLKTEAEMKASE	1776.93	8					1776.96
60−108 (**D**)LKKHGTVVLTALGGILKKKGHHEAELKPLAQ-SHATKHKIPIKYLEFIS(**D**)/61−109	5426.5 (ave)	27			5426.1	5426.8	5426.6
61−108LKKHGTVVLTALGGILKKKGHHEAELKPLAQS-HATKHKIPIKYLEFIS	5311.4 (ave)	26					5312.2
110−153 AIIHVLHSKHPG**D**FGA**D**AQGAMTKALELFRN**D**-IAAKYKELGFQG	4768.5 (ave)	25	4766.7	4768.1	4768.4	4768.6	4768.1
110−140 AIIHVLHSKHPG**D**FGA**D**AQGAMTKALELFRN	3346.9 (ave)	19		3346.6	3346.8	3345.8	3345.5
109−140 (**D**)AIIHVLHSKHPG**D**FGA**D**AQGAMTKALELFRN(**D**)/110−141	3461.9 (ave)	21			3461.9		
109−125 (**D**)AIIHVLHSKHPG**D**FGA(**D**)/110−126	1813.93	9		1813.94	1813.87	1813.92	1813.88
110−125 AIIHVLHSKHPG**D**FGA	1698.91	6		1699.02	1698.91	1698.90	1698.89
109−121 (**D**)AIIHVLHSKHPG(**D**)/110−122	1423.78	3		1423.77	1423.82	1423.79	1423.78
110−121 AIIHVLHSKHPG	1308.76	1			1308.74	1308.77	1308.75
123−153 FGA**D**AQGAMTKALELFRN**D**IAAKYKELGFQG	3362.9 (ave)	20	3361.8	3360.6	3362.7	3362.6	3363.5
122−140 (**D**)FGA**D**AQGAMTKALELFRN(**D**)/123−141	2055.0	13		2055.08	2055.0	2055.03	2055.05
123−140 FGA**D**AQGAMTKALELFRN	1939.97	12				1940.8	1940.93
127−153 AQGAMTKALELFRN**D**IAAKYKELGFQG	2970.55	18	2971.8 (ave)	2970.53	2970.53	2970.54	2970.53
126−140 (**D**)AQGAMTKALELFRN(**D**)/127−141	1664.84	5		1664.95	1664.82	1664.84	1664.84
127−140 AQGAMTKALELFRN	1549.82	4			1549.85	1549.81	1549.88
142−153 IAAKYKELGFQG	1324.73	2	1324.73	1324.73	1324.72	1324.73	1324.73
sequence coverage (%)			28.8	68.0	100	100	98

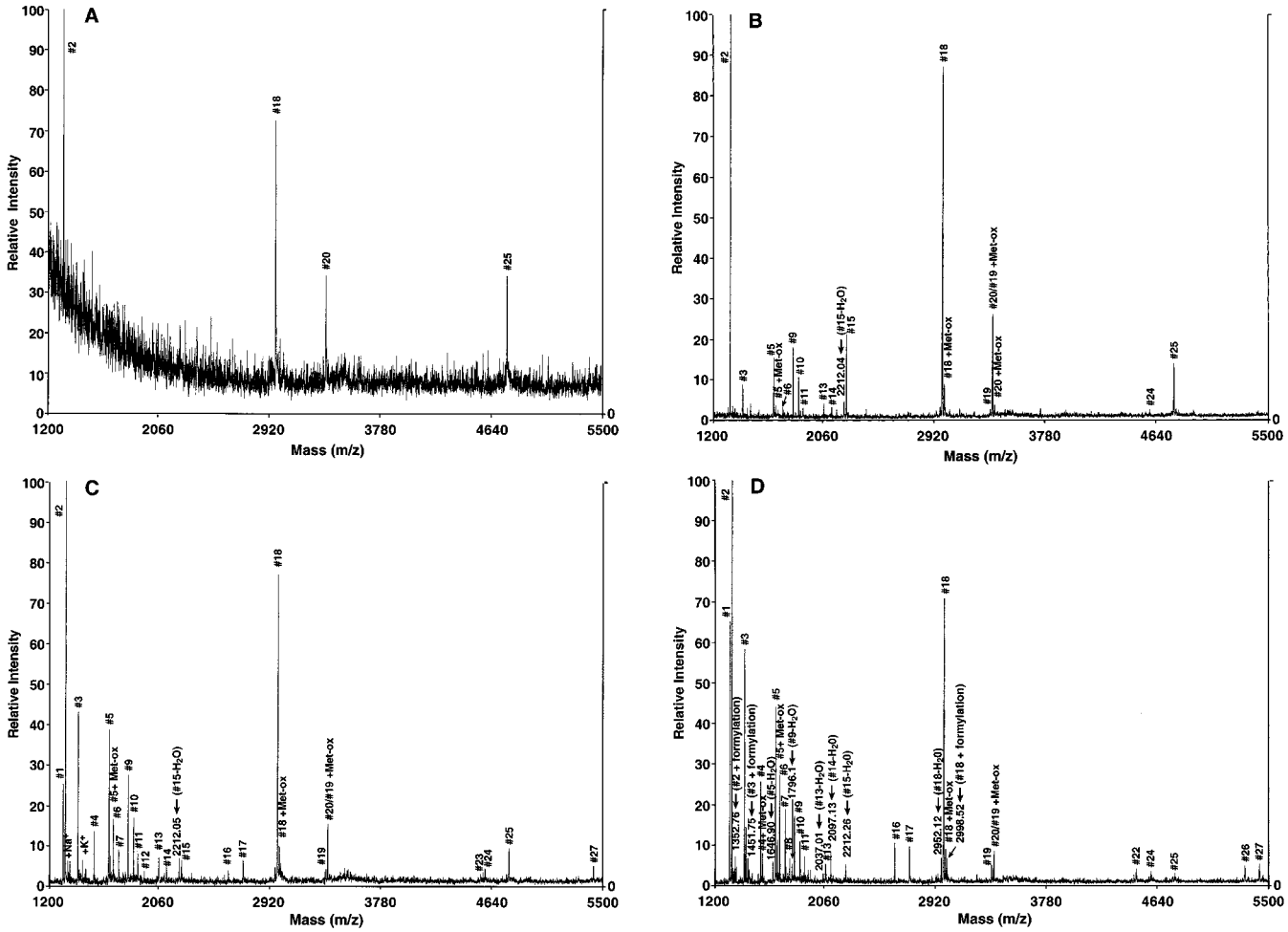


Figure 2. MALDI-TOF mass spectra of formic acid cleavage products of horse apomyoglobin: (A) 5 min, 108 °C; (B) 30 min, 108 °C; (C) 120 min, 108 °C; and (D) 240 min, 108 °C. Because no cleavage products were observed below *m*/*z* 1200, the spectrum display range was truncated at *m*/*z* 1200 for clarity.

EXPERIMENTAL SECTION

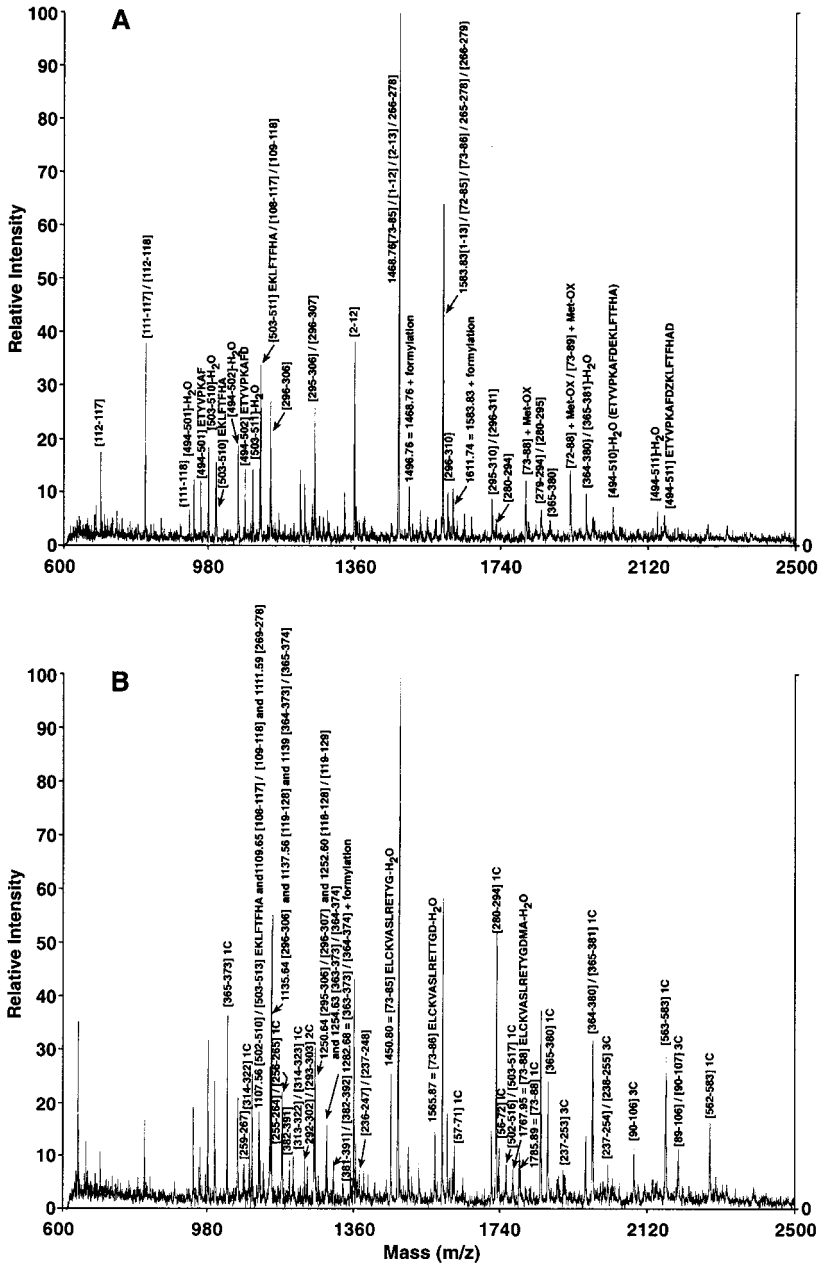
Materials and Reagents. Apomyoglobin (horse heart), bovine serum albumin, ammonium bicarbonate, and 2,5-dihydroxybenzoic acid (DHB) were purchased from Sigma (St. Louis, MO). Dithiothreitol (DTT) was purchased from Research Organics, Inc. (Cleveland, OH). Formic acid (GR), HPLC-grade water, methanol, and acetonitrile were obtained from EM Science (Gibbstown, NJ). Trifluoroacetic acid (99%) was obtained from Aldrich (Milwaukee, WI). Trypsin (modified, sequence grade) and Asp-N were obtained from Roche Diagnostics (Indianapolis, IN).

Gel Electrophoresis and Destaining. Apomyoglobin and bovine serum albumin (60−100 pmol) were dissolved in NOVEX tricine SDS sample buffer (InVitrogen, Carlsbad, CA) and subjected to SDS-PAGE at 90 V for 2 h using a 16% NOVEX tricine gel. The proteins were visualized by Coomassie Blue staining. The protein bands were excised and destained using a 25 mM NH_4HCO_3 solution in 50% methanol. The clear gel was washed extensively with HPLC water and dehydrated in acetonitrile, then dried under vacuum. Additional protein samples were from an ongoing protein identification project involving proteins that copurify with Ty1-virus-like particles (Ty1-VLPs) from *Saccharomyces cerevisiae*. These unknown proteins were purified using a 10% NOVEX Tris−glycine gel (125 V for 1 h 45 min) using the process described above.

Enzymatic In-Gel Digestion. The clean, dry gel band was rehydrated in a minimal volume of freshly prepared trypsin at a concentration of 125 ng/μL dissolved in 50 mM NH_4HCO_3. Then the gel was ground into small pieces using a plastic stirring tube (BioSpec Products, Bartlesville, OK), and 50 mM NH_4HCO_3 was added to cover the gel pieces. The digestion was carried out at 37 °C for 2 h (apomyoglobin) or overnight (BSA). To extract the peptides, 30 μL of 50% acetonitrile in 5% trifluoroacetic acid was added. The tube was vigorously mixed for 5 min and centrifuged, and the supernatant was carefully removed using a gel-loading pipet tip. This step was repeated with an additional 30 μL of 50% acetonitrile in 5% trifluoroacetic acid. The extracted peptides were pooled and dried in a SpeedVac. A 10 μL portion of 50% acetonitrile in water was added to dissolve the peptides for mass spectrometric analysis.

In-gel digestion of apomyoglobin with the endoprotease Asp-N was performed similarly. The concentration of Asp-N in the digestion buffer (50 mM NH_4HCO_3) was ~20 ng/μL, and the sample was incubated for 2 h at 37 °C.

In-Gel Cleavage Using Formic Acid. Instead of using Pyrex glass hydrolysis tubes[8] sealed by a blowtorch under vacuum, the cleavage reaction was carried out inside a microfuge tube. Formic acid was chosen as the hydrolysis medium, because it is a mild



DTHKSEIAHR FKDLGEEHFK GLVLIAFSQY LQQCPFDEHV KLVNELTEFA

KTCVADESHA GCEKSLHTLF GDELCKVASL RETYGDMADC CEKQEPERNE

202

CFLSHKDDSP DLPKLKPDPN TLCDEFKADE KKFWGKYLYE IARRHPYFYA

5346.1

PELLYYANKY NGVFQECCQA E DKGACLLPK IETMREKVLT SSARQRLRCA

7365.7

SIQKFGERAL KAWSVARLSQ KFPKAEFVEV TKLVTD**LTKV HKECCHGDLL**

ECADDRADLA KYICDNQDTI SSKLKECCDK PLLEKSHCIA EVEKDAIPEN

LPPLTADFAE DKDVCKNYQE AKDAFLGSFL YEYSRRHPEY AVSVLLRLAK

EYEATLEECC AKDDPHACYS TVFDKLKHLV DEPQNLIKQN CDQFEKLGEY

GFQNALIVRY TRKVPQVSTP TLVEVSRSLG KVGTRCCTKP ESERMPCTED

6564.6

YLSLILNRLC VLHEKTPVSE KVTKCCTESL VNRRPCFSAL TPD**ETYVPKA**

4878.8

FDEKLFTFHA DICTLPDTEK QIKKQTALVE LLKHKPKATE EQLKTVMENF

VAFVDKCCAA DDKEACFAVE GPKLVVSTQT ALA

Figure 3. MALDI-TOF mass spectra of formic acid cleavage products of bovine serum albumin: (A) without DTT, (B) with DTT, and (C) peptide mass mapping from B. The bold letters represent the covered sequence, and the underlined letters represent peptides unrecovered, the *m*/*z* values of which are listed underneath. In B, for simplicity and clarity, only the new peaks due to inclusion of DTT are labeled.

acid and a good buffer in comparison to hydrochloric acid. In addition, formic acid is an excellent solvent and denaturant for proteins, which offers potential for cleavage of insoluble proteins. Approximately 200 μL of freshly prepared 2% formic acid was added to the dry protein gel band in a microfuge tube. (DTT (3 mg) may be added to the mixture to improve sequence coverage by breaking disulfide bonds between peptides.) After complete rehydration of the gel, the mixture was incubated at 108 °C in a Thermolyne Dri-Bath. The incubation time was from 5 min to 4 h for apomyoglobin and 2 h for all other samples. For apomyoglobin, room temperature was also tested for an incubation time of 22 h. The sample was then cooled, and the supernatant was transferred to a fresh microfuge tube and dried. A 10 μL portion of 50% acetonitrile in water was added to dissolve the peptides for mass spectrometric analysis.

MALDI-MS Analysis. A MALDI-TOF mass spectrometer with delayed extraction (Voyager-DE PRO, Perseptive Biosystems, Framingham, MA) was used to acquire the mass spectra. The matrix solution was a 2-fold dilution of saturated 2,5-dihydroxybenzoic acid in 50% acetonitrile in water. Aliquots of 0.5 μL of the peptide mixture and 0.5 μL of the matrix solution were mixed on the sample plate and air-dried prior to analysis. All spectra were obtained in reflectron mode and averaged over 150 laser shots. Masses were calibrated internally with peptides from trypsin autolysis or with calculated cleavage peptide products.

Protein Identification. A database search was performed against NCBInr.02.06.2001 using a local copy of ProteinProspector v.3.2.1. The enzyme specificity under Digest was modified following instructions in the user's manual (http://prospector.ucsf.edu/ucsfhtml3.2/instruct/allman.htm#Enzyme) so that formic acid cleavage at either Asp-X or X-Asp or both was reflected in the MS-Fit program. The allowed mass tolerance was 100 ppm for MS-Fit.

RESULTS AND DISCUSSION

Apomyoglobin. Apomyoglobin from horse heart was used as a test protein. The cleavage products were monitored by mass spectrometry at different incubation times from 5 min to 4 h. The results are summarized in Table 1. Cleavage occurred as early as 5 min at 108 °C, resulting in 29% sequence coverage; incubation at room temperature overnight did not yield any cleavage product. Complete sequence recovery was achieved after 1 h of incubation. The mass spectra are shown in Figure 2. Prolonged reaction times resulted in less missed cleavage, although not necessarily better sequence coverage, probably because of the difficulty associated with short peptide recovery. For example, the peak at m/z 1742.9 (peak 7), corresponding to a double cleavage product at both the C-terminal of Asp^4 and the N-terminal of Asp^{20}, was below the noise level of the spectrum after 30 min of incubation (Figure 2B) but became evident after longer incubation times (Figure 2C,D). Other similar observations are summarized in Table 1.

The spectra were very clean and showed impressive cleavage specificity at aspartyl residues. In the first 2 h of incubation, 28 peaks with satisfying signal-to-noise ratio were observed, 27 of which can be assigned to expected cleavage products at aspartyl residues. The remaining peak was at m/z 2212.1, −18 mass units away from m/z 2230.1, the major cleavage product containing two more cleavage sites, GLSDGEWQQVLNVWGKVEAD (see Table 1, peak 15). Therefore, this peak may arise from the cyclic anhydride or imide intermediate before a double cleavage event.[5,7] After prolonged incubation, that is, 4 h, more similar peaks due to elimination of water were observed, for example, peaks at m/z 2037.0 and 2097.1, as labeled in the spectrum (Figure 2D). In addition, formylation[9] adducts of major cleavage products were also observed. For example, peaks at m/z 1352.76, 1451.75, and 2998.52 were due to formylation of peaks 2, 3, and 18, respectively (Figure 2D). Although these observed chemistries do not alter cleavage specificity, it is advantageous to limit the incubation time to no more than 2 h for a clean cleavage reaction.

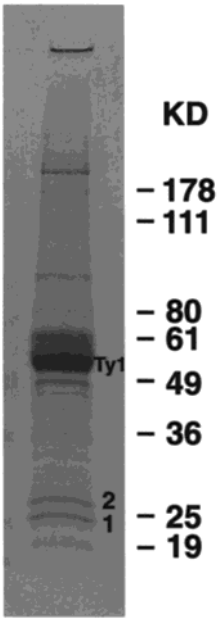


Figure 4. SDS-gel of proteins that cofractionate with Ty1-virus-like particles from *S. cerevisiae*. Bands 1 and 2 were analyzed.

To further demonstrate the efficiency of formic acid cleavage, we performed parallel experiments using tryptic digestion as well as Asp-N digestion for 2 h at 37 °C (data not shown). Tryptic digestion retrieved about 86% of the sequence of apomyoglobin with three small, basic peptides (below m/z 1000) undetected. Asp-N is a commercially available enzyme with cleavage sites involving aspartyl residues and is usually used as a complement for a tryptic digestion. Our experiment with Asp-N resulted in 66% sequence coverage and some Asp-N autolysis products. The unrecovered sequence piece was peptide [60−108], DLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEFIS, which has an average m/z of 5426.5. This peptide was evident after 60 min of incubation with formic acid (Table 1 and Figure 1).

Bovine Serum Albumin. We used bovine serum albumin (BSA) to test the feasibility of formic acid cleavage for average-sized proteins. Of the BSA sequence, 30% was recovered directly from the supernatant after 2 h of formic acid treatment (Figure 3A). Considering that there are as many as 35 cysteine residues in BSA, DTT was included in the reaction mixture, and the sequence coverage improved dramatically to 62% (Figure 3B,3C). Upon careful examination of the spectra, more cysteine-containing peptides were retrieved with the inclusion of DTT in the reaction mixture. This result was consistent with the observation by Gobom

(9) Miller, D. L.; Papayannopoulos, I. A.; Styles, J.; Bobin, S. A.; Lin, Y. Y.; Biemann, K.; Iqbal, K. *Arch. Biochem. Biophy.* **1993**, *301*, 41−52.

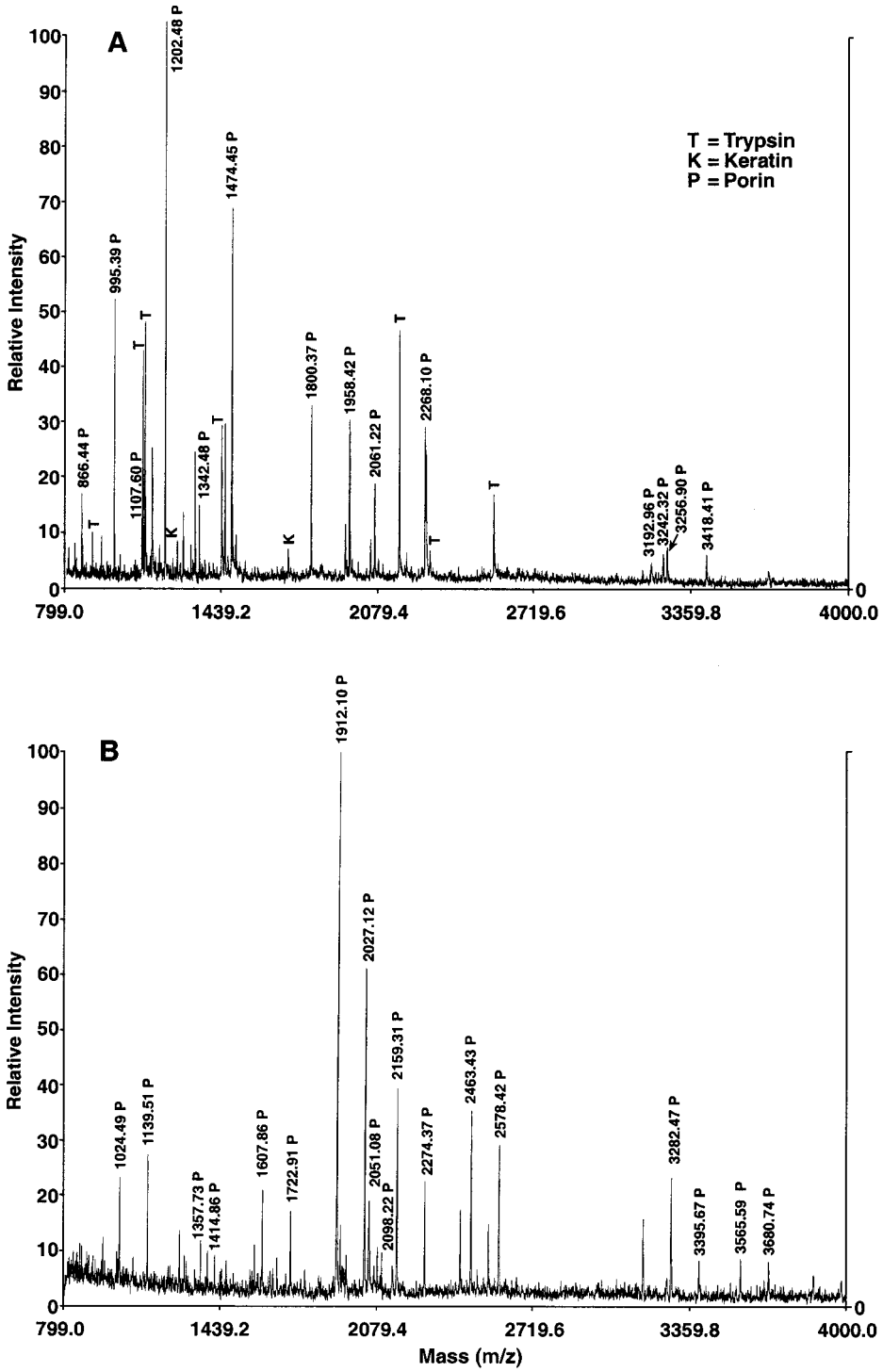


Figure 5. MALDI-TOF mass spectra of (A) trypsin digest and (B) formic acid cleavage products of band 1. Note: In the spectra, a lower mass accuracy than that used in generating Table 2 is allowed for some high-mass peaks; therefore, more peaks are labeled derived from the protein in the spectra than the masses matched indicated in Table 2.

et al. using vapor-phase hydrolysis.[4] In Figure 3C, the uncovered sequences are underlined and include five peptides, [87−88], [130−170], [171−236], [392−450], and [451−493], the masses of which were either too low (below m/z 300) to be detectable or too high (above m/z 4500) for efficient recovery from the cleavage mixture. In fact, when expanding the acquisition mass range to above m/z 4500, we observed three of the four high-mass peptides in linear mode (data not shown). The cleavage of BSA, therefore, was complete and efficient.

Furthermore, the cleavage was very specific at aspartyl residues. Although we observed a few minor peaks due to either formation of pyroglutamic acid at N-terminal glutamic acid,[9] the cyclic intermediate,[5,7] or formylation,[10] these peaks do not compromise the cleavage specificity at aspartyl residues. They can be easily sorted out by subtracting 18 mass units (for pyroglutamic acid and cyclic intermediate) or adding 28 mass units (for

(10) Goodlett, D. R.; Armstrong, F. B.; Creech, R. J.; van Breemen, R. B. *Anal. Biochem.* **1990**, *186*, 116−120.

Table 2. MS-Fit Search Results from Parallel Trypsin Digestion and Formic Acid Cleavage of Gel Band 1 in Figure 4

rank	MOWSE score	no. masses matched (%)	protein MW (Da)/pI	species	accession no.	protein name
			MS-Fit Search Results for Tryptic Digest of Band 1[a]			
1	8.26×10^3	7/31 (22)	30428.6/7.70	*S. cerevisiae*	6324273	(Z71331) ORF YNL055c
2	1.95×10^3	6/31 (19)	30429.6/6.83	*S. cerevisiae*	130685	(X02324) porin (aa 1−283)
3	1.95×10^3	6/31 (19)	30503.8/6.83	*S. cerevisiae*	173166	(M34907) voltage-dependent anion-selective channel (VDAC) protein
			MS-Fit Search Results for Formic Acid Cleavage of Band 1[b]			
1	1.00×10^{12}	15/36 (41)	30428.6/7.70	*S. cerevisiae*	6324273	(Z71331) ORF YNL055c
1	1.00×10^{12}	15/36 (41)	30429.6/6.83	*S. cerevisiae*	130685	(X02324) porin (aa 1−283)
1	1.00×10^{12}	15/36 (41)	30503.8/6.83	*S. cerevisiae*	173166	(M34907) voltage-dependent anion-selective channel (VDAC) protein

[a] NCBInr.02.06.2001. MW 19 000−35 000 Da, 147 830 entries. Full pI range, 616977 entries. Species search, 8767 entries; Combined MW, pI, species searches, 1735 entries. MS-Fit, 11 entries. Considered modifications, oxidation of M. Minimum number peptides to match, 4. Peptide mass tolerance (monoisotopic), 100 ppm. Maximum missed cleavages, 2. Cysteines unmodified. Peptide N terminus, hydrogen. C terminus, free acid. Input number peptide masses, 31. [b] NCBInr.02.06.2001. MW 19000−35000 Da, 147830 entries. Full pI range, 616977 entries. Species search, 8767 entries. Combined MW, pI, species searches, 1735 entries. MS-Fit, 10 entries. Considered modifications, oxidation of M. Minimun number peptides to match, 4. Peptide mass tolerance (monoisotopic), 100 ppm. Maximum missed cleavages, 2. Cysteines unmodified. Peptide N terminus, hydrogen. C terminus, free acid. Input number peptide masses, 36.

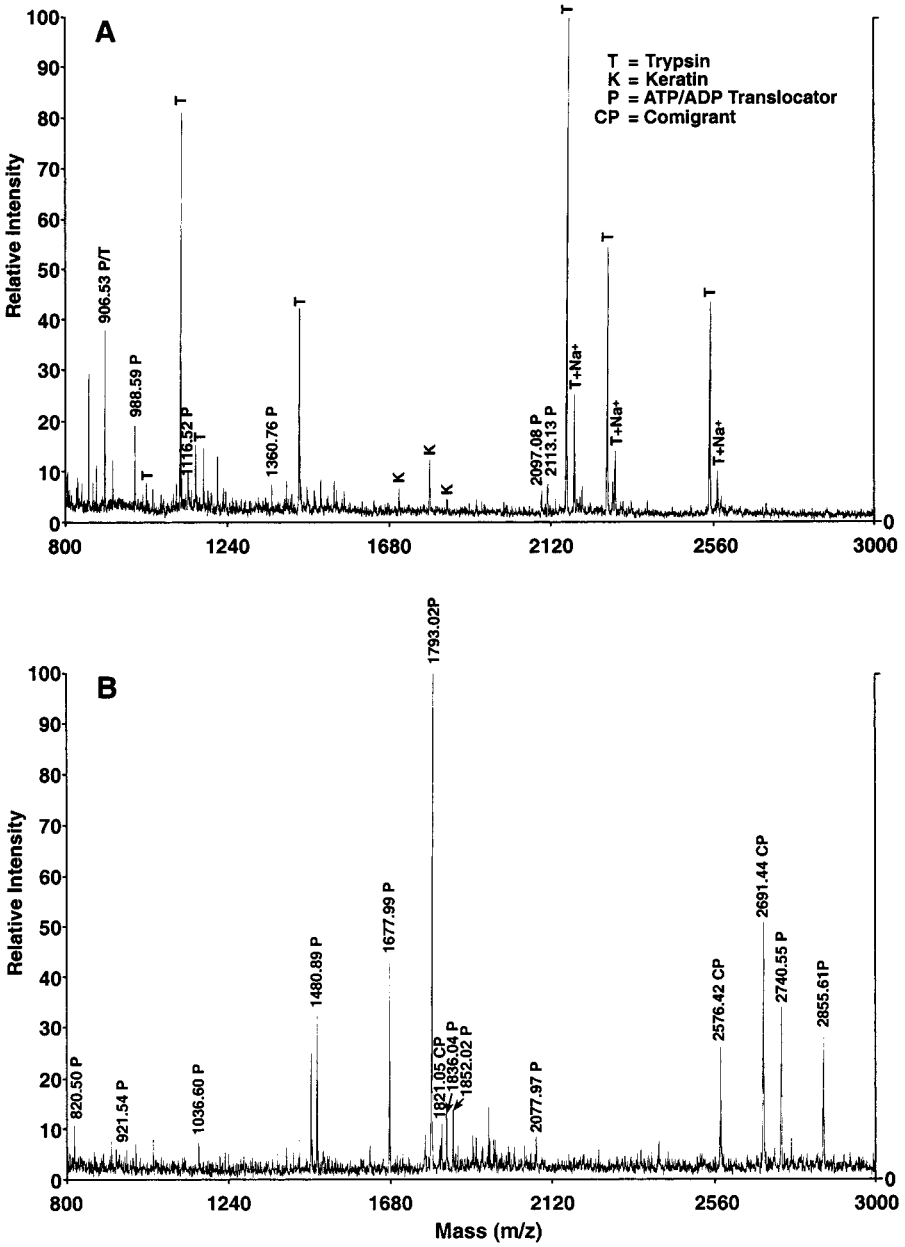


Figure 6. MALDI-TOF mass spectra of (A) trypsin digest and (B) formic acid cleavage products of band 2.

formylation) to the primary specific cleavage products. As a result, more than 95% of the total observed peaks can be attributed to specific cleavage at an aspartyl residue, and 62% of sequence recovery is based only on the primary specific cleavage products.

A parallel experiment with trypsin digestion resulted in 45% sequence coverage. In addition to BSA, a 120 KD protein was also tested using this method, which resulted in 37% sequence coverage, as compared to 18% from trypsin digestion (data not shown). Thus, for large proteins, such as BSA, the overall matched masses and sequence coverage generated from formic acid cleavage should provide reliable protein identification.

Research Samples. We then applied formic acid cleavage to two proteins that copurify with Ty1-virus-like particles from *S. cerevisiae*.[11] The samples were Coomassie Blue-stained gel bands with molecular weights of ~30 Kda (Figure 4). Formic acid cleavage and tryptic digestion were applied to both samples. Figure 5 shows the mass spectra of the cleavage products. For band 1, an MS-Fit search for proteins from formic acid cleavage retrieved the same multiple entries of Por1P variants (accession 6324273, 130685, 173166, NCBInr.0206.2001) as from tryptic digestion, as shown in Table 2, and 53% sequence coverage was obtained from formic acid cleavage. This result indicates that formic acid cleavage provides protein identification as reliably as tryptic digestion.

Interestingly, for band 2, tryptic digestion did not generate an informative mass spectrum (Figure 6A), because the major peaks were from trypsin autolysis. In contrast, formic acid cleavage produced a rich and much cleaner spectrum, as shown in Figure 6B, from which we identified ATP/ADP translocator Pet9P (accession 6319441, NCBInr.0206.2001), an abundant mitochondrial membrane protein, as the major component. The other comigrated component was identified as ORF YGR201C (accession 6321640, NCBInr.0206.2001). With this information, we again checked the spectrum of tryptic digestion and were able to sort out 8 tryptic peptides from the ATP/ADP translocator protein. Most of these peaks were below the noise level of the spectrum, except for two relatively intense peaks at m/z 906.53 and 988.59. Because the m/z value of the former is very close to that of a trypsin autolysis product (m/z 906.51), it is not possible to assign this peak unambiguously to a tryptic peptide from the identified protein. This example demonstrates the advantage of formic acid cleavage in simplifying data interpretation.

In summary, our study has shown that formic acid cleavage at aspartyl residues is a feasible method for protein identification. Because aspartyl residues are more common, it offers the advantage of a much wider range of substrates and nontoxicity than the well-established cynogen bromide cleavage, which only cleaves at the rarer methionyl residue. As an alternative to proteolysis by enzymatic digestion, formic acid cleavage is fast, clean, and specific. No buffer is introduced, which eliminates chemical noise. The one-pot reaction reduces sample handling, thereby decreasing contamination and sample loss. This feature is very amenable for high-throughput analyses. More importantly, because most proteins can be dissolved in formic acid, this cleavage method holds the potential for cleavage of insoluble proteins. We also expect that as a universal digester, acid cleavage will not discriminate against low-abundance proteins. Further studies are underway to explore these possibilities.

This project has been funded in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. CO-56000. We also acknowledge assistance from the Advanced Biomedical Computing Center of SAIC-Frederick for modifying ProteinProspector and calculating the relative abundance of amino acids. In addition, we acknowledge Ms. Lindsey Smith for editorial assistance. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

Received for review June 1, 2001. Accepted September 5, 2001.

AC010619Z

(11) Garfinkel, D. J.; Boeke, J. D.; Fink, G. R. *Cell* **1985**, *4*, 507−517.