RAPID COMMUNICATIONS IN MASS SPECTROMETRY
Rapid Commun. Mass Spectrom. 2008; **22**: 1041–1046
Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/rcm.3467

Specificity of immobilized porcine pepsin in H/D exchange compatible conditions

Yoshitomo Hamuro[1]*, Stephen J. Coales[1], Kathleen S. Molnar[1], Steven J. Tuske[1] and Jeffrey A. Morrow[2]

[1]ExSAR Corporation, 11 Deer Park Drive, Suite 103, Monmouth Junction, NJ 08852, USA
[2]Sierra Analytics Inc., 5815 Stoddard Road, Suite 601, Modesto, CA 95356, USA

Received 20 November 2007; Revised 21 January 2008; Accepted 24 January 2008

Statistical analysis of data from 39 proteins (13 766 amino acid residues) digested with immobilized porcine pepsin under conditions compatible with hydrogen/deuterium (H/D) exchange (<1°C, <30 s) was performed to examine pepsin cleavage specificity. The cleavage of pepsin was most influenced by the amino acid residue at position P1. Phe and Leu are favored residues each with a cleavage probability greater than 40%. His, Lys, Arg, or Pro residues prohibit cleavage when found at the P1 position. Pro also cannot be at position P2 (cleavage probability <0.3%). Occupation of the P3 position by His, Lys, or Arg, or occupation of the P2′ position by Pro, also leads to very little cleavage (cleavage probability <1.7%). The average cleavage probability over the entire data set was 13.6%, which is slightly lower than the value previously obtained by Powers *et al*. (14.8%). This is due, in part, to the larger protein sizes used in the current study. While the specificity of pepsin was similar to that previously observed, higher selectivity was observed in the present study due to less experimental variation in the conditions used to generate our database. Copyright © 2008 John Wiley & Sons, Ltd.

Pepsin is a digestive protease (EC 3.4.23.1) released into the stomach to degrade food proteins into peptides. High-resolution X-ray crystal structures of porcine pepsin have been solved[1] and the specificity of pepsin has been studied extensively using mostly small peptide substrates.[2–4]

Recently, pepsin utilization has been increasing in, among other areas, the fields of bottom-up proteomics[5–7] and hydrogen/deuterium (H/D) exchange.[8–14] In these studies, peptic fragments were identifed by tandem mass spectrometric (MS/MS) fragmentation patterns and sequence information using programs such as SEQUEST and MASCOT.[15] Unlike tryptic digestion, relatively weak specificities are known for pepsin digestion. This promiscuity in pepsin cleavage prevents peptide identifications via cleavage positions alone.

Powers *et al*. performed model peptide studies, as well as statistical analysis of pepsin specificity derived from published data.[16] While those results indicate a preference for aromatic residues at the P1 position as well as other residue-specific trends, cleavage specificity was obscured due, in part, to the diverse experimental conditions used to generate the peptic database. In their statistical study, Powers *et al*. surveyed all of the results published in *Biochemistry*, *Journal of Biological Chemistry*, and *Biochemistry Journal* from 1969–1972, irrespective of the digestion conditions employed.

*Correspondence to: Y. Hamuro, ExSAR Corporation, 11 Deer Park Drive, Suite 103, Monmouth Junction, NJ 08852, USA.
E-mail: yhamuro@exsar.com

Here, we present a statistical analysis of pepsin digestion data accumulated at the laboratory of Virgil Woods at UCSD and our laboratory at ExSAR Corporation. All digestions were performed using an immobilized pepsin column at low temperature (<1°C) and short digestion time (<30 s) in an automated liquid chromatography/mass spectrometry (LC/MS) system. The present results, generated under well-controled H/D exchange compatible conditions, show that pepsin has higher specificity than previously indicated.

EXPERIMENTAL

Digestion of proteins by immobilized pepsin

Each protein sample was diluted to approximately 10 μM. A volume of 20 μL of the diluted protein sample was added to 30 μL of one of the chilled, acidified buffers listed in Table 1. The acidified protein solution was immediately passed over a pepsin column (66 or 104 μL bed volume) at a flow rate of 200 μL/min (carrier solution = 0.05% trifluoroacetic acid (TFA)) at <1°C. Column packing material was made by immobilizing porcine pepsin (Sigma, St. Louis, MO, USA) on Poros 20 AL media (Applied Biosystems, Foster City, CA, USA) at 30 mg/mL by Schiff base chemistry and sodium cyanogen borohydride ($NaCNBH_3$) reduction per the manufacturer's instructions.[17] On average, each protein molecule is in contact with the pepsin column for about 20 or 30 s.

Table 1. List of acidic solutions used and pH

Acidic solution	pH
0.5 M urea, 1 M TCEP	1.3
2 M GuHCl, 1 M TCEP	1.4
8 M urea, 1 M TCEP	2.5
6.4 M GuHCl, 0.8% formic acid	1.9
3.2 M GuHCl, 0.8% formic acid	2.1
1.6 M GuHCl, 0.8% formic acid	2.3
3.2 M GuHCl, 0.8% formic acid	2.5*

*pH adjusted with NaOH.
TCEP: tris(2-carboxyethyl)phosphine hydrochloride; GuHCl: guanidine hydrochloride.

Separation/analysis of peptic fragments

The digested peptic fragments were immediately captured by a reversed-phase trap column. Subsequently, the peptide fragments were eluted from the trap column and separated by a C18 column with a linear gradient of 12% acetonitrile – 38% acetonitrile in 23 min (0.05% TFA in water; flow rate 5–50 μL/min). Mass spectrometric analyses were carried out with a Thermo Finnigan LCQ™ mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) with capillary temperature of 200°C.

Sequence identification of peptic fragments

To quickly identify pepsin-generated peptides for each digestion condition employed, spectral data were acquired in a data-dependent MS/MS mode with dynamic exclusion. SEQUEST (Thermo Fisher Scientific) was used to identify the sequence of each of the dynamically selected precursor ions. This tentative peptide identification was confirmed by visual inspection of the precursor ion charge state identified by SEQUEST for each peptide.[17–19] This procedure generated data that was interpreted as a pepsin digestion map for each protein sample analyzed (e.g., Fig. 1).

Selection of digestion data

For the statistical analysis, the immobilized porcine pepsin digestion data of 39 proteins containing a total of about 13 766 amino acid residues were used. These data comprise the published works from the Woods Laboratory at UCSD (dual specific A-kinase anchoring protein 2 (DAKAP2),[17] protein kinase A regulatory subunit RIIb (PKA RIIb),[18] and protein kinase A regulatory subunit RIa (PKA RIa)[19]), and the data generated at ExSAR Corporation. Both laboratories used very similar automated immobilized pepsin digestion LC/MS systems as described above. Cleavage data for proteins with sequence coverage of less than 90% or with glycosylations were not used.

Analysis of pepsin cleavages

For example, horse cytochrome c has 104 amino acid residues with 103 cleavage sites theoretically available for pepsin cleavage. The LCMS, SEQUEST analysis described above identified thirteen cleavage sites: cleavage after F10, E20, F36, F46, T47, L64, M65, E66, Y67, M80, F82, L94, and A96. There are six Leu residues in the sequence (L32, L35, L64, L68, L94, and L98) and pepsin cleaved after two of them (L64 and L94). Therefore, the probability of Leu being at the P1 position is 33% (= 2/6) in cytochrome c. Similarly, there was only one cleavage before a Leu residue (cleavage between Y67-L68) and thus the probability of Leu being at the P1′ position is 17% (= 1/6). A similar analysis was carried out for all other amino acids in all other protein samples. The probability for each residue being found in sites P2, P3, P4, P2′, P3′, and P4′ during an observed pepsin cleavage was determined as well (Table 2). The observed cleavage at the first three and the last three amide bonds of each protein sample was not included in the analysis, as these sites may give biased information by not filling the entire pepsin active site. Therefore, the average cleavage probability of horse cytochrome c was $13/(103 - 6) = 13.4\%$.

The pepsin specificity for each P1–P1′ amino acid combination was also determined (Table 3). For example, the sequence Tyr-Leu occurs twice in cytochrome c (Y67-L68 and Y97-L98). Pepsin cleaved between Y67-L68 but not between Y97-L98. Therefore, the probability of the pepsin cleavage between Tyr-Leu in cytochrome c is 50%. The probability of pepsin cleavage in all 400 (= 20×20) combinations of P1–P1′ was monitored in all other protein samples.

RESULTS AND DISCUSSION

Digestion of protein samples with immobilized porcine pepsin

A diverse sampling of proteins was digested by immobilized porcine pepsin. First, porcine pepsin was immobilized on aldehyde media by Schiff base chemistry and reduction with $NaCNBH_3$ per the manufacturer's instructions.[17] The media was packed in a column (66 or 104 μL bed volume) and the pepsin column was placed in an automated system which also includes a reversed-phase trap column and a C18 column.[14] Each protein sample was passed over the pepsin column at a flow rate of 200 μL/min and temperature of <1°C, immediately after the addition of acidic buffer with denaturant (Table 1). The duration of contact between each protein molecule and the pepsin column was 20 s for the 66 μL column and 30 s for the 104 μL column.

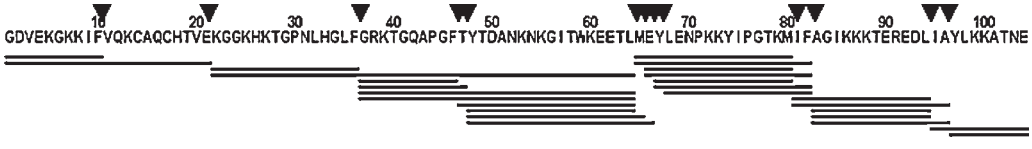

Figure 1. Pepsin digest peptide coverage map of cytochrome c. Bold lines indicate the peptides identified. Triangles indicate the pepsin cleavage sites.

DOI: 10.1002/rcm

The 39 proteins used for this analysis were selected on the basis of posttranslational modifications (PTMs) and sequence coverage completeness: no PTMs except N-terminal acetylations and disulfide bonds were present and sequence coverage was 90% or better by immobilized pepsin digestion. Only the published data from Woods laboratory at UCSD and the data generated at our laboratory (ExSAR Corporation) were used, as these two laboratories use very similar automated LC/MS analysis systems. The proteins with PTMs (except N-terminal acetylations and disulfide bonds) were excluded from this analysis, because the PTMs may influence the pepsin cleavage. Data with poor sequence coverage were also excluded as the results may present a bias towards a lower cleavage probability. The acidifying condition for each protein sample varied slightly (Table 1). For each protein sample, the acidifying buffer was optimized to have the highest sequence coverage and highest resolution. The pH of the acidifying buffer ranged from 1.3 to 2.5.

On average, pepsin cleaved 13.6% of all potential cleavage sites, with a range from 9% to 20% depending on the protein sample. A total of about 13 800 amino acid residues were present in the 39 protein samples studied. The LC/MS analysis after immobilized porcine pepsin digestion identified numerous peptic fragments which were generated by over 1830 pepsin cleavages. The cleavage at the first three and the last three amide bonds of each protein sample were excluded (see Experimental section).

Phe, Leu, and Met are favored at the P1 position

Pepsin cleavage occurs over 40% of the time after Phe or Leu (or when residue Phe or Leu is at position P1, Table 2). Cleavage after Met also occurs more than 30% of the time, although this residue is often not recognized as a favored cleavage site for pepsin (e.g., ExPASy PeptideCutter and http://prowl.rockefeller.edu/recipes/contents.htm). This lack of recognition may be partially due to the relatively low abundance of this amino acid. Pepsin also cleaved better than average after Cys, Glu, Trp, and Tyr residues.

No cleavage after His, Lys, Pro, and Arg

Little or no cleavage was observed after charged residues in acidic conditions (His, Lys, and Arg) or the cyclic imino acid (Pro). It is striking that no pepsin cleavage was observed after Pro or Arg, despite the fact that Pro and Arg appeared in this study 637 times and 737 times, respectively (Table 2). Immobilized pepsin cleaved only once after His in 405 chances and twice in 763 chances after Lys. Immobilized pepsin also cleaved poorly after aliphatic β-branched amino acids (Ile and Val) and after Gly.

Aromatic residues are favored at position P1′

Pepsin cleaves more than 25% of the time when aromatic residue Tyr, Trp, or Phe is at the P1′ position, while the probability of pepsin cleavage before Gly is only 3.2% (Table 2). The influence of the P1′ position is not as strong as that of the P1 position. Aliphatic β-branched residues (Ile and Val) are slightly favored at the P1′ position (better than average 13.6%) unlike the P1 position. Interestingly, positively charged residues (His, Lys, and Arg) and Pro are not as detrimental to pepsin cleavage at the P1′ site as they are at P1.

Pro is forbidden at the P2 position

Pepsin cleaved only once in 637 chances when Pro was at the P2 position (0.2% cleavage). Pro is also disfavored at the P2′ (1.2%) or P3′ position (2.7%). By contrast, other residues at P2, P2′ and P3′ generally have a weak to moderate influence on pepsin cleavage.

His, Lys, and Arg are disfavored at the P3 position

When positively charged residues are at the P3 position, pepsin cleaves less than 1.7% of the time. It is interesting that these positively charged residues have detrimental effects at the P1 or P3 position and little influence when located at the P2 or P1′ position.

The P4 and P4′ positions have little influence

While Pro is the most favored residue and Cys is the least favored residue for the P4 position, the difference in cleavage probability is less than two-fold. Similarly, the difference in cleavage probability is less than two-fold at P4′ with Pro being the most favored residue and Thr being the least favored residue. The present results are in line with the previous results. Antonov *et al*. concluded that pepsin has five subsites (P3–P2′)[4] and Powers *et al*. determined that it has seven subsites (P4–P3′).[16]

Pepsin specificity for positions P1-P1′

Pepsin specificity was also determined for possible combinations of two amino acids at P1–P1′ positions ($400 = 20 \times 20$) in order to investigate any possible synergy on pepsin cleavage for residues at these two positions (Table 3a). All amino acid combinations except Trp-Trp linkage were observed at least once in the database used for this study (Table 3b). The effect on pepsin cleavage of residues at the P1 and P1′ positions is in general found to be additive. When P1 is occupied by a favorable residue such as Phe or Leu and P1′ is occupied by a favorable residue such as Tyr or Phe, there is a very good chance of the combination being favorable for pepsin cleavage.

A few exceptions were observed. Neither of the aliphatic β-branched amino acids (Ile and Val) were favorable residues at the P1 position with average cleavage probability being 2% and 4%, respectively. However, only when the P1′ position was occupied by Trp did pepsin cleave the amide bond with better than expected average probability (Ile-Trp for 30% and Val-Trp 24%). A combination of residues containing amide and/or carboxylic acid side chains at P1–P1′, on the other hand, appears to be disfavored: Asn-Glu, Gln-Asp, Asp-Asn, and Asn-Gln linkages were never cleaved by pepsin in a total of more than 100 occurrences. Gly is the least favored residue at the P1′ position with pepsin cleaving only 3% of the time. However, when Phe is preceding Gly, pepsin cleaves better than average (28%).

Pepsin cleavage of SOD1 mutants

SOD1 (superoxide dismutase 1) wild-type (WT) and L38V mutants were digested by immobilized pepsin under

DOI: 10.1002/rcm

RCM

Table 2. Pepsin specificity

	P4	P3	P2	P1	P1'	P2'	P3'	P4'	
F	17.8%	6.5%	6.8%	45.8%	28.1%	7.6%	13.8%	13.8%	542
L	12.0%	15.3%	11.1%	44.2%	18.1%	10.9%	17.3%	14.7%	1375
M	12.7%	21.3%	12.2%	35.0%	20.7%	12.2%	19.8%	14.1%	331
C	10.8%	20.7%	20.1%	23.4%	16.8%	12.1%	13.1%	16.5%	214
E	12.7%	14.7%	19.1%	19.9%	10.5%	17.9%	14.9%	11.2%	955
W	15.8%	9.2%	5.1%	16.7%	27.6%	11.6%	15.2%	15.2%	198
Y	17.8%	7.1%	8.6%	16.2%	33.1%	13.7%	13.1%	15.7%	451
D	16.7%	12.3%	17.5%	13.1%	10.5%	11.8%	11.7%	13.1%	799
A	11.8%	19.5%	16.6%	12.9%	16.4%	17.5%	10.0%	11.6%	989
Q	11.8%	12.1%	15.4%	9.7%	6.1%	12.7%	14.6%	12.9%	607
N	14.2%	9.9%	21.7%	9.0%	9.0%	8.5%	11.5%	14.3%	467
T	14.6%	18.9%	14.2%	6.7%	8.8%	15.6%	14.5%	11.0%	657
S	11.4%	19.5%	18.4%	5.3%	9.0%	17.6%	12.5%	13.5%	851
G	13.1%	19.1%	9.4%	4.1%	3.2%	8.2%	7.3%	15.1%	933
V	13.0%	17.7%	18.1%	4.0%	18.6%	22.1%	17.4%	12.0%	876
I	13.2%	19.0%	17.2%	2.4%	24.3%	18.3%	20.2%	12.8%	706
K	12.7%	1.6%	11.0%	0.3%	7.7%	14.3%	17.0%	12.7%	763
H	11.7%	1.7%	7.7%	0.2%	8.6%	2.6%	9.1%	11.5%	405
R	13.3%	1.2%	13.3%	0.0%	9.8%	21.1%	15.3%	13.9%	737
P	19.4%	13.7%	0.2%	0.0%	5.6%	1.3%	2.7%	20.7%	637

Percentage of cleavage when each amino acid residue occupies each position. The last column is the number of each amino acid observed in this study.

identical conditions (data not shown). While peptide 21-53 was observed in both WT SOD1 and L38V, peptide 21-37 was observed only in SOD1 WT and not in the L38V digest. What this means is that pepsin cleaved at F20-E21, G37-L38, and N53-T54 in SOD1 WT and that it cleaved at F20-E21 and N53-T54 but not at G37-V38 in L38V mutant. The probability of Leu or Val being at the P1′ position is almost identical (Table 2) and this table cannot predict the different cleavage patterns. On the other hand, the P1–P1′ combination matrix (Table 3) shows that the Gly-Leu bond was cleaved by pepsin 13% of 77 occurrences while the Gly-Val bond was never cleaved by pepsin in 59 chances. These SOD1 mutant digestions may exemplify the presence of synergy between positions P1 and P1′.

Comparison with the previous data

The specificity of pepsin has been investigated in the past using small model peptides.[2–4] Powers *et al*. also statistically analyzed pepsin specificity using published data during 1967–1972.[16] In their study, 177 proteins or peptides consisting 7089 amino acid residues – the average size about 40 amino acids per molecule – were used. On average 14.8%

Table 3a. Pepsin specificity by P1–P1′ positions

	F	L	M	C	E	W	Y	D	A	Q	N	T	S	G	V	I	K	H	R	P	Ave.
Y	65%	68%	30%	25%	70%	71%	36%	48%	48%	47%	39%	23%	31%	17%	11%	7%	0%	0%	0%	0%	33%
F	85%	84%	64%	75%	53%	40%	33%	37%	38%	24%	28%	21%	5%	9%	8%	8%	0%	0%	0%	0%	28%
W	60%	60%	50%	57%	63%	-	50%	17%	17%	33%	38%	17%	23%	17%	24%	30%	0%	0%	0%	0%	28%
I	65%	63%	62%	40%	36%	15%	20%	42%	30%	40%	25%	18%	13%	9%	9%	0%	2%	0%	0%	0%	24%
M	83%	58%	42%	0%	29%	33%	25%	20%	30%	20%	0%	11%	0%	11%	6%	0%	0%	0%	0%	0%	21%
V	50%	53%	61%	21%	31%	10%	16%	25%	28%	23%	16%	11%	12%	0%	4%	2%	0%	0%	0%	0%	19%
L	64%	56%	66%	36%	21%	29%	25%	12%	16%	7%	4%	8%	4%	13%	7%	1%	0%	0%	0%	0%	18%
C	20%	54%	0%	50%	36%	22%	0%	33%	6%	33%	25%	0%	20%	12%	0%	18%	0%	0%	0%	0%	17%
A	55%	54%	38%	18%	35%	8%	9%	13%	14%	7%	9%	7%	6%	4%	3%	2%	0%	0%	0%	0%	16%
E	42%	45%	29%	20%	9%	24%	19%	6%	6%	2%	0%	0%	4%	0%	4%	6%	0%	0%	0%	0%	10%
D	44%	46%	38%	0%	11%	21%	17%	5%	5%	0%	5%	2%	2%	4%	0%	0%	0%	0%	0%	0%	10%
R	42%	34%	26%	29%	9%	13%	16%	9%	8%	6%	4%	0%	5%	3%	0%	0%	0%	0%	0%	0%	10%
N	42%	45%	7%	0%	13%	0%	11%	0%	4%	5%	0%	0%	4%	0%	0%	0%	0%	0%	0%	0%	9%
S	52%	42%	22%	0%	4%	6%	14%	2%	5%	0%	0%	5%	2%	0%	3%	0%	0%	0%	0%	0%	9%
T	31%	27%	25%	35%	3%	29%	4%	12%	5%	3%	6%	9%	3%	2%	0%	0%	0%	0%	0%	0%	9%
H	43%	33%	29%	17%	6%	0%	10%	15%	22%	15%	0%	11%	3%	0%	0%	0%	0%	0%	0%	0%	9%
K	47%	33%	32%	0%	12%	13%	0%	2%	3%	3%	0%	4%	0%	2%	0%	2%	0%	0%	0%	0%	8%
Q	33%	26%	17%	0%	11%	0%	20%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	6%
P	17%	24%	18%	20%	8%	0%	6%	0%	5%	0%	0%	0%	2%	0%	11%	2%	0%	5%	0%	0%	6%
G	28%	7%	8%	6%	1%	0%	10%	0%	0%	0%	0%	0%	0%	0%	0%	0%	2%	0%	0%	0%	3%
Ave.	46%	44%	35%	23%	20%	17%	16%	13%	13%	10%	9%	7%	5%	4%	4%	2%	0%	0%	0%	0%	14%

The probability of pepsin cleavage between two amino acids.
For example, there is 65% chance of cleavage between F-Y, while there is 33% chance of cleavage between Y-F.
Sequence W-W was never observed in the entire analysis (Table 3b).

DOI: 10.1002/rcm

Table 3b. The number of occurrences of each combination

	F	L	M	C	E	W	Y	D	A	Q	N	T	S	G	V	I	K	H	R	P	total
Y	17	47	10	4	40	7	11	25	29	17	18	22	26	36	27	27	28	8	25	26	450
F	13	45	11	8	36	5	18	49	39	21	25	33	38	43	25	37	21	13	34	27	541
W	10	15	4	7	8	0	6	6	18	6	16	12	13	6	21	10	12	5	15	9	199
I	23	54	21	15	61	13	20	53	40	35	28	28	40	43	53	38	42	17	47	33	704
M	6	45	12	4	24	9	8	20	27	15	4	18	18	19	18	21	22	14	13	12	329
V	28	92	18	14	61	10	31	48	67	35	25	64	59	59	69	44	48	31	36	38	877
L	47	171	29	25	91	14	52	85	99	70	50	64	98	77	88	71	70	38	83	56	1378
C	10	24	3	2	11	9	7	9	18	6	8	9	10	17	13	11	19	8	12	8	214
A	38	114	32	11	63	13	23	62	94	44	43	43	67	73	64	55	52	15	46	37	989
E	36	92	34	10	101	17	27	52	66	42	27	50	57	66	56	32	56	22	56	57	956
D	39	74	16	6	55	14	47	44	66	25	22	42	51	55	57	37	64	12	38	38	802
R	24	76	19	21	57	8	25	33	62	36	23	35	40	59	45	28	33	28	48	38	738
N	19	49	14	5	48	1	19	29	24	19	16	17	27	30	23	25	23	13	36	30	467
S	52	69	18	9	45	17	35	49	64	41	32	44	48	68	64	47	43	27	45	37	854
T	42	70	16	17	38	7	26	26	43	31	17	35	35	53	44	48	37	10	36	27	658
H	14	33	7	6	18	8	10	13	23	20	9	9	31	22	26	20	19	71	19	17	395
K	19	84	19	10	57	15	19	48	65	35	34	25	49	60	56	42	51	13	36	27	764
Q	24	78	6	12	38	11	20	39	29	39	10	21	33	47	46	43	30	13	37	29	605
P	23	59	17	10	36	7	18	48	40	26	28	34	49	38	36	41	38	19	29	43	639
G	58	84	25	18	67	13	29	61	76	44	32	52	62	62	45	29	55	28	46	48	934
total	542	1375	331	214	955	198	451	799	989	607	467	657	851	933	876	706	763	405	737	637	13493

of amide bonds were cleaved by pepsin. In the present study, 39 proteins consisting 13 766 amino acid residues – average size about 353 amino acids long – were used. On average 13.6% of amide bonds were cleaved by immobilized pepsin.

The previous study and the current study agree well as a general trend (Table 4). For example, the three most favorable amino acids at the P1 position are Phe, Leu, and Met in both studies. Similarly, positively charged amino acids (His, Lys, and Arg) and the cyclyc imino acid (Pro) are the least favored at the P1 position.

On the other hand, the present study shows sharper specificity than the previous study. For example, pepsin cleaved with a probability of less than 0.3% when His, Lys, Pro, or Arg occupies the P1 position in the present study, while pepsin cleaved 6%, 3%, 1%, or 6% of the time, respectively, in the previous study (Table 4). This higher specificity is probably due to the narrower experimental variations in the present database. In the present study, all proteins were digested by immobilized porcine pepsin with very similar conditions (1°C for 20–30 s at pH 1.3–2.5 in the presence of denaturant). On the other hand, the digestion conditions in the study of Powers *et al*. may vary significantly, as that study combined porcine pepsin digestion results published in *Biochemistry*, *Journal of Biological Chemistry*, and *Biochemistry Journal* from 1969–1972, irrespective of the digestion conditions employed.

Table 4. Pepsin specificity at P1 and P1′ positions in the present and previous studies[16]

	P1			P1′	
aa	present	previous	aa	present	previous
F	46%	51%	Y	33%	34%
L	44%	41%	F	28%	29%
M	35%	43%	W	28%	21%
C	23%	11%	I	24%	26%
E	20%	24%	M	21%	18%
W	17%	40%	V	19%	23%
Y	16%	24%	L	18%	20%
D	13%	26%	C	17%	11%
A	13%	16%	A	16%	21%
Q	10%	14%	E	11%	13%
N	9%	12%	D	11%	16%
T	7%	11%	R	10%	11%
S	5%	9%	N	9%	8%
G	4%	7%	S	9%	8%
V	4%	8%	T	9%	12%
I	2%	5%	H	9%	11%
K	0%	3%	K	8%	15%
H	0%	6%	Q	6%	8%
R	0%	6%	P	6%	9%
P	0%	1%	G	3%	8%

Protein size and cleavage probability

Larger proteins tend to give lower pepsin cleavage probabilities when compared to smaller proteins (Fig. 2). A larger number of peptides are produced from a large protein than from a small protein, assuming pepsin acts similarly. A larger number of peptides may lead to chromatographic crowding and thus ion suppression in

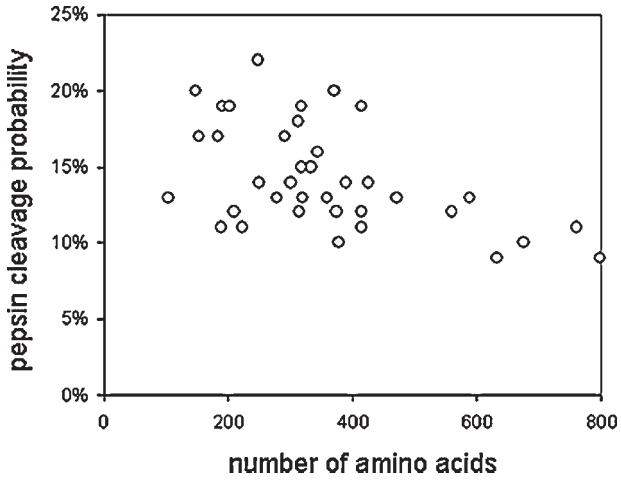

Figure 2. Protein size versus pepsin cleavage probability.

DOI: 10.1002/rcm

the mass analysis. Therefore, there is higher chance of failing to detect less populated and/or difficult to ionize peptides in a larger protein. This contributes to the lower average cleavage probability in the current study than that observed in the previous study.

CONCLUSIONS

The cleavage specificity of immobilized porcine pepsin was examined statistically with very narrow experimental variations. Pepsin cleaved after Phe and Leu favorably as previously observed. Immobilized pepsin rarely cleaved right after His, Lys, Pro, or Arg (being the P1 position) and/or one residue after P (being the P2 position) in H/D exchange compatible low temperature and short reaction time. If cleavage is observed in these situations, a more strict examination of the peptide identity may be necessary. The residue at the P1 position clearly has the strongest influence on pepsin cleavage, while the residues on positions P3 to P3′ have some effects. The effects of residues at the P1 and P1′ positions are generally additive, with a few exceptions. The residues at the P4 and P4′ positions have minimal influence on the pepsin cleavage.

Acknowledgements

SOD1 WT and L38V were kindly provided by Professor Jeffrey Agar of Brandeis University.

REFERENCES

1. Sielecki AR, Fedorov AA, Boodhoo A, Andreeva NS, James MN. *J. Mol. Biol.* 1990; **214**: 143.
2. Fruton JS. *Adv. Enzymol.* 1970; **33**: 401.
3. Trout GE, Fruton JS. *Biochemistry* 1969; **8**: 4183.
4. Antonov VK. *Adv. Exp. Med. Biol.* 1977; **95**: 179.
5. MacCoss MJ, Wu CC, Yates JR 3rd. *Anal. Chem.* 2002; **74**: 5593.
6. Tabb DL, MacCoss MJ, Wu CC, Anderson SD, Yates JR 3rd. *Anal. Chem.* 2003; **75**: 2470.
7. Bogdanov B, Smith RD. *Mass Spectrom. Rev.* 2005; **24**: 168.
8. Zhang Z, Smith DL. *Protein Sci.* 1993; **2**: 522.
9. Engen JR, Smith DL. *Anal. Chem.* 2001; **73**: 256A.
10. Mandell JG, Falick AM, Komives EA. *Proc. Natl. Acad. Sci. USA* 1998; **95**: 14705.
11. Hoofnagle AN, Resing KA, Ahn NG. *Annu. Rev. Biophys. Biomol. Struct.* 2003; **32**: 1.
12. Del Mar C, Greenbaum EA, Mayne L, Englander SW, Woods VL Jr. *Proc. Natl. Acad. Sci. USA* 2005; **102**: 15477.
13. Englander J, Del Mar C, Li W, Englander S, Kim J, Stranz D, Hamuro Y, Woods VL Jr. *Proc. Natl. Acad. Sci. USA* 2003; **100**: 7057.
14. Hamuro Y, Coales SJ, Southern MR, Nemeth-Cawley JF, Stranz DD, Griffin PR. *J. Biomol. Techniques* 2003; **14**: 171.
15. Balgley BM, Laudeman T, Yang L, Song T, Lee CS. *Mol. Cell. Proteomics* 2007; **6**: 1599.
16. Powers JC, Harley AD, Myers DV. *Adv. Exp. Med. Biol.* 1977; **95**: 141.
17. Hamuro Y, Burns LL, Canaves JM, Hoffman RC, Taylor SS, Woods VL Jr. *J. Mol. Biol.* 2002; **321**: 703.
18. Hamuro Y, Zawadzki KM, Kim JS, Stranz D, Taylor SS, Woods VL Jr. *J. Mol. Biol.* 2003; **327**: 1065.
19. Hamuro Y, Anand G, Kim J, Juliano C, Stranz D, Taylor S, Woods VL Jr. *J. Mol. Biol.* 2004; **340**: 1185.

DOI: 10.1002/rcm