Electrospray mass spectrometric evidence for the occurrence of two major variants in native pig pepsin A

Brian N. GREEN,* Arwyn T. JONES† and Norman B. ROBERTS‡§

*VG Organic, Tudor Road, Altrincham, Cheshire WA14 5RZ, U.K., and Departments of †Physiology and ‡Clinical Chemistry, University of Liverpool, Liverpool L69 3BX, U.K.

Native pig pepsin was analysed by negative ion electrospray mass spectrometry in order to rationalize anomalies between the published sequences. Outstanding variations in otherwise identical sequences indicate that amino acid residue 242 is either Asp or Tyr, and in some determinations an additional Ile is inserted

at position 230. Mass spectrometric evidence is consistent with the presence, in the native enzyme, of two variants in comparable abundance, with either Asp or Tyr at residue 242. There is no evidence for the additional Ile at position 230.

INTRODUCTION

Pepsin is a member of the family of gastric aspartic proteinases, comprising pepsins and gastricsin(s), and plays a major role in the digestion of food in vertebrates. One of these, pig pepsin (EC 3.4.23.1), has been sequenced in several laboratories (Table 1) [1–7]. Although there is consensus on the major part of the sequence, variations occur at amino acid 242 (Asp or Tyr), and an Ile is inserted in two sequences at residue 230 [1,2].

Furthermore, although chromatographic and electrophoretic heterogeneity was observed by some earlier workers [8,9], more recently there has been no direct mention of the occurrence of variants, except perhaps by Sielecki et al. [6], where two peaks were observed during chromatographic purification of the native pepsin.

Electrospray mass spectrometry (ESMS) can determine the molecular masses of proteins with a precision of better than 0.01% [10]. Generally with this technique, proteins are analysed by forming a series of multiply charged (protonated) positive

ions. This is because the majority of proteins contain a sufficient number of basic amino acids, and hence obtain a sufficient number of positive charges to bring the mass-to-charge (m/z) ratios of the ions formed well into the range (4000) of commercial quadrupole mass spectrometers. However, a feature of the gastric aspartic proteinases is the paucity of basic amino acids in their sequences. Pig pepsin contains only two Arg, one Lys and one His and hence only three or four sites for protonation. Thus, in common with several aspartic proteinases, it is unsuitable for ESMS analysis as positive ions. It does, however, contain large numbers of acidic residues (13 Glu and 28 or 29 Asp), which makes it well suited for analysis as negative (deprotonated) multiply charged ions. Loo et al. [10] analysed pig pepsin in this way and measured the molecular mass as 34605 Da, i.e. 21 Da higher than that predicted from the sequence that they used to calculate the molecular mass [4].

Here we determine by ESMS the molecular masses of two major pig pepsin A variants (A1 and A2) isolated from a commercial sample by high performance anion exchange

Table 1 Differences between various published sequences of pig pepsin

See the text for details. The molecular mass is calculated assuming that Ser-68 is phosphorylated.

Source and reference	Residue		Total number of residues	Molecular mass (Da)
	229- 230	242		
Protein sequence				
Tang et al. [1]	Ala-Ile-Asn	Asp	327	34696.9
cDNA sequence				
Tsukagoshi et al. [2]	Ala-Ile-Asn	Asp	327	34696.9
Lin et al. [3]	Ala- -Asn	Tyr	326	34631.9
Crystal structure				
Adad-Zapatero et al. [4]	Ala- -Asn	Asp	326	34583.8
Cooper et al. [5]	Ala- -Asn	Asp	326	34583.8
Sielecki et al. [6]	Ala- -Asn	Asp	326	34583.8
Hartsuck et al. [7]	Ala- -Asn	Asp	326	34583.8

Abbreviation used: ESMS, electrospray mass spectrometry.

§ To whom correspondence should be addressed.

chromatography (HPAEC). The masses show, within experimental error (0.01 %), that they agree with the values predicted from the sequences with residue 242 as either Asp or Tyr, provided that Ile-230 is omitted. Furthermore, ESMS data for a CNBr digest of the native (unseparated) pepsin A is consistent with the presence of both variants in comparable amounts.

EXPERIMENTAL

Samples were analysed in the negative ion mode on a VG Quattro (VG Organic, Altrincham, U.K.) at a concentration of 5–10 pmol/ μ l of propan-2-ol/water (1:1, v/v) containing 0.3 % ammonia. They were introduced into the electrospray source at 5 μ l/min. Typically, scanning was in the range m/z 900–1700 and data were summed over approx. 5 min to obtain the final spectra. These data, on an m/z scale, were transformed on to a true molecular mass scale using maximum entropy (MaxEnt)-based software supplied with the instrument [11]. Mass scale calibration employed the multiply charged negative ion series from separate introductions of horse heart myoglobin (cat. no. M1882; Sigma Chemical Co., Poole, Dorset, UK.; sequence molecular mass = 16951.5 Da). Molecular masses are based on the following atomic weights of the elements: C = 12.011, H = 1.00794, N = 14.00674, O = 15.9994, P = 30.97376 and S = 32.066 [12].

Pig pepsin A (Sigma P-6887) was separated into two major components on a 7.5 cm $\times$ 0.75 cm TSK DEAE-5PW anion exchange column (Anachem, Luton, U.K.) using a linear gradient of 0–0.8 M NaCl. The collected fractions were desalted by extensive dialysis against 100 mM HCl, pH 4.0, at 4 °C, and lyophilized before mass spectrometric analysis.

CNBr digestion was carried out (24 h at 20 °C) on pepsin A (10 mg/ml) in 50 % (v/v) aqueous formic acid. The CNBr concentration was 6 mg/ml. After diluting with 10 vol. of water, the fragment peptides were purified by retaining them in a 3000 Da cut-off microconcentrator (Amicon Ltd., Upper Mill, Stonehouse, Glos., U.K.), followed by washing the retentate with 2 $\times$ 300 μ l of 1 % (v/v) formic acid. The washed retentate was dissolved in water and diluted (5–10 pmol/ μ l) with water/propan-2-ol (1:1, v/v) containing 0.3 % (v/v) ammonia for electrospray analysis.

RESULTS AND DISCUSSION

Figure 1 is the anion exchange chromatogram of native pig pepsin showing the two variants A1 and A2 clearly resolved. Figure 2 shows typical electrospray spectra of the pepsin as received (Figure 2a) and of components A1 (Figure 2b) and A2 (Figure 2c) separated as shown in Figure 1. A second commercial sample of the native pepsin (cat. no. 195367; ICN Biomedicals Ltd., High Wycombe, Bucks., U.K.) gave an identical spectrum to that shown in Figure 2(a). These spectra were obtained by processing the original multiply charged spectra with maximum entropy (MaxEnt) based software [11], in order to take advantage of the resolution enhancement afforded by this technique. Without this enhancement, components which were close together in molecular mass were insufficiently resolved from one another to allow accurate determination of their molecular masses.

Comparing the spectra from the unseparated and separated samples illustrates the limitations of analysing the mixture, even after resolution enhancement by the MaxEnt software. Although the spectrum from the unseparated sample (Figure 2a) shows the lower mass component clearly, there is some uncertainty about higher mass components. However, the spectra from the isolated pepsins A1 and A2 (Figures 2b and 2c) show almost identical detail, but are displaced in molecular mass by 49 Da. Peaks

Figure 1 Anion exchange chromatograph of native pig pepsin showing separation of the two major variants, A1 and A2

higher in mass than the major peaks can be attributed to alkali metal adducts, predominantly sodium.

The results from seven repeat analyses of pepsins A1 and A2 are summarized in Table 2. The mean molecular masses determined for the major components are within 2 Da (0.006 %) of the values predicted from the two 326-amino-acid sequences given in Table 2 after adding 80 to account for the phosphorylated Ser-68 [1]. Both pig pepsin variants therefore appear to be phosphorylated, in contrast to ESMS studies on five human pepsin variants, where only two minor variants were tentatively assigned as being phosphorylated [13]. Also, the mean difference in molecular mass between A1 and A2 was 48.4 (Asp to Tyr = 48.1 Da). Thus the molecular mass measurements on the intact variants are consistent with residue 242 being Asp in A1 and Tyr in A2. Minor components, measured to have a molecular mass 183.9 Da higher than the major components, can be attributed to the pepsins with two additional amino acid residues (Ala plus Leu; 184.2 Da) from the activation peptide [14], i.e. corresponding to the protein containing residues –2 to 326.

To isolate the section of the sequence containing the variants, the unseparated pepsin was digested with CNBr; the relevant part of the resulting spectrum is shown in Figure 3. This is the MaxEnt processed spectrum from the digest. It shows three out of the five expected fragment peptides designated C1–C5. These are C3 (residues 200–246; calculated molecular masses 4408.7 Da and 4456.8 Da for residue 242 as Asp and Tyr respectively), C4 (residues 247–290; molecular mass 4636.1 Da) and C5 (residues 291–326; molecular mass 4112.7 Da). Their mean measured masses (two determinations on each of two CNBr digests) were: C3 (Asp-242), 4408.6 $\pm$ 0.4 Da (mean $\pm$ S.D.); C3 (Tyr-242), 4456.4 $\pm$ 0.5 Da; C4, 4636.1 $\pm$ 0.3 Da; C5, 4112.4 $\pm$ 0.2 Da. Except for C5, the calculated masses assume that the C-termini are homoserine lactones. The masses of all three fragments closely agree with the masses predicted from the sequences. In particular, C3 clearly shows two major peaks of comparable intensity, whose masses are consistent with those predicted on the basis that residue 242 is either Asp or Tyr.

Unfortunately, the assignments for C3 are not mass spectrometrically unique, since there are four aspartic acid residues in

Figure 2 MaxEnt processed spectra from (a) native pig pepsin A and from pepsins A1 (b) and A2 (c) isolated by anion exchange chromatography

Molecular masses are given in Da.

Table 2 Molecular mass determinations from seven repeat analyses on pig pepsins A1 and A2

Measured mass results are means $\pm$ S.D. Sequence masses for pepsins A1 and A2 are taken from [4] and [3] respectively.

Pepsin	Residues	Molecular mass (Da)		
		Measured	From sequence	Measured minus sequence
A1	1 to 326	34585.5 $\pm$ 0.8	34583.8	1.7
A1	—2 to 326	34769.3 $\pm$ 3.6	34768.1	1.2
A2	1 to 326	34633.9 $\pm$ 0.5	34631.9	2.0
A2	—2 to 326	34817.7 $\pm$ 1.4	34816.2	1.5

this fragment and also a Val which could be exchanged for a Phe, all of which could give the same mass difference of 48 Da. Nevertheless, in view of the general consensus on the sequence of pig pepsin by other workers, except for residue 242 and perhaps residue 230, it seems reasonable to accept that both Asp and Tyr can occur as residue 242 in the native enzyme. Certainly, it would explain why some workers found Asp and another found Tyr, since the result would depend on which variant was isolated in the initial purification of the native pepsin. The occurrence of the two variants in comparable amounts also explains why Loo et al. [10] measured the molecular mass as 35605 Da, close to the mean mass of the two variants (35608 Da).

In the structure of pig pepsin, residue 242 forms part of a highly flexible β hairpin loop region incorporating residues 241–245 [5]. This loop also has dyad structural equivalence to the active-site flap which, via electrostatic contacts, helps to maintain favourable positioning of the substrate in the active site prior to peptide bond hydrolysis. The 241–245 loop, however, adopts a more solvent-exposed orientation than the active-site flap, and its high flexibility results in poor crystallographic structural definition, making it difficult to predict whether it makes any

contacts with substrate residues [5–7]. A similar single amino acid change (Asp-244 to Gly) in the same loop region of calf chymosin leads to significant proteolytic differences between the two variants [15]. The chymosin loop, comprising residues 241–249, is also flexible and is believed to influence substrate specificity by interacting electrostatically with substrate residues. The topologically equivalent loop has been implicated as one factor involved in the high cleavage specificities of two other aspartic proteinases, renin [16] and the HIV proteinase [17]. Kinetic studies are now required to investigate the functional proteolytic implications of the Asp to Tyr difference between the two pig pepsin variants.

In conclusion, high performance anion exchange chromatography and ESMS have unequivocally demonstrated the heterogeneity of a commercial pig pepsin preparation. In addition, ESMS has provided a rational explanation for the discrepancies between the published sequences of pig pepsin by suggesting the occurrence of two variants with amino acid 242 as either Asp or Tyr. The two variants may correspond to pepsins B and C of Ryle [9], or to pig pepsin peaks I and III of Rajagopalan et al. [8], none of which were further characterized. Since only one gene

Figure 3 Part of the MaxEnt processed spectrum from a CNBr digest of pig pepsin A, showing three of the five peptide fragments (C3–C5)

C3 (residues 200–246) contains the variants, with residue 242 being either Asp or Tyr. Molecular masses are given in Da. Masses in parentheses are calculated from the sequences, assuming that C3 and C4 form the homoserine lactone at their C-termini. Peaks F and H can be assigned as formylated peptides and homoserines respectively.

encoding pig pepsin has been identified [18], heterogeneity involving single substitutions may possibly be attributed to allelic differences at a single locus.

REFERENCES

- Tang, J., Sepulveda, P., Marciniyszyn, J., Jr., Chen, K., Huang, W.-Y., Tao, N., Liu, D. and Lanier, J. P. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3437–3439
- Tsukagoshi, N., Ando, Y., Tomita, Y., Uchida, R., Takemura, T., Sasaki, T., Yamagata, H., Udaka, S., Ichihara, Y. and Takahashi, K. (1988) *Gene* **65**, 285–292
- Lin, X. L., Wong, R. N. and Tang, J. (1989) *J. Biol. Chem.* **264**, 4482–4489
- Abad-Zapatero, C., Rydal, T. J. and Erickson, J. (1990) *Proteins Struct. Funct. Genet.* **8**, 62–81
- Cooper, J. B., Khan, G., Taylor, G., Tickle, I. J. and Blundell, T. L. (1990) *J. Mol. Biol.* **21**, 199–222
- Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S. and James, M. G. (1990) *J. Mol. Biol.* **214**, 143–170
- Hartsuck, J. E., Koelsch, G. and Remington, S. J. (1992) *Proteins Struct. Funct. Genet.* **13**, 1–25
- Rajagopalan, T. J., Moore, S. and Stein, W. H. (1966) *J. Biol. Chem.* **241**, 4940–4950
- Ryle, A. P. (1970) *Methods Enzymol.* **19**, 316–336
- Loo, J. A., Loo, R. R. O., Light, K. J., Edmonds, C. G. and Smith, R. D. (1992) *Anal. Chem.* **64**, 81–88
- Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A. and Skilling, J. (1992) *Rapid Commun. Mass Spectrom.* **6**, 707–711
- IUPAC Commission on Atomic Weights and Isotopic Abundances (1993) *J. Phys. Chem. Ref. Data* **22**, 1571–1584
- Jones, A. T., Green, B. N. and Roberts, N. B. (1995) in *Aspartic Proteinases: Structure, Function, Biology, and Biomedical Implications* (Takahashi, K., ed.), pp. 83–89, Plenum Press, New York
- Tang, J., Sepulveda, P., Marciniyszyn, J., Jr., Chen, K. C. S., Huang, W.-Y., Tao, N., Liu, D. and Lanier, J. P. (1973) *Proc. Natl. Acad. Sci. U.S.A.* **70**, 3437–3439
- Gilliland, G. L., Oliva, M. T. and Dill, J. (1992) in *Structure and Function of the Aspartic Proteinases* (Dunn, B. M., ed.), pp. 23–37, Plenum Press, London
- Dhanaraj, V., Dealwis, E. G. and Frazdo, C. (1992) *Nature (London)* **357**, 446–472
- Blundell, T. L., Lapatto, D. E. and Whittle, P. J. (1990) *Trends Biochem. Sci.* **15**, 425–430
- Meijrink, P. H., Bebelman, J., Oldenburg, A. N., Defize, J., Planta, R. J., Eriksson, A. W., Pals, G. and Mager, W. H. (1993) *Eur. J. Biochem.* **213**, 1283–1296