Expression of soluble cloned porcine pepsinogen A in *Escherichia coli*

Takuji TANAKA and Rickey Y. YADA*

Department of Food Science, University of Guelph, Guelph, Ontario N1G 2W1, Canada

A system for the production of soluble porcine pepsinogen A (EC 3.4.23.1) was developed by fusing the pepsinogen and thioredoxin genes and then expressing the fused product (Trx-PG) in *Escherichia coli*. The expressed fusion protein was purified using a combination of ion-exchange and hydrophobic chromatography. Trypsin digestion of the fusion protein yielded pepsinogen which was one residue longer than the intrinsic length. Acidification of either the fusion protein or pepsinogen (tryptic digestion of Trx-PG) yielded recombinant pepsin A (r-pepsin). When compared with commercial porcine pepsin A, r-pepsin had similar milk-clotting and proteolytic activities, kinetic parameters and pH dependency. The above results indicate that an expression system was developed which yielded fully active soluble pepsin(ogen) from *Escherichia coli*.

INTRODUCTION

Porcine pepsin A (EC 3.4.23.1) belongs to the aspartic proteases and is characterized by having two catalytic aspartic acid residues in the binding site. X-ray crystallography of the aspartic proteases [1–10] has shown that these proteases are composed of two similar domains and that each domain contributes a catalytic aspartate which is located at the bottom of a cleft formed between the two domains.

Although the three-dimensional structure of pepsin has been successfully elucidated, only a few studies have appeared regarding putative functional residues and regions [11,12]. This, in part, has been due to difficulties in expressing cloned pepsinogen (pepsin is obtained from the activation of pepsinogen) which has been in the form of inclusion bodies. Similarly, all cloned aspartic proteases using the *Escherichia coli* system, with the exception of human immunodeficiency virus protease [13,14], have been expressed as inclusion bodies [12,15–19]. Inclusion body proteins require that they be unfolded and then refolded in order that a 'properly' folded and functional protein is obtained. Refolding, however, does not ensure that all the protein molecules have the identical and correct structure. By contrast, if the protein can be expressed in the soluble fraction and is identical in specific activity to the native protein, it is assumed that every protein molecule is folded correctly.

Fusion protein systems have been used to obtain soluble proteins [20,21]. However, a common drawback to these systems is the requirement to remove the fused portion which is usually accomplished through limited proteolysis. Several proteases have been used for this purpose and include: enterokinase, thrombin and blood coagulation factor Xa. These proteases are generally expensive, and therefore, not economical for large-scale purification of the target protein. In the case of pepsin, two possible strategies to recover this protein or pepsinogen from the fusion protein without the use of expensive proteases exist: one is the autocatalytic cleavage of fusion pepsinogen to produce pepsin, and the other is the tryptic digestion of the fused protein to recover pepsinogen. The fusion pepsinogen could cleave the fused protein autocatalytically under acidic conditions since pepsinogen excises its prosegment, which is placed between the fused protein and pepsin. Alternatively, since pepsin has only one lysine and two arginine residues in its sequence, tryptic digestion would be expected to retain the intact pepsin portion by hydrolysing the fused protein in the linker region or in the prosegment region.

The development of an expression system to produce soluble protein would greatly aid in studies whereby the effects of mutation(s) on structure–function of both pepsin and pepsinogen could be examined. In the present study, a thioredoxin fusion protein system [21] for pepsinogen expression was investigated in attempts to improve the solubility of cloned pepsin(ogen).

MATERIALS AND METHODS

Materials

The plasmid, pTrxFus, was obtained from Invitrogen (San Diego, CA, U.S.A.). Restriction enzymes and modification enzymes for DNA manipulation were from Boehringer Mannheim Canada (Laval, Québec, Canada) and Life Technologies (Gaithersburg, MD, U.S.A.). DEAE-Sepharose CL-6B and Butyl-Sepharose 4 Fast Flow were purchased from Pharmacia Biotech (Uppsala, Sweden). Trypsin and porcine pepsinogen A were purchased from Sigma Chemical (St. Louis, MO, U.S.A.). All other chemicals were of the purest grade commercially available.

Construction of expression plasmid

Porcine pepsinogen gene was fused with thioredoxin gene as follows. The pepsinogen gene was excised from pGBT-T19-pp [12] with *Eco*RI restriction enzyme as a 1.4 kb fragment and then blunt-ended. This gene was introduced into the blunt-ended *Sal*I site of pTrxFus. Positive clones were screened using *Nde*I digestion and confirmed by Western blot analysis. The resultant plasmid was named pTFP1000.

Purification of fused pepsinogen

Plasmid pTFP1000 was used to transform *Escherichia coli* GI724 according to the method described by Hanahan [22]. The resulting

Abbreviations used: Trx-PG, *E. coli* thioredoxin–porcine pepsinogen fusion protein; r-pepsinogen, recombinant pepsinogen; r-pepsin, recombinant pepsin A; n-pepsin, native commercial porcine pepsin A.

* To whom correspondence should be addressed.

transformants were cultured in 1.5 litres of induction medium (1 × M9 salts, 0.2 % Casamino acids, 0.5 % glucose, 1 mM $MgCl_2$ and 150 mg/ml ampicillin) at 22 °C. When the absorbance, at 550 nm, reached 0.5, tryptophan (0.1 mg/ml) was added. Eight hours after induction, cells were harvested by centrifugation at 8000 ***g*** for 10 min at 4 °C. The cells were disrupted by sonication and centrifuged at 18000 ***g*** for 20 min at 4 °C to yield crude cell extracts. The crude extracts were partially purified by 40 % satd. ammonium sulphate precipitation. After dialysis, against 20 mM Tris/HCl (pH 7.5) buffer, the proteins were applied on to a DEAE-Sepharose CL-6B column (5 cm × 25 cm; flow rate 2.5 ml/min). The column was washed with a 0.2 M NaCl solution and eluted with a 0.4 M NaCl solution. The proteins in the eluant were recovered using 80 % satd. ammonium sulphate. The recovered proteins were then applied on to a Butyl-Sepharose 4 Fast Flow column (2.5 cm × 20 cm; equilibrated with 0.8 M ammonium sulphate/20 mM Tris/HCl (pH 7.5); flow rate 0.8 ml/min). The thioredoxin–porcine pepsinogen fusion protein expressed in *E. coli* (Trx-PG) was eluted with a linear gradient of 0.8 M to 0 M ammonium sulphate. Protein concentration in each step was measured using the Lowry method [23].

Purification of pepsinogen and pepsin from Trx-PG

Pepsinogen was obtained by digesting the Trx-PG preparation with 1 % (w/w) trypsin at 37 °C for 8 h. The digest was applied on to a Sephadex G-50 column (0.5 cm × 15 cm). Proteins eluting in the void volume were collected.

In order to obtain pepsin, either Trx-PG or tryptic digested Trx-PG were acidified with a 10 % volume of 40 mM sodium acetate/HCl (pH 1.3) and neutralized with a 10 % volume of 200 mM sodium acetate (pH 5.3). Pepsin molecules were purified using the same protocol as described for pepsinogen.

The purity of the purified pepsinogen and pepsin samples was analysed by Western blotting and Coomassie Blue staining of an SDS/polyacrylamide gel. Protein concentration in each step was measured by the Lowry method [23] and UV absorption ($A_{280}^{1\%} = 13$) [12].

N-terminal sequences of pepsinogen and pepsin were determined by the Edman degradation method using a gas-phase microsequencer.

Characterization of recovered pepsin

Milk-clotting activity of the recombinant pepsin preparation was determined using a 0.32 % (w/v) powdered skim milk solution. Skim milk powder was dissolved in 20 mM sodium acetate buffer (pH 5.3) and 0.8 mM $CaCl_2$. The pepsin solution was added to 1 ml of the skimmed milk solution in an optical cuvette and the absorbance of the mixture was measured at 505 nm [11,24]. One unit of the milk-clotting activity was defined as the amount of protein which gave a 0.4 unit change in absorbance over 1 s.

Proteolytic activity was determined as the rate of release of soluble peptides from a 1 % (w/v) denatured haemoglobin solution [25]. One unit was defined as the amount of protein which gave a change of 1 absorbance unit (due to soluble peptides) at 280 nm in 1 min.

Kinetic constants were determined using the synthetic substrate KPAEFF(NO_2)AL, [F(NO_2) is *p*-nitrophenylalanine] as described by Dunn et al. [26] and Fusek et al. [27]. The incubation mixtures covered a range of substrate concentrations (0.01–0.2 mM) in 0.1 M citrate/HCl, pH 2.1. The reaction was initiated by adding the protein to the pre-incubated mixture. The steady-state kinetic parameters K_m and k_0 were determined by a non-linear least-squares fit of the data using the method described by Sakoda and Hiromi [28].

The pH dependency of the activity of pepsin was analysed using the above synthetic substrate. The reaction was carried out in 90 mM sodium acetate/HCl buffer (pH 1.1–6.0) using 0.1 mM peptide substrate. Hydrolysis was measured as described by Dunn et al. [26] and Fusek et al. [27].

RESULTS AND DISCUSSION

In attempts to determine the amount of fusion protein as a function of the total protein, SDS/PAGE followed by both Coomassie Blue and silver staining were attempted; however, due to the poor staining ability of the fusion protein by both stains an accurate determination could not be made. The amount of the expressed fusion protein was estimated to be 5–15 % of the whole protein in pTFP1000/*E. coli* GI724 as judged from Coomassie Blue staining and Western blot analysis (Figure 1).

Culturing temperature had a substantial effect on the amount of expressed protein. Protein expression was examined at 16, 22 and 30 °C, at various incubation times, using SDS/PAGE. Although culturing at 30 °C yielded larger total amounts of Trx-PG as compared with 16 and 22 °C, most of the protein was not recovered in the soluble fraction. Culturing at 16 °C yielded small amounts of cells and expressed Trx-PG, even after 24 h of incubation; the Trx-PG was not detectable on the SDS/PAGE gel. The yield of soluble fusion protein reached a maximum at 22 °C (Figure 1), as evidenced by stain intensity.

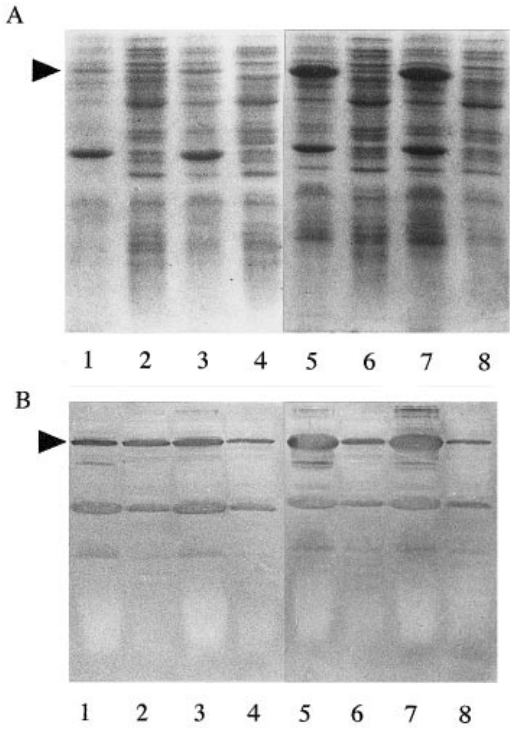

Figure 1 Comparison between expressions at 22 and 30 °C

(**A**) SDS/polyacrylamide gel of cell extracts (soluble fraction) and cell debris (insoluble fraction) from pTFP1000/GI724. (**B**) Western blot of the same gel. The protein extracted from the same volume of culture was applied to each lane. Arrows indicate Trx-PG. Lane 1, cell extracts (22 °C, 7 h); lane 2, cell debris (22 °C, 7 h); lane 3, cell extracts (22 °C, 22 h); lane 4, cell debris (22 °C, 22 h); lane 5, cell extracts (30 °C, 8 h); lane 6, cell debris (30 °C, 8 h); lane 7, cell extracts (30 °C, 23 h); lane 8, cell debris (30 °C, 23 h).

Table 1 Purification of fusion pepsinogen

Method	Protein concentration (mg/ml)	Volume (ml)	Total protein content (mg)
Crude cell extracts	6.9	155	1069
$(NH_4)_2SO_4$ precipitation	12.5	21	263
DEAE-Sepharose	6.5	22	143
Butyl-Sepharose	0.75	12.2	9.15

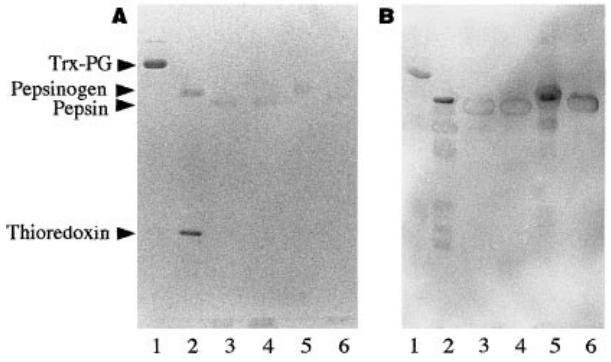

Figure 2 Electrophoresis of Trx-PG and its derivatives

(**A**) SDS/12%-polyacrylamide gel stained with Coomassie Blue G-250. (**B**) Western blot of the same gel. One μg of protein was applied to each lane. Lane 1, Trx-PG; lane 2, tryptic-digested Trx-PG; lane 3, tryptic-digested Trx-PG followed by acid treatment; lane 4, acid-treated Trx-PG; lane 5, native porcine pepsinogen; lane 6, acid-treated native pepsinogen.

Table 2 Kinetic analysis of recombinant and commercial pepsin

Abbreviation: ND, not determined.

Enzyme	Milk-clotting activity (units/mg)	Proteolytic activity (units/mg)	K_m (mM)	k_0 (s^{-1})
r-Pepsin*	27.9 ± 1.5	19.8 ± 0.3	0.033 ± 0.005	65.4 ± 3.1
r-Pepsin†	27.7 ± 1.6	ND	ND	ND
n-Pepsin	28.3 ± 0.8	21.1 ± 0.5	0.026 ± 0.004	79.5 ± 3.8

* r-Pepsin purified directly from Trx-PG.
† r-Pepsin purified from r-pepsinogen.

The purification of the fusion pepsinogen (Trx-PG) was accomplished using two chromatographic separations (Table 1). Unfortunately Trx-PG was only slightly activated using acidification prior to purification (the activity was not high enough to determine the specific activity, i.e. less than 0.001 unit/mg). The low activation of Trx-PG by acidification, at this step, may have been caused by co-precipitation with high-pI proteins or inhibition by other proteins. Due to this low activation and the poor staining, the specific yield of the Trx-PG at each purification step could not be determined. The purity of the Trx-PG preparation was, however, estimated to be over 95% from the Coomassie Blue-stained SDS/PAGE gel which showed some additional minor bands, while Western blotting analysis showed a major Trx-PG band (Figure 2).

Limited proteolysis of Trx-PG by trypsin resulted in two bands on the SDS/PAGE gel (Figure 2). The migration of the observed fragments was as expected, i.e. bands with molecular masses similar to pepsinogen and thioredoxin were observed. N-terminal analysis of the large fragment (i.e. pepsinogen) from trypsin digest was F-M-L-V-K-V-P-L-V-R, indicating that the cleavage site was one amino acid residue upstream from the intrinsic first methionine residue. After cleavage, recombinant pepsinogen (r-pepsinogen) was separated on a gel-filtration column. SDS/PAGE and Western blotting confirmed the presence of r-pepsinogen (Figure 2). The r-pepsinogen preparation had no milk-clotting activity prior to acidification and showed the same activity (27.7 units/mg) as native commercial porcine pepsin A (n-pepsin) after acidification (Table 2).

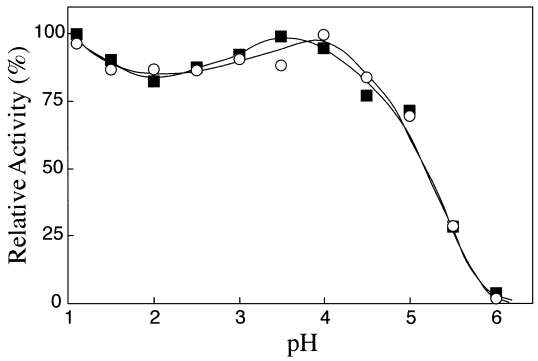

Figure 3 pH dependency of hydrolysis of synthetic peptide substrate with pepsins

Hydrolysis of the synthetic substrate was measured at the appropriate pH. Data are reported as percentage relative activity, as a percentage of the highest activity over the pH range examined for the enzyme in question. Open circles and closed boxes represent r-pepsin and n-pepsin respectively.

Acidification of the Trx-PG to pH 1.3 precipitated out impurities from the solution. Following acidification, gel filtration of the preparation removed the cleaved pepsinogen prosegment (which was fused to thioredoxin) yielding recombinant pepsin (r-pepsin). r-Pepsin was confirmed by SDS/PAGE and Western blotting (Figure 2). The r-pepsin purified directly from the fusion protein had similar milk-clotting activity to pepsin obtained from the r-pepsinogen preparation (Table 2). Further kinetic analyses were therefore conducted only on r-pepsin purified through the direct acidification of Trx-PG. The N-terminal sequence of r-pepsin was I-G-D-P-E-P-L which was consistent with n-pepsin [12].

n-Pepsin was purified from its zymogen using the same method as that for r-pepsin, and was used as a reference. The milk-clotting, proteolytic activities and kinetic parameters of various pepsins are summarized in Table 2. Both r- and n-pepsins showed similar milk-clotting and proteolytic activities, resulting in similar milk-clotting to proteolysis activity ratios (1.41 and 1.34 respectively). The kinetic analyses of r- and n-pepsin are shown in Table 2. Michaelis and rate constants for both pepsins were similar. Figure 3 shows the results of pH dependency of both r- and n-pepsins. Both pepsins maintained 80% activity in the range from pH 1.1 to 4.0, and almost no activity at pH 6.0.

In conclusion, it was confirmed that: (i) the fusion pepsinogen expression system produced soluble Trx-PG (pepsinogen fused with thioredoxin); (ii) r-pepsinogen and r-pepsin were similar to their respective native proteins; and (iii) r-pepsin could be obtained either through direct acidification of Trx-PG or indirectly through the acidification of r-pepsinogen preparation from Trx-PG following trypsinolysis. The flexibility to produce

either pepsinogen or pepsin from the fusion protein provides an excellent means by which mutation(s) of these proteins on structure–function can be examined.

We are grateful to Drs. Jordan Tang and Xin-li Lin for supplying both the porcine pepsinogen cDNA and the pepsin antibody. The chromogenic substrate was synthesized at the Institute of Molecular Biology and Biotechnology of McMaster University, Hamilton, Ontario, Canada. N-terminal sequences were determined at the Biotechnology Service Centre, Department of Clinical Biochemistry, University of Toronto, Toronto, Ontario, Canada. This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1 Andreeva, N. S., Zdanov, A. S., Gustchina, A. E. and Fedorov, A. A. (1984) J. Biol. Chem. **259**, 11353–11365
2 Jaskólski, M., Miller, M., Rao, J. K. M., Leis, J. and Wlodawer, A. (1990) Biochemistry **29**, 5889–5898
3 Navia, M. A., Fitzgerald, P. M. D., McKeever, B. M., Leu, C.-T., Heimbach, J. C., Herberm, W. K., Sigal, I. S., Darke, P. L. and Springer, J. P. (1989) Nature (London) **337**, 615–620
4 Newman, M., Safro, M., Frazao, C., Khan, G., Zdanow, A., Tickle, I. J., Blundell, T. L. and Andreeva, N. (1991) J. Mol. Biol. **221**, 1295–1309
5 Newman, M., Watson, F., Roychowdhury, P., Jones, H., Badasso, M., Cleasby, A., Wood, S. P., Tickle, I. J. and Blundell, T. L. (1993) J. Mol. Biol. **230**, 260–283
6 Suguna, K., Bott, R. R., Padlan, E. A., Subramanian, E., Sheriff, S., Cohen, G. H. and Davies, D. R. (1987) J. Mol. Biol. **196**, 877–900
7 James, M. N. G. and Sielecki, A. R. (1983) J. Mol. Biol. **163**, 299–361
8 Siekecki, A. R., Hayakawa, K., Fujinaga, M., Murphy, M. E. P., Fraser, M., Muir, A. K., Carilli, C. T., Lewicki, J. A., Baxter, J. D. and James, M. N. G. (1989) Science **243**, 1346–1351
9 Gilliland, G. L., Winborne, E. L., Nachman, J. and Wlodawer, A. (1990) Protein **8**, 82–101
10 Tong, L., Pav, S., Pargellis, C., Dô, F., Lamarre, D. and Anderson, P. (1993) Proc. Natl. Acad. Sci. U.S.A. **90**, 8387–8391
11 Cottrell, T. J., Harris, L. J., Tanaka, T. and Yada, R. Y. (1995) J. Biol. Chem. **271**, 19974–19978
12 Lin, X.-l., Wong, R. N. S. and Tang, J. (1989) J. Biol. Chem. **264**, 4482–4489
13 Seelmeier, S., Schmidt, H., Turk, V. and von der Helm, K. (1988) Proc. Natl. Acad. Sci. U.S.A. **85**, 6612–6616
14 Darke, P. L., Leu, C.-T., Davis, L. J., Heimbach, J. C., Diehl, R. E., Hill, W. S., Dixon, R. A. F. and Sigal, I. S. (1989) J. Biol. Chem. **264**, 2307–2312
15 Nishimori, K., Kawaguchi, Y., Hidaka, M., Uozumi, T. and Beppu, T. (1982) Gene **19**, 337–344
16 Kaytes, P. S., Theriault, N. Y., Poorman, R. A., Murakami, K. and Tomich, C.-S. C. (1986) J. Biotechnol. **4**, 205–218
17 Masuda, T., Nakano, E., Hirose, S. and Murakami, K. (1986) Agric. Biol. Chem. **50**, 271–279
18 Imai, T., Cho, T., Takamatsu, H., Hori, H., Saito, M., Masuda, T., Hirose, S. and Murakami, K. (1986) J. Biochem. (Tokyo) **100**, 425–432
19 Chen, Z., Koelsch, G., Han, H.-P., Wang, Z.-J., Lin, X.-l., Hartsuck, J. A. and Tang, J. (1991) J. Biol. Chem. **266**, 11718–11725
20 Nilsson, B., Abrahmsen, L. and Uhlén, M. (1985) EMBO J. **4**, 1075–1080
21 LaVallie, E. R., DiBlasio, E. A., Kovacic, S., Grant, K. L., Schendel, P. F. and McCoy, J. M. (1993) Bio/Technology **11**, 187–193
22 Hanahan, D. (1983) J. Mol. Biol. **166**, 557–580
23 Lowry, O. H., Rosenbrough, N. J., Farr, A. L. and Randall, R. J. (1951) J. Biol. Chem. **193**, 265–275
24 Peter, M. (1976) Anal. Biochem. **73**, 258–261
25 Mizobe, F., Takahashi, K. and Ando, T. (1973) J. Biochem. (Tokyo) **73**, 61–68
26 Dunn, B. M., Jimenez, M., Parten, B. F., Veller, M. J., Rolph, C. E. and Kay, J. (1986) Biochem. J. **237**, 899–906
27 Fusek, M., Lin, X.-l. and Tang, J. (1990) J. Biol. Chem. **265**, 1496–1501
28 Sakoda, M. and Hiromi, K. (1976) J. Biochem. (Tokyo) **80**, 547–555

Received 11 September 1995/22 November 1995; accepted 4 December 1995