The *N*-terminal amino acid sequence of pig kidney endopeptidase-24.11 shows homology with pro-sucrase-isomaltase

Ian S. FULCHER,* Darryl J. C. PAPPIN† and A. John KENNY*‡

*MRC Membrane Peptidase Research Group and †SERC Sequence Unit, Department of Biochemistry, University of Leeds, Leeds LS2 9JT, U.K.

Endopeptidase-24.11 (EC 3.4.24.11), a widely distributed ectoenzyme, was isolated from pig kidneys by detergent solubilization of membranes and immuno-affinity chromatography. In all, 12 preparations of the enzyme were submitted to solid-phase sequencing, yielding a consensus sequence of 25 amino acid residues of the *N*-terminal segment. Some samples were treated with either trypsin or *Staphylococcus aureus* V8 proteinase before sequencing. There were four lysine and one arginine residues in the first nine positions. This segment was susceptible to hydrolysis by trypsin and, in some samples, to endogenous proteinases. From residue 19 onwards, the sequence became intensely hydrophobic. There was a striking homology with the *N*-terminal sequence of pro-sucrase-isomaltase. From Lys⁷ to Leu²⁰ there were seven identical amino acid residues and four conservative substitutions. We suggest that endopeptidase-24.11 is topologically similar to this glycosidase, the *N*-terminus at the cytoplasmic face and the hydrophobic segment serving the roles of both signal peptide and hydrophobic anchor.

INTRODUCTION

Endopeptidase-24.11 (EC 3.4.24.11) is a typical ectoenzyme (subunit M_r 87000) in kidney microvilli and is also present on many other cell surfaces (Gee *et al.*, 1985). When reconstituted into liposomes the junctional peptide, which links the globular protein to the hydrophobic anchor, was found to be 2 nm in length: too short to permit release of the enzyme by treatment with proteinases (Kenny *et al.*, 1983). Many other microvillar enzymes can be released by papain or trypsin, and comparisons of the *N*-terminal residues of the detergent- and proteinase-released forms of sucrase-isomaltase (EC 3.2.1.10/48), dipeptidyl peptidase IV (EC 3.4.14.5), aminopeptidase A (EC 3.4.11.7), aminopeptidase N (EC 3.4.11.2) and γ -glutamyltransferase (EC 2.3.2.2) have supported the view that these ectoenzymes are anchored by a hydrophobic region close to the *N*-terminus [for reviews, see Kenny & Maroux (1982) and Semenza (1986)]. In two instances, sucrase-isomaltase and dipeptidyl peptidase IV, the finding of similar *C*-terminal residues in the two forms has indicated that the anchorage is confined to the *N*-terminal region. Recently, full-length sequences derived from cDNA clones have been reported for γ -glutamyltransferase (Laperche *et al.*, 1986) and pro-sucrase-isomaltase, a polypeptide of 1827 residues (Hunziker *et al.*, 1986). In each case an intensely hydrophobic sequence of 22 and 20 amino acid residues was located near the *N*-terminus, thus supporting the conclusions of the earlier work.

The resistance of endopeptidase-24.11 to release by proteinase treatment has precluded this experimental approach to an understanding of its mode of insertion in the membrane. In the present paper we report the sequence of 25 residues at the *N*-terminus of the detergent-solubilized enzyme and show that it has

considerable homology with the corresponding portion of pro-sucrase-isomaltase.

EXPERIMENTAL

Purification of endopeptidase-24.11

Pig kidney microsomal membranes were solubilized by Triton X-100 and the enzyme isolated by immuno-affinity chromatography (Gee *et al.*, 1983). Two monoclonal antibodies were used: GK7C2 in the earlier purifications, but when GK4A9 was obtained (Gee & Kenny, 1985), this was preferred, since it gave a higher yield for each column run (typically 5 mg of enzyme). Each sample of endopeptidase-24.11 was shown to be single-banded on SDS/polyacrylamide-gel electrophoresis. Some samples were also shown to contain a single antigenic component by cross-immunoelectrophoresis using a polyvalent antiserum raised to kidney microvillar membranes (Fulcher & Kenny, 1983).

Treatment with proteinases

Some samples of the detergent form of endopeptidase-24.11 were treated with trypsin (1:10, w/w) for 2 h and repurified on DEAE-cellulose as previously described (Fulcher & Kenny, 1983). Two samples were treated with proteinase V8 from *Staphylococcus aureus* (Miles Scientific) under the same conditions as those used for trypsin treatment, except that the buffer was 0.1 M-NaHCO₃/Na₂CO₃, pH 8.0, and the incubation time was 1 h. The reaction was stopped by adding di-isopropyl fluorophosphate to 0.1 mM.

Preparation of enzyme samples for sequencing

Samples (2–5 mg) were analysed exhaustively against 0.1% SDS in water (Milli Q; Millipore) and freeze-dried.

‡ To whom correspondence and reprint requests should be sent.

Xaa-Pro-Lys-Pro-Lys-Lys-Lys-Gln-Arg-Trp-Thr-Pro-Leu-Glu-Ile-Ser-Leu-Glu-Val-Leu-Val-Leu-Val-Leu-Val-Xaa-(Ile)-

Fig. 1. Sequence of the *N*-terminal region of endopeptidase-24.11

The sequence shown was derived from studies on 12 samples of the purified enzyme. 1, The sequence of five of the untreated samples ('Untreated¹') started at residue 1 (not identified); these samples had been isolated on a GK4A9 column. 2, Two untreated samples ('Untreated²') sequenced from Lys⁶; these had been isolated on a GK7C2 column. Two samples subjected to treatment with proteinase V8 also sequenced from Lys⁶, whereas three samples treated with trypsin gave sequences starting at Trp¹⁰. (See the text for details.)

Each sample was dissolved in 70 μ l of 0.2 M-NaHCO₃ containing 0.25% (w/v) SDS and added to 20 mg of *p*-phenylene di-isothiocyanate-glass (17 nm pore size, 200–400 mesh, controlled-pore glass; Sigma), prepared by the method of Wachter *et al.* (1973). The mixture was incubated for 90 min at 56 °C under N₂, then washed with water and methanol to remove non-covalently bound material.

Solid-phase sequencing

The glass-coupled peptide was then sequenced by automated solid-phase Edman degradation (Laursen, 1971) using the microsequence facility built by the SERC (Science and Engineering Research Council) Sequence Unit in the Department of Biochemistry, University of Leeds. Anilinothiazolinone amino acids were converted in 0.05 ml of 20% trifluoroacetic acid for 12 min at 70 °C (under N₂) and the corresponding phenylthiohydantoin amino acids identified by reverse-phase h.p.l.c. using a method adapted from Zimmerman *et al.* (1977). The phenylthiohydantoin amino acids were quantified at 265 nm by a Spectra Physics SP-4100 computing integrator; serine and threonine residues were confirmed by the detection of their dehydro derivatives at 313 nm. The yield of the phenylthiohydantoin amino acid released at the second cycle of each sequence run was calculated. The efficiency of coupling was assessed for five samples by determining the uncoupled protein. In three samples the beads were hydrolysed in 6 M-HCl at 110 °C for 24 h and the residual coupled protein was determined by amino acid analysis. The coupling efficiencies were in the range 76–81%.

RESULTS AND DISCUSSION

N-Terminal sequence of different preparations of endopeptidase-24.11

A total of twelve samples of the endopeptidase were sequenced. Fig. 1 shows the consensus sequence of 27 amino acid residues, derived from four types of sample. The sequence of untreated samples fell into two patterns: in five instances it started at residue 1 (never identified), but in two it commenced at Lys⁶. The latter two samples had been purified on a GK7C2 column, whereas the

former five were isolated by a different monoclonal antibody, GK4A9. For reasons which are not clear, it seems that the use of the GK7C2 column, which required several runs for each sample, may have permitted limited proteolysis to occur at the Lys⁵–Lys⁶ bond. Curiously, the same limited proteolytic attack was noted in two samples treated with proteinase V8. Since this proteinase is known to hydrolyse only at Glu–Xaa bonds (Drapeau *et al.*, 1972), the hydrolysis at Lys⁵–Lys⁶ must have been due to the action of another proteinase (not in this case attributable to the use of GK7C2, since these samples were purified on a GK4A9 column). The effect of trypsin was to shorten the polypeptide chain so that these samples sequenced from the Trp¹⁰ residue resulting from the tryptic hydrolysis of the Arg⁹–Trp¹⁰ bond. Hydrolysis adjacent to the four lysine residues may also have occurred, but none of these fragments were sought for. We had previously reported some minor changes in the properties of the detergent-treated enzyme after trypsin treatment (Fulcher & Kenny, 1983). The treated protein retained its amphipathic properties, but exhibited a lower pI on isoelectric focusing and on ion-exchange chromatography. The loss of five basic amino acid residues would adequately account for these observations. The identification of two truncated forms of the enzyme emphasizes the susceptibility of this region (residues 1–9) to proteolytic attack.

We cannot be certain that our sequence has started at the *N*-terminal residue. Earlier attempts (results not shown) to sequence this protein (using 20–40 nmol samples) by a Beckman 890c spinning-cup sequencer, gave the same sequence for three trypsin-treated samples (i.e. starting at Trp¹⁰) and the initial sequencing yields were at best 20%. Two untreated samples were also run; one gave no sequence, the other commenced at Lys⁶, but the yields at each cycle were at best 5%. These observations had raised the possibility that the *N*-terminal residue might have been acylated in the native enzyme. The sequences reported here were all obtained on samples covalently coupled to glass. The yield was assessed at the second cycle; the mean for all runs was $19.6 \pm 1.6\%$ (S.E.M.) and the range 15–30%. For the five runs where Pro² was the residue assessed (i.e. starting at the putative *N*-terminus), the yield was $19.2\% \pm 1.3\%$, and for the seven runs where the sequence began at Lys⁶ or Trp¹⁰ the

Fig. 2. Homology of the *N*-terminal sequence of endopeptidase-24.11 with that of isomaltase

The sequences of the *N*-terminal regions have been aligned to reveal maximum homology. Identical residues are 'boxed' in with continuous lines. Conservative substitutions are 'boxed' in with dotted lines. Data for rabbit isomaltase are from the cDNA-deduced sequence for pro-sucrase-isomaltase (Hunziker *et al.*, 1986); that for rat isomaltase is from Hauri *et al.* (1982) and that for rabbit aminopeptidase N from Feracci *et al.* (1982). The *N*-terminal sequence of pig isomaltase (Sjöström *et al.*, 1982) is identical with that of rabbit, except for the following substitutions: Lys³→Arg; Arg⁴→Lys; Val¹⁸→Ala; Phe²¹→Leu; Ile²²→Ser. Had it been added to the Figure, the Lys⁷ box would have been drawn in with a dotted line.

Fig. 3. Hydropathy plots of endopeptidase-24.11 and the homologous segment of isomaltase

The hydropathy values of the homologous segments [(a), 25 residues for endopeptidase-24.11 and (b) 22 for pro-sucrase-isomaltase] have been computed as described by Kyte & Doolittle (1982) using a window size of 11 residues. The unidentified residue 1 of endopeptidase-24.11 was given an arbitrary identification as Ala.

yield was $19.9 \pm 2.7\%$. The second-cycle residue on each of these runs (Thr, Lys and Pro) happens to be somewhat labile in the conversion procedure, possibly covalently linked to glass or inefficiently released in the degradation

step. These considerations would suggest that the potential yield for 'uncomplicated' residues would be considerably higher than the 20% observed; hence it seems unlikely that the *N*-terminus of the native protein is acylated.

Homology with other brush-border hydrolases

In Fig. 2 the sequences of the isomaltase subunit of rabbit intestinal sucrase-isomaltase (Frank *et al.*, 1978; Hunziker *et al.*, 1986) and that from rat intestine (Hauri *et al.*, 1982) and aminopeptidase N from rabbit intestine (Feracci *et al.*, 1982) are aligned with that endopeptidase-24.11 to reveal maximum homology. For reasons of clarity, the sequence of pig isomaltase (Sjöström *et al.*, 1982) has been omitted from Fig. 2. The only difference from that of rabbit which would affect the homology shown is the reversal of the order of Lys³–Arg⁴. All three microvillar hydrolases show some general similarities: a positively charged *N*-terminal sequence leading either directly (in the case of aminopeptidase N), or via an intermediate sequence (in other cases), to a strongly hydrophobic sequence.

The homology between aminopeptidase N and the other two enzymes is not impressive, being confined to three hydrophobic residues. That between endopeptidase-24.11 and isomaltase is worthy of further comment. There are seven identical residues over a 14-residue stretch and there are four conservative substitutions in the same stretch (Arg/Lys, Trp/Phe, Thr/Ser, Ser/Thr). However, the substitution of Glu (residue 18) in the endopeptidase for Ile in isomaltase is non-conservative. It is clear from the full sequence of isomaltase, derived from the cDNA cloning, that the only potential transmembrane domain (comprising 20 residues) starts

one residue before this Ile (Hunziker *et al.*, 1986). If our hydrophobic sequence, starting with Val¹⁹, also represents the transmembrane domain of endopeptidase-24.11, the anchor starts two residues later than that of isomaltase. However, we were able to identify positively only seven consecutive hydrophobic residues before the yields became too low.

They hydropathy plots (Kyte & Doolittle, 1982) of both sequences are shown in Fig. 3. They reinforce the comments above on the general similarities of the *N*-terminal segments of these proteins and suggest that the transmembrane domain of endopeptidase-24.11 might begin at residue 12.

The identity of the sequence Leu¹³-Glu-Ile-Ser-Leu¹⁷ with residues 9–11 of isomaltase (except for the substitution of Ser for Thr) led us to an unsuccessful search for any evidence of carbohydrate moieties attached to the Ser residue, which in earlier runs has not been positively identified. This was relevant because the corresponding residue in isomaltase (only later identified as Thr) appeared to be glycosylated in both the rabbit (Frank *et al.*, 1978) and pig enzymes (Sjöström *et al.*, 1982). This information was crucial to the model then proposed for the membrane insertion of sucrase-isomaltase, since it placed the *N*-terminus at the cell surface and required that there should be two transmembrane sequences. However, the full-length sequence revealed only one hydrophobic domain of 20 residues that could serve as an anchor, and the attribution of carbohydrate chains to Thr¹² was withdrawn (Hunziker *et al.*, 1986).

The cDNA-derived sequences of pro-sucrase-isomaltase also excluded the existence of a cleaved signal sequence. It was therefore argued that the *N*-terminal segment, including the hydrophobic anchor of the mature protein, serves the function of a signal peptide, which in this instance is not cleaved by signal peptidase (Hunziker *et al.*, 1986). This mode of biogenesis for those brush-border hydrolases known to be anchored by a region near the *N*-terminus, which had been proposed much earlier in a rather speculative way (Kenny &

Booth, 1978; Semenza, 1979), has now been confirmed in respect of pro-sucrase-isomaltase as well as for γ -glutamyltransferase (Laperche *et al.*, 1986). The extent of homology we have observed with isomaltase suggests that the model may well apply to endopeptidase-24.11

REFERENCES

- Drapeau, G. R., Boily, Y. & Houmard, J. (1972) *J. Biol. Chem.* **247**, 6720–6728
- Feracci, H., Maroux, S., Bonicel, J. & Desnuelle, P. (1982) *Biochim. Biophys. Acta* **684**, 133–136
- Frank, G., Brunner, J., Hauser, H., Wacker, H. & Semenza, G. (1978) *FEBS Lett.* **96**, 182–187
- Fulcher, I. S. & Kenny, A. J. (1983) *Biochem. J.* **211**, 743–753
- Gee, N. S. & Kenny, A. J. (1985) *Biochem. Soc. Trans.* **13**, 101–102
- Gee, N. S., Matsas, R. & Kenny, A. J. (1983) *Biochem. J.* **214**, 377–386
- Gee, N. S., Bowes, M. A., Buck, P. & Kenny, A. J. (1985) *Biochem. J.* **228**, 119–126
- Hauri, H. P., Wacker, H., Rickli, E. E., Bigler-Meier, B., Quaroni, A. & Semenza, G. (1982) *J. Biol. Chem.* **257**, 4522–4528
- Hunziker, W., Speiss, M., Semenza, G. & Lodish, H. F. (1986) *Cell* **46**, 227–234
- Kenny, A. J. & Booth, A. G. (1978) *Essays Biochem.* **14**, 1–44
- Kenny, A. J. & Maroux, S. (1982) *Physiol. Rev.* **62**, 91–128
- Kenny, A. J., Fulcher, I. S., McGill, K. A. & Kershaw, D. (1983) *Biochem. J.* **211**, 755–762
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132
- Laperche, Y., Bulle, F., Aissani, T., Chobert, M. N., Aggerbeck, M., Hanoune, J. & Guellaen, G. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 937–941
- Laursen, R. A. (1971) *Eur. J. Biochem.* **20**, 89–102
- Semenza, G. (1979) *Ciba Found. Symp.* **70**, 133–146
- Semenza, G. (1986) *Annu. Rev. Cell Biol.* **2**, in the press
- Sjöström, H., Norén, O., Christiansen, L. A., Wacker, H. & Spiess, M. (1982) *FEBS Lett.* **148**, 321–325
- Wachter, E., Machleidt, W., Hofner, H. & Otto, J. (1973) *FEBS Lett.* **35**, 97–102
- Zimmerman, C. L., Appella, E. & Pisano, J. J. (1977) *Anal. Biochem.* **77**, 569–573