Homologues of insulinase, a new superfamily of metalloendopeptidases

Neil D. RAWLINGS* and Alan J. BARRETT

Department of Biochemistry, Strangeways Research Laboratory, Worts Causeway, Cambridge CB1 4RN, U.K.

On the basis of a statistical analysis of an alignment of the amino acid sequences, a new superfamily of metalloendopeptidases is proposed, consisting of human insulinase, *Escherichia coli* protease III and mitochondrial processing endopeptidases from *Saccharomyces* and *Neurospora*. These enzymes do not contain the 'HEXXH' consensus sequence found in all previously recognized zinc metalloendopeptidases.

INTRODUCTION

The majority of metalloendopeptidases (metalloproteinases) bind an essential zinc atom at a -His-Glu-Xaa-Xaa-His- (or HEXXH) site similar to that found in thermolysin (Jongeneel *et al.*, 1989). The purpose of the present paper is to draw attention to the existence of a separate group of metalloendopeptidases that are structurally unrelated to thermolysin, and do not contain the HEXXH consensus sequence. This group contains human insulinase, *Escherichia coli* protease III and endopeptidases concerned in the processing of precursor proteins as they enter mitochondria.

Human insulinase is an endopeptidase that acts on insulin, glucagon and other polypeptides up to about M_r 7000 (Kirschner & Goldberg, 1981). Activity is inhibited by EDTA and 1,10-phenanthroline (Goldberg *et al.*, 1981; Kayalar & Wong, 1989), indicating that the enzyme is a metalloendopeptidase. The reported inhibition by *N*-ethylmaleimide and some other thiol-blocking reagents (Ansorge *et al.*, 1984) is not inconsistent with the enzyme being a metalloendopeptidase, also being seen with thimet peptidase (EC 3.4.24.15) (Barrett & Brown, 1990). The amino acid sequence of human insulinase was determined from the cDNA by Affholter *et al.* (1988), who showed that the enzyme is homologous with an enzyme from *E. coli* known as protease III (Finch *et al.*, 1986) or protease Pi (Goldberg *et al.*, 1981). [In the present paper the terms 'homology', 'homologue' and 'homologous' are used exclusively to indicate evolutionary relationship as defined by Reeck *et al.* (1987).] The bacterial enzyme also is inhibited by chelating agents, and can be re-activated with Zn^{2+} (Cheng & Zipser, 1979). The specificity of protease III resembles that of insulinase, as does that of an enzyme from *Drosophila* that also has a similar *N*-terminal amino acid sequence (Garcia *et al.*, 1988).

Mitochondrial endopeptidases that remove the leader peptide from proteins transported into the mitochondria are known from a variety of species, including rat, ox, the yeast *Saccharomyces cerevisiae* and the ascomycete *Neurospora crassa*; these enzymes are known to be inhibited by EDTA and 1,10-phenanthroline, and to be re-activated by bivalent cations (Miura *et al.*, 1982; McAda & Douglas, 1982; Böhni *et al.*, 1983; Kumamoto *et al.*, 1986; Hawlitschek *et al.*, 1988; Schneider *et al.*, 1990). In yeast and *Neurospora*, two enzymes are involved in processing the leader peptides of mitochondrial proteins, the

matrix-processing peptidase and the processing-enhancing protein, which act sequentially (Schmidt *et al.*, 1984; Tropschug *et al.*, 1988). Similar two-step processing events have been observed in the rat, mouse and human mitochondria (Sztul *et al.*, 1987, 1988; Kalousek *et al.*, 1988; Schmidt *et al.*, 1984).

Neurospora processing-enhancing protein is also subunit II of the ubiquinol-cytochrome *c* reductase complex (Schulte *et al.*, 1989), and the sequences of subunits I and II of this complex from the yeast are also homologous (reviewed by Weiss *et al.*, 1990).

METHODS

The FASTP program of Lipman & Pearson (1985) was used to compare sequences for possible relationships, with the KTUP value set to 1. The RDF program from the same source was used to test the statistical significance of possible homologies, with the number of randomizations of the shuffled sequence set to 100, and the KTUP value set to 1. An aligned score of 6.0 standard deviation units above the mean score of the alignments with the shuffled sequences was taken as proof of an evolutionary sequence relationship, that is to say homology.

Sequences were aligned progressively by application of the rules of Doolittle (Feng & Doolittle, 1990).

RESULTS AND DISCUSSION

Fig. 1 shows the alignment of the *N*-terminal portions of the insulinase and protease III molecules with the mitochondrial proteins; insulinase and protease III have long *C*-terminal extensions that are not shown.

Table 1 shows the statistical significance of the similarities between the sequences as calculated with the Lipman & Pearson (1985) FASTP program. These confirm the evolutionary relationship of insulinase with *E. coli* protease III, and of the mitochondrial processing endopeptidases with the cytochrome *c* reductase subunits, but, in addition, significant relationships are seen for the alignments of insulinase and protease III with *Saccharomyces* processing-enhancing protein and *Neurospora* processing-enhancing protein. It follows from this that all of these proteins are members of a single superfamily, in the terminology of Dayhoff *et al.* (1983).

*To whom correspondence should be addressed.

Fig. 1. Alignment of sequences of members of the insulinase superfamily

Sequences are numbered according to that of insulinase, and are truncated at residue 539; an alignment of the C-termini of insulinase and protease III is given by Affholter *et al.* (1988). – indicates an insertion in a sequence, and * at position 292 + 1 in *Neurospora* matrix-processing peptidase represents a segment of 66 residues that is not found in the other sequences. Residues identical with those in insulinase are printed in white on black. The abbreviations and references to sequence data are as follows: INSHU, human insulinase (Affholter *et al.*, 1988); PIII, *E. coli* protease III (Finch *et al.*, 1986; but with residues 391–398 as described by Clavarié-Martin *et al.*, 1987); YMPP, *Saccharomyces* matrix-processing peptidase (Pollock *et al.*, 1988; Jensen & Yaffe, 1988); YPEP, *Saccharomyces* processing-enhancing protein (Witte *et al.*, 1988); NMPP, *Neurospora* matrix-processing peptidase (Schneider *et al.*, 1990); NPEP, *Neurospora* processing-enhancing protein/cytochrome *c* reductase subunit I (Schulte *et al.*, 1989); YCYTCI, *Saccharomyces* cytochrome *c* reductase subunit I (Tzagoloff *et al.*, 1986); YCYTCII, *Saccharomyces* cytochrome reductase subunit II (Oudshoorn *et al.*, 1987).

Table 1. Comparison of the amino acid sequences of proteins of the insulinase superfamily

The values at lower left are percentages of identical residues in the alignment of Fig. 1. The values at upper right are the scores (SD units) from the FASTP program, values above 6.00 being printed in **bold italic** type. Abbreviations are as described in the legend to Fig. 1.

Test sequence	Shuffled sequence							
	INSHU	PIII	YMPP	YPEP	NMPP	NPEP	YCYTCI	YCYTCII
INSHU	—	48.23	1.67	6.28	1.70	7.24	1.59	0.98
PIII	25	—	0.34	10.38	1.46	9.47	0.59	1.49
YMPP	12	8	—	42.92	50.17	37.28	12.10	11.90
YPEP	15	15	25	—	10.53	98.67	30.93	11.46
NMPP	10	9	43	21	—	21.67	7.14	9.23
NPEP	16	13	25	52	23	—	45.04	10.45
YCYTCI	8	8	17	23	15	32	—	1.02
YCYTCII	10	10	15	15	11	17	14	—

The enzymes of the insulinase superfamily do not contain the HEXXH sequence that contributes to the binding of zinc in all previously characterized metalloendopeptidases, and further experimental work will be required to identify the metal-binding residues in these proteins. The only amino acids previously recognized as zinc-binding are histidine, glutamic acid and cysteine (Vallee & Auld, 1990). For the insulinase homologues, cysteine can quickly be eliminated, because protease III has only one cysteine residue, which is in its C-terminal extension (residue 764). Conserved histidine residues occur at positions 108, 213 and 264, and conserved glutamic acid residues at 133, 145 and 189, but in all cases these are missing from one or two of the sequences. If the yeast cytochrome *c* reductase subunits I and II are not endopeptidases, then His-108, His-213 and Glu-189 are candidate metal ligands for the others. However, these do not show the 'short/long' spacing characteristic of ligands of the great majority of zinc metalloenzymes, including the thermolysin-related endopeptidases (Vallee & Auld, 1990). Whether the essential metal of the enzymes of the insulinase superfamily is zinc or another metal, the involvement of ligands other than histidine, glutamic acid and cysteine clearly cannot be excluded.

Note added in proof (received 15 February 1991)

Since the submission of this paper, the sequences of two more members of the insulinase superfamily have been published. These are an insulin-degrading enzyme from *Drosophila melanogaster* (Kuo *et al.*, 1990) and a mitochondrial processing endopeptidase from rat liver (Kleiber *et al.*, 1990). Both of these sequences retain the amino acids proposed above to be the zinc-binding residues (His-108, Glu-189 and His-213, human insulinase numbering). Kleiber *et al.* (1990) propose an alternative zinc-binding site, namely His-108, Glu-111 and His-112 (a reversal of the thermolysin-like HEXXH motif), but this is unlikely because His-112 is absent from all three mitochondrial processing proteinases.

REFERENCES

- Affholter, J. A., Fried, V. A. & Roth, R. A. (1988) *Science* **242**, 1415–1418
- Ansorge, S., Bohley, P., Kirschke, H., Langner, J. & Wiedersanders, B. (1984) *Biomed. Biochim. Acta* **43**, 39–46
- Barrett, A. J. & Brown, M. A. (1990) *Biochem. J.* **271**, 701–706
- Böhni, P. C., Daum, G. & Schatz, G. (1983) *J. Biol. Chem.* **258**, 4937–4943
- Cheng, Y.-S. Y. Zipser, D. (1979) *J. Biol. Chem.* **254**, 4698–4706
- Claverie-Martin, F., Diaz-Torres, M. R. & Kushner, S. R. (1987) *Gene* **54**, 185–195
- Dayhoff, M. O., Barker, W. C. & Hunt, L. T. (1983) *Methods Enzymol.* **91**, 524–545
- Feng, D.-F. & Doolittle, R. F. (1990) *Methods Enzymol.* **183**, 375–387
- Finch, P. W., Wilson, R. R., Brown, K., Hickson, I. D. & Emmerson, P. T. (1986) *Nucleic Acids Res.* **14**, 7695–7703
- Garcia, J. V., Fenton, B. W. & Rosner, M. R. (1988) *Biochemistry* **27**, 4237–4244
- Goldberg, A. L., Sreedhara Swamy, K. H., Chung, C. H. & Larimore, F. S. (1981) *Methods Enzymol.* **80**, 680–702
- Hawltischek, G., Schneider, H., Schmidt, B., Tropschug, M., Hartl, F.-U. & Neupert, W. (1988) *Cell* **53**, 795–806
- Jensen, R. E. & Yaffe, M. P. (1988) *EMBO J.* **7**, 3863–3871
- Jongeneel, C. V., Bouvier, J. & Bairoch, A. (1989) *FEBS Lett.* **242**, 211–214
- Kayalar, C. & Wong, W. T. (1989) *J. Biol. Chem.* **264**, 8928–8934
- Kalousek, F., Hendrick, J. P. & Rosenberg, L. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 7536–7540
- Kirschner, R. J. & Goldberg, A. L. (1981) *Methods Enzymol.* **80**, 702–711
- Kleiber, J., Kalousek, F., Swaroop, M. & Rosenberg, L. E. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 7978–7982
- Kumamoto, T., Ito, A. & Omura, T. (1986) *J. Biochem. (Tokyo)* **100**, 247–254
- Kuo, W. L., Gehm, B. D. & Rosner, M. R. (1990) *Mol. Endocrinol.* **4**, 1580–1591
- Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441
- McAda, P. C. & Douglas, M. G. (1982) *J. Biol. Chem.* **257**, 3177–3182
- Miura, S., Mori, M., Amaya, Y. & Tatibana, M. (1982) *Eur. J. Biochem.* **122**, 641–647
- Oudshoorn, P., Van Steeg, H., Swinkels, B. W., Schoppink, P. & Grivell, L. A. (1987) *Eur. J. Biochem.* **163**, 97–103
- Pollock, R. A., Hartl, F.-U., Cheng, M. Y., Ostermann, J., Horwich, A. & Neupert, W. (1988) *EMBO J.* **7**, 3493–3500
- Reeck, G. R., de Haën, C., Teller, D. C., Doolittle, R. F., Fitch, W. M., Dickerson, R. E., Chambon, P., McLachlan, A. D., Margoliash, E., Jukes, T. H. & Zuckerhandl, E. (1987) *Cell* **50**, 667
- Schmidt, B., Wachter, E., Sebald, W. & Neupert, W. (1984) *Eur. J. Biochem.* **144**, 581–588
- Schneider, H., Arretz, M., Wachter, E. & Neupert, W. (1990) *J. Biol. Chem.* **265**, 9881–9887
- Schulte, U., Arretz, M., Schneider, H., Tropschug, M., Wachter, E., Neupert, W. & Weiss, H. (1989) *Nature (London)* **339**, 147–149
- Sztul, E. S., Hendrick, J. P., Kraus, J. P., Wall, D., Kalousek, F. & Rosenberg, L. E. (1987) *J. Cell Biol.* **105**, 2631–2639
- Sztul, E. S., Chu, T. W., Strauss, A. W. & Rosenberg, L. E. (1988) *J. Biol. Chem.* **263**, 12085–12091
- Tropschug, M., Nicholson, D. W., Hartl, F.-U., Köhler, H., Pfanner, N., Wachter, E. & Neupert, W. (1988) *J. Biol. Chem.* **263**, 14433–14440
- Tzagoloff, A., Wu, M. & Crivellone, M. (1986) *J. Biol. Chem.* **261**, 17163–17169
- Vallee, B. L. Auld, D. S. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 220–224
- Weiss, H., Leonard, K. & Neupert, W. (1990) *Trends Biochem. Sci.* **15**, 178–180
- Witte, C., Jensen, R. E., Yaffe, M. P. & Schatz, G. (1988) *EMBO J.* **7**, 1439–1447