Methods in Protein Sequence Analysis
Jörnvall/Höög/Gustavsson (eds.)

THE YEAST PROHORMONE-PROCESSING KEX2 PROTEASE, AN ENZYME WITH SPECIFICITY FOR PAIRED BASIC RESIDUES

R.S. Fuller, C. Brenner, P. Gluschankof, and C.A. Wilcox

Department of Biochemistry, Beckman Center for Molecular and Genetic Medicine, Stanford University School of Medicine, Stanford, CA, 94305-5307, U.S.A.

SUMMARY: The Kex2 protease of the yeast *Saccharomyces cerevisiae* and two recently discovered human homologues define a novel subfamily of subtilisin-related serine proteases that may be responsible for cleaving prohormones and other precursor polypeptides at paired basic residues in the eukaryotic secretory pathway. Here, the pathway of post-translational modification of Kex2 protease, its domain structure and primary sequence specificity are discussed.

INTRODUCTION

Widespread application of the methods of Molecular Biology has led to a vast increase in the data base of protein primary structures. In the absence of information about post-translational modifications that give rise to the final, mature form of a gene product, an amino acid sequence inferred from nucleotide sequence is incomplete. Only direct chemical determination of N-terminal and C-terminal amino acid sequence, and of the nature and location

of covalent modifications can define primary structure unambiguously. Nevertheless, knowledge of the enzymology and physiology of post-translational modifications can be helpful when predicting the structure of a polypeptide known from nucleotide sequence data alone.

Proteolytic Processing at Paired Basic Residues: A common post-translational modification of polypeptides targeted to the secretory pathway of eukaryotic cells is endoproteolytic cleavage, most often at the carboxyl side of a pair of basic residues, especially Lys-Arg and Arg-Arg. Diverse molecules undergo such processing, including precursors of both secreted peptides and proteins [such as peptide hormones, neuropeptides, growth factors, the vitamin K-dependent coagulation factors, and serum albumins (for reviews, see Sossin et al., 1989; Andrews et al., 1989; Furie and Furie, 1988)] and integral plasma membrane proteins such as the insulin receptor (Yoshimasa et al., 1988) and several viral envelope glycoproteins (Perez and Hunter, 1987; Klenk and Rott, 1988). Although conclusive identification of enzymes catalyzing such reactions in higher eukaryotic cells has proven difficult, the Kex2 protease of the yeast *Saccharomyces cerevisiae* has been shown on the basis of both genetic and biochemical criteria to be required *in vivo* for processing both pro-α-factor, precursor of the α-mating factor, and pro-killer toxin (Julius et al., 1984; reviewed in Fuller et al., 1988). Kex2 protein, partially purified from overproducing cells, is a transmembrane, Ca^{2+}-dependent, neutral serine protease that specifically cleaves peptide substrates at the carboxyl side of Lys-Arg and Arg-Arg dipeptides (Fuller et al., 1986; Fuller et al., 1989a; Mizuno et al., 1989). Kex2 protease can process mammalian precursors such as proinsulin (Thim et al., 1986) and pro-opiomelanocortin (Thomas et al., 1989) accurately, and the existence of human homologues of Kex2 protease (Fuller et al., 1989b; van den Ouweland et al., 1990; Smeekens and Steiner, 1990) suggests conservation of this enzymatic mechanism throughout eukaryotic evolution.

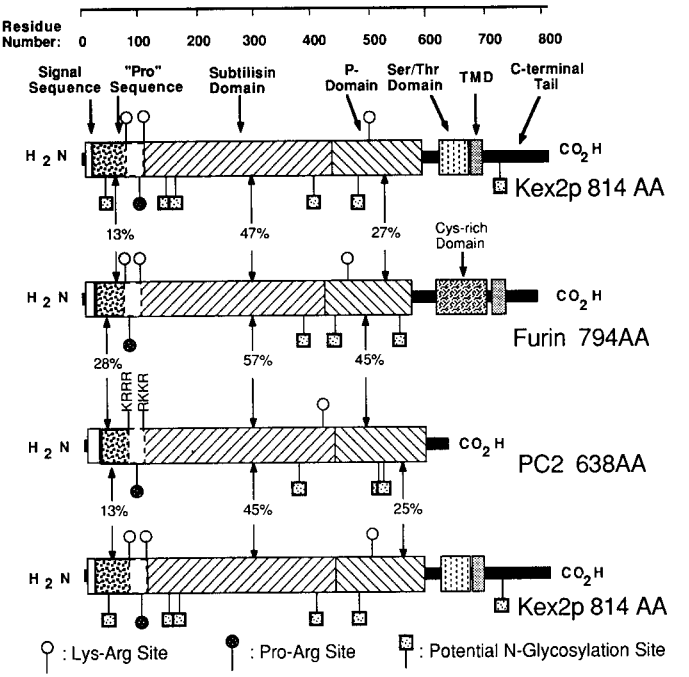

Figure 1. Schematic comparison of Kex2 protein and human homologues. TMD, transmembrane domain; AA, amino acids. Percentage identity between domains is noted by the arrows.

Structural Features of Kex2 Protein (See Fig. 1): The sequence of the *KEX2* gene predicts an 814 amino acid polypeptide with an N-terminal signal peptide and a single transmembrane domain near the C-terminus followed by a highly charged, 115 residue C-terminal cytosolic "tail" (Fuller et al., 1988, Mizuno et al., 1988) that is required for intracellular retention of the protein within a late compartment of the yeast Golgi apparatus (Fuller et al., 1989b; K. Redding and R.S. Fuller, submitted). A C-terminal deletion that leaves just the first 614 residues of the molecule intact still possesses Ca^{2+}-dependent proteolytic activity (Fuller, et al., 1989a). Within this N-terminal region lies the likely catalytic domain, a 295 amino acid sequence ("Subtilisin Domain") that is about 30% identical to members of the subtilisin class of serine proteases (Fuller et al., 1988, Mizuno et al., 1988, Fuller et al., 1989b). About 120 residues intervene between the signal

peptide and the subtilisin domain. In subtilisin, an analogous region comprises a "pro-" sequence that is cleaved autoproteolytically upon secretion (Power et al., 1986; Ikemura and Inouye, 1989). Two Lys-Arg dipeptides are found at residues 79-80 and 108-109 in Kex2 protein, providing possible sites for autoproteolytic activation. Mature Kex2 protein is a 125-135 kDal glycoprotein that contains only 3-5 kDal of Asn-linked (N-linked) oligosaccharide, but a substantial amount of Ser/Thr-linked (O-linked) carbohydrate (Fuller et al., 1989b), perhaps primarily in a Ser/Thr rich region just N-terminal to the transmembrane domain (Fuller et al., 1989a).

RESULTS AND DISCUSSION

Pathway of Maturation of Kex2 Protease: The pathway of post-translational modification and maturation of Kex2 protein in the yeast secretory pathway is summarized in Fig. 2. The primary translation product of the *KEX2* gene, immunoprecipitated from an *in vitro* translation of polyA$^+$ RNA from a Kex2-overproducing strain, migrates in SDS-PAGE at 110 kDal rather than the predicted size of 90 kDal, probably due to the highly acidic C-terminal cytosolic tail (Fuller et al. 1989a). The inferred primary sequence, along with homology to subtilisin, suggested that Kex2 protein is synthesized as a pre-pro-protein. Hypothetically, prepro-Kex2 protein should undergo signal peptide cleavage, followed by cleavage of the N-terminal pro-peptide, likely by autoproteolysis at the Lys-Arg sites mentioned previously. An increase in the molecular weight of Kex2 protein in a strain containing a temperature-sensitive mutation in signal peptidase (Böhni et al., 1988) indicated that the signal peptide of Kex2 protein is normally cleaved. Rapid pulse-chase labeling (pulse time ~1 min) of cells using a mixture of ^{35}S-labeled amino acids, followed by immunoprecipitation of labeled Kex2 protein, allowed observation of short-lived intermediates in the endoplasmic reticulum (ER). Because the antibody used was directed against the C-terminal 100 residues of Kex2 protein, the rapid conversion

($t_{1/2}$ ~ 1 min) of the earliest form visulalized, I_1, (125 kDal) into a second species, I_2, (~115 kDal) could be attributed to N-terminal proteolytic cleavage. The fact that cleavage occurred in the ER was established by analysis of a conditional secretory mutant (*sec18*) in which transport from the ER to the Golgi complex is blocked at the non-permissive temperature (Kaiser and Schekman, 1990).

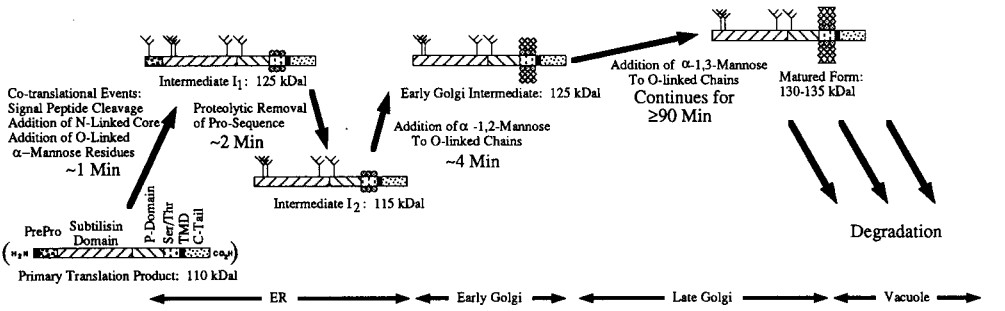

Figure 2. Pathway of maturation of Kex2 protein. Refer to text for details.

From its size relative to the primary translation product, I_1 already has undergone extensive modification, presumably co-translationally. The increase in apparent molecular weight between I_1 and the primary translation product is due to addition of both N-linked and O-linked carbohydrated, as demonstrated by the effects of the glycosylation inhibitors tunicamycin and 2-deoxyglucose and by digestion with endoglycosidases. Transport of species I_2 to the Golgi complex is marked by rapid conversion ($t_{1/2}$ ~ 2-3 min) to a slower migrating form (Early Golgi Intermediate, ~125 kDal), principally by addition of mannose in an α-1,2-linkage to the O-linked mannose residues added in the ER (Kukuruzinska et al., 1987). The mature protein, localized to late Golgi structures (K. Redding and R.S. Fuller, submitted), undergoes a gradual increase in molecular weight. This modification is

probably due to addition of terminal α-1,3-linked mannose residues to the O-linked chains because it is blocked by the *mnn1* mutation (Kukuruzinska et al., 1987). Ultimately, Kex2 protein is degraded in the vacuole, a process which is blocked by mutations in the *PEP4* and *PRB1* genes, which encode the major vacuolar proteases (Jones, 1984).

Kex2 Protease and Its Human Homologues Define a Subfamily of Subtilisin-Related Proteases: The recent discovery of human (Fuller et al., 1989b, Smeekens and Steiner, 1990) and mouse (Seidah et al., 1990) genes that encode polypeptides homologous to Kex2 protein strengthens the likelihood that a homologous family of enzymes may be involved in precursor processing in eukaryotes. A schematic comparison of Kex2 protein with the two human sequences is presented in Fig. 1. Like Kex2 protein, furin and PC2 both have likely N-terminal signal peptides and a potential "pro" sequence punctuated by paired basic or multi-basic sites prior to a domain of ~300 residues that is about 30% identical to subtilisin. The greatest similarities between Kex2 protein and the human homologues are in the subtilisin domains, in which furin and PC2, respectively, are about 47% and 45% identical to Kex2 protein. Beyond the subtilisin domain, each sequence contains an additional ~ 155 residues ("P-domain") of somewhat lower homology that is not found in other members of the subtilisin family. In the subtilisin domain and P-domains, the mammalian sequences are somewhat more homologous to one another than to Kex2 protein.

Aside from Kex2 protease (and possibly the mammalian homologues), the subtilisin family comprises a group of relatively non-specific degradative enzymes, found both in prokaryotes (the intracellular and secreted subtilisins of the *Bacilli* and thermitase from *Thermoactinomyces vulgaris*) and in eukaryotes (proteinase K from *Tritarachium album* Limber and vacuolar protease B from *Saccharomyces cerevisiae*) (Moehle et al., 1987). Thus it is striking that within their subtilisin domains, Kex2 protease, furin and PC2 are all more homologous to one another than any of the three is to the degradative subtilisins. Classification of Kex2 protease, furin and PC2 as a subtilisin subfamily rests on

three additional arguments. First, there is conservation among the three of specific residues near the critical catalytic residues that differ from the degradative subtilisin consensus (Fuller et al. 1989b; Smeekens and Steiner, 1990). Second, each has pairs or other multiples of basic residues just prior to the subtilisin domain. These may be sites for autoproteolytic activation and therefore may provide information about the primary sequence specificities of the putative proteases. Finally, as described below, the P-domain represents a motif that is unique to these three sequences and which has been found to be essential for function in the case of Kex2 protease.

Minimal Sequences Required for Proteolytic Activity: To map the C-terminal endpoint of the active site domain of Kex2 protein, random C-terminal deletions of the *KEX2* structural gene were created using *Bal31* nuclease digestion, beginning at the endpoint of the most extensive existing deletion (amino acid 614). The expectation was that the functional endpoint would correspond to the end of the subtilisin domain (approximately residue 438). Deletion mutants were analyzed by expression in yeast using both bioassays indicative of Kex2 function *in vivo* (production of mature α-factor) and direct biochemical assay of secreted proteolytic activity using the fluorogenic peptide substrate, boc-Gln-Arg-Arg-MCA. The smallest active deletion encoded 593 residues of Kex2 sequences plus 5 additional residues, while the largest inactive deletion encoded 592 residues of Kex2 sequence plus two additional residues. Remarkably the deletion endpoint lay 155 residues beyond the end of the subtilisin domain (at Tyr_{438}), precisely at the endpoint of homology between Kex2 protein and the human homologues (see Fig. 1). We refer to this 155 residue region as the "P-domain" because of its presence in this subfamily of subtilisin homologues which may be involved in processing.

By placing the gene encoding the secreted, soluble form of Kex2 protease under the control of the powerful, constitutive *TDH3* promoter on a multicopy vector, pAB23 (Fuller et al., 1989a), the active protease was secreted into culture medium at 1-2 mg/l.

This minimally-sized active form of the enzyme has been purified to near homogeneity. Future studies will concentrate on kinetics of interaction of the purified enzyme with a variety of substrates.

Analysis of substrate sequence specificity of Kex2 protease by an *in vivo* assay: Previous studies have suggested that the principal determinants of specificity for Kex2 protease are the P_1 and P_2 substrate residues (Julius et al., 1984b; Achstetter and Wolf, 1985; Fuller et al., 1989a; Mizuno et al., 1989). Kex2 protease cleaves C-terminally to Lys-Arg and Arg-Arg sites with similar kinetics in small peptides (Fuller et al., 1989a), but cleaves substrates containing Lys-Lys or a single Arg residue poorly (Julius et al., 1984b). While only Lys-Arg sites are found in pro-α-factor, pro-killer toxin contains both Lys-Arg and Arg-Arg sites, and, in addition, a Pro-Arg site at which processing may also depend on Kex2 protease (Fuller et al., 1988).

To complement purely biochemical approaches, a system has been developed to characterize the substrate specificity of Kex2 protease *in vivo*. Mating of haploid yeast cells of the α-mating type with haploid cells of the **a**-mating type requires that α-cells secrete fully processed α-factor. Mating can be measured quantitatively over 6 or 7 orders of magnitude (Hartwell, 1980), providing a sensitive assay for processing of pro-α-factor by Kex2 protease *in vivo*.

Cassette mutagenesis (Wells, 1985) was used to create a library of mutations at the P_2 site in pro-α-factor, and the effects of individual mutations on the relative efficiency of cleavage by Kex2 protease were determined by the quantitative mating assay. The following relative order of cleavage preference by Kex2 protease was observed: Lys-Arg = 1.0 >> (Thr-Arg > Pro-Arg > Ile-Arg > Asn-Arg) = 10^{-2}-10^{-4} >> (Phe-Arg, Leu-Arg) ≤ 4×10^{-6}. The Phe-Arg and Leu-Arg substitutions were as inactive as negative controls, suggesting that these molecules were not cleaved at all by Kex2 protease. Mating by strains containing the Thr-Arg, Pro-Arg, Ile-Arg, and Asn-Arg substitutions, although feeble, was

dependent on Kex2 protease, as could be shown by testing strains deleted of the *KEX2* gene.

These data are consistent with a stringent requirement for Lys-Arg (or Arg-Arg) for efficient processing. No simple picture emerges for substitutions that are measurably, but weakly, cleaved, although certain observations can be made. With the β-branched residues Thr and Ile at P_2, some cleavage is observed, whereas none was observed with the γ-branched residue, Leu. It would appear that very bulky hydrophobic residues (Phe) may be excluded from the P_2 position. Poor cleavage of Pro-Arg is interesting, given the possiblity that Kex2 protease may cleave such a site in a natural substrate, pro-killer toxin. Published data on cleavage of Pro-Arg-containing substrates by partially purified Kex2 protease *in vitro* is conflicting (Achstetter and Wolf, 1985; Mizuno, et al. 1989). We have detected inhibition of Kex2 protease by D-Phe-Pro-Arg-chloromethylketone (C.Brenner and R.S. Fuller, unpublished), but a 100-fold higher concentration was required for inhibition than in the case of Ala-Lys-Arg-chloromethylketone (Fuller et al., 1989a). If the Pro-Arg site in pro-killer toxin is cleaved by Kex2 protease, it seems likely that the "presentation" of the site due to secondary or tertiary structure of the substrate may play an important role.

Perspectives: Detailed studies of the yeast Kex2 protease should shed light on the mechanism and specificity of precursor processing in the eukaryotic secretory pathway. The ability to purify large amounts of a soluble form of Kex2 protease will permit thorough studies of the primary sequence specificity of the enzyme and the interaction of the enzyme with natural substrate molecules, and may provide a novel reagent for protein chemistry as well. An *in vivo* assay for cleavage specificity provides a powerful complementary tool for probing substrate recognition by Kex2 protease.

Acknowledgements: The authors wish to acknowledge support for this work from a Lucille P. Markey Scholar Award and an N.I.H. grant (GM 39697) to R.S.F., and an EMBO postdoctoral fellowship to P.G.

REFERENCES

Achstetter, T. and Wolf, D.H. (1985) EMBO J. 4, 173-177.
Andrews, P.C., Brayton, K.A. and Dixon, J.E. (1989) In: Regulatory Peptides, Experientia Supplementum. Vol. 56 (J.M. Polak, Ed), Birkhäuser Verlag/ Basel, pp. 192-209.
Böhni, P. C., Deshaies, R. J., and Schekman, R. W. (1988) J. Cell Biol. 106, 1035-1042.
Fuller, R.S., Brake,A.J. and Thorner, J. (1986) In: Microbiology--1986 (L. Lieve, Ed), American Society for Microbiology/ Washington, D.C., pp. 273-278.
Fuller, R.S., Sterne, R.E., and Thorner, J. (1988) Ann. Rev. Physiol. 50, 345-362.
Fuller, R.S., Brake, A., and Thorner, J. (1989a) Proc. Natl. Acad. Sci. USA 86, 1434-1438.
Fuller, R.S., Brake, A.J. and Thorner, J. (1989b) Science 246, 482-486.
Furie, B. and Furie, B.C. (1988) Cell 53, 505-518.
Hartwell, L.H. (1980) J. Cell. Biol. 85, 811-822.
Ikemura, H. and Inouye, M. (1988) J. Biol. Chem. 263, 12,959-12,963.
Jones, E.W. (1984) Ann. Rev. Genet. 18, 233-270.
Julius, D., Brake, A., Blair, L., Kunisawa, R. and Thorner, J. (1984) Cell 37, 1075-1089.
Kaiser, C.A. and Schekman, R. (1990) Cell 61, 723-733.
Klenk, H.D. and Rott, R. (1988) Adv. Virus Research 34, 247-281.
Kukuruzinska, M. A., Bergh,M. L. E., and Jackson, B. J. (1987) Ann. Rev. Biochem. 56, 915-944.
Mizuno, K., Nakamura, T., Ohshima, T., Tanaka, S., and Matsuo, H. (1988) Biochem. Biophys. Res. Commun. 156, 246-254.
Mizuno, K., Nakamura,T., Oshima, T., Tanaka, S. and Matsuo, H. (1989) Biochem. Biophys. Res. Comm. 59, 305-311.
Moehle, C.M., Tizard, R., Lemmon, S.K., Smart, J., and Jones, E.W. (1987) Mol. Cell. Biol. 7, 4390-4399.
Perez, L.G., and Hunter, E. (1987) J. Virology 61, 1609-1614.
Power, S.D., Adams, R.M., and Wells, J.A. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 3096-3100.
Seidah, N.G., Gaspar, L., Mion, P., Marcinkiewicz, M. Mbikay, M. and Chrétien, M. (1990) DNA and Cell Biol. 9, 415-424.
Smeekens, S.P. and Steiner, D.F. (1990) J. Biol. Chem. 265, 2997-3000.
Sossin, W.S., Fisher, J.M., and Scheller, R.H. (1989) Neuron 2, 1407-1417.
Thim, L., Hansen,M.T., Norris, K., Hoegh, I., Boll, E., Forstrom, J., Ammerer, G. and Fiil, N.P. (1986) Proc. Natl. Acad. Sci. USA 83, 6766-6770.
Thomas, G., Thorne, B.A., Thomas, L., Allen, R.G., Hruby, D.E., Fuller, R. and Thorner, J. (1988) Science 241, 226-230.
van den Ouweland, A.M.W., van Duijnhoven, H.L.P., Keizer, G.D., Dorssers, L.C.J., and Van de Ven, W.J.M. (1990) Nuc. Acids Res. 18, 664.
Wells, J.A., Vasser, M., and Powers, D.B. (1985) Gene 34, 315-323.
Yoshimasa, Y., Seino, S., Whittaker, J., Kakehi, T., Kosaki, A., Kuzuya, H., Imura, H., Bell, G.I. and Steiner, D.F. (1988) Science 240, 784-787.