Expression, processing and secretion of a proteolytically-sensitive insect diuretic hormone by *Saccharomyces cerevisiae* requires the use of a yeast strain lacking genes encoding the Yap3 and Mkc7 endoproteases found in the secretory pathway

Kathrin S. COPLEY*, Sheri M. ALM*, David A. SCOOLEY* and William E. COURCHESNE†

*Department of Biochemistry, University of Nevada, Reno, NV 89557, U.S.A., and †Department of Microbiology, University of Nevada, Reno, NV 89557, U.S.A.

A system is described for the heterologous expression of peptides in *Saccharomyces cerevisiae*. A synthetic gene encoding a precursor of the 41 amino acid *Manduca sexta* diuretic hormone (Mas-DH) was expressed at 0.8 mg/l purified peptide. A precursor of a mutant peptide of Mas-DH, Mas-DH[K22Q] was also expressed. The peptides were purified, then treated with peptidylglycine α-amidating enzyme to generate the α-amidated, mature, form of Mas-DH or Mas-DH[K22Q], which were biologically active. Successful expression of full-length Mas-DH + Gly depended upon the use of a protease-deficient yeast strain. In wild-type strains, Mas-DH + Gly was recovered only as proteolytic fragments, even in the presence of various protease inhibitors. Expression of Mas-DH + Gly in strains deficient in either the Mkc7 or the Yap3 protease reduced proteolysis, while no proteolysis of Mas-DH + Gly was detectable in a strain lacking both proteases. This protease-deficient strain may prove of general utility for expression of peptides. Analysis of recovered proteolytic fragments revealed a complex pattern of cleavage sites. Both the Yap3 and Mkc7 proteases preferred to cleave at a single Glu-Lys-Glu-Arg site. Analysis of secondary cleavage sites showed that Yap3 preferred to cleave after either Lys or Arg and Mkc7 after Lys. This paper is the first report on the in vivo activity and specificity of Yap3 and Mkc7 expressed at physiological levels.

INTRODUCTION

There are four major types of expression systems in use today, *E. coli*, yeast, mammalian cells, and baculovirus-infected insect cells. Important considerations for the production of biologically active peptides include the desire for high levels of expression of the recombinant peptide, the requirement for proper post-translational modification, secretion of peptide for easier purification, and proper processing from a precursor form at the N-terminus. We sought to develop an expression system for the diuretic hormone of *Manduca sexta* (Mas-DH), a 41 amino acid peptide with an amidated carboxyl terminus. The baculovirus system has been tried using a synthetic gene encoding Mas-DH [1]; expression of intact peptide was never shown. We chose an *S. cerevisiae* expression system, because the peptide should be produced with its normal N-terminal residue, and be secreted from the cell at high levels [2], making purification relatively easy. Since no peptidylglycine α-amidating enzyme (PAM) activity has been found in yeast, an in vitro α-amidation reaction using commercially available PAM was chosen to complete the synthesis of biologically active Mas-DH. The C-terminal Gly is required for α-amidation of the peptide by PAM, a bifunctional enzyme found in the brain of *M. sexta* [3] and many vertebrates [4].

Some peptides less than 100 amino acids in length have been successfully expressed and secreted in yeast-based expression systems; however, the majority of these peptides are heavily constrained with disulphide bonds. Peptides with no disulphide bonds have been difficult to express without significant proteolysis. We describe a system for the successful expression of a nonproteolysed precursor (Mas-DH + Gly) of the diuretic hormone of *Manduca sexta*. Success required elimination of both the Yap3 and Mkc7 endoproteases from the secretory pathway. The sites cleaved by each of these proteases in vivo was identified.

EXPERIMENTAL

Preparation of Mas-DH + Gly expression vector

The Mas-DH + Gly oligonucleotide sequence was synthesized at the UNR DNA synthesis laboratory using the preferred codons for *S. cerevisiae*. The Mas-DH + Gly oligonucleotide (sense strand) was 168 bases long and encoded an amino acid linker region containing a Kex2 cleavage site followed immediately by the coding region for Mas-DH + Gly. Restriction sites were engineered into the sequence for ligation into the vector and gene manipulation. The synthetic oligonucleotide was purified by polyacrylamide gel electrophoresis (10 % gel, 7.5 M urea). Primers complementary to the 5’ and 3’ ends of the synthetic Mas-DH + Gly oligonucleotide were synthesized [5’-AGCGAATTCTTCAAACCTGATG-3’ (5′ primer) and 5’-GGCATCGATCATTAACCAAT-3’ (3′ primer)] and used with the 16 nucleotide sense strand as the template to amplify the Mas-DH + Gly

Abbreviations used: DH, diuretic hormone; Mas-DH, *M. sexta* diuretic hormone; CRF, corticotropin-releasing factor; PAM, peptidylglycine α-amidating enzyme; TFA, trifluoroacetic acid; BSA, bovine serum albumin; RPLC, reversed-phase liquid chromatography; ESI-MS, electrospray ionization mass spectrometry; MS-saline, *M. sexta* saline; IBMX, isobutylmethylxanthine; DTT, dithiothreitol; GAPDH, glyceraldehyde phosphate dehydrogenase.

1 Present address: Amylin Pharmaceuticals, 9737 Towne Centre Dr., San Diego, CA 92121.

2 To whom correspondence should be addressed.
gene fragment using standard PCR techniques (94 °C, 1 min; 63 °C, 2 min; 72 °C, 2 min; 30 cycles and Taq DNA Polymerase (Perkin-Elmer)]. The PCR product was ligated into pT7 Blue (Novagen) and transformed into bacterial strain JM110 for amplification and sequencing. The DNA sequence was determined using standard methods [5]. After confirming the sequence, the gene was excised from pT7 Blue using EcoR1 and Bsp106 I restriction enzymes (Stratagene) and purified by agarose gel electrophoresis. The gene fragment was then ligated into the YE-pIK expression plasmid (a 2p plasmid-based E. coli-yeast shuttle vector provided by Dr. I. V. Karpichev), which contains the TRP1 gene for selection in yeast. The YE-pIK plasmid also contains a glyceraldehyde phosphate dehydrogenase (GAPDH) promoter for high-level expression of any gene under its control. Secretion of the Mas-DH + Gly peptide from the cell was directed by the α-factor pre-pro sequence. The synthetic Mas-DH + Gly fragment was ligated into the vector immediately after and inframe with the α-factor pre-pro sequence, completing the synthetic Mas-DH gene in the expression vector pKSC2.

Expression of Mas-DH + Gly

The expression plasmid, pKSC2, was transformed by a Li+ acetate-based method into several different S. cerevisiae strains to optimize expression of the peptide (Table 1 below). Transformed cells were grown on synthetic media [1.7 g/l yeast nitrogen base without amino acids or (NH4)2SO4, and supplemented with 4 g/l (NH4)2SO4 and 20 g/l glucose]. Amino acids and nucleotides were added to supplement auxotrophies, except Trp which was omitted to allow selection for cells retaining the plasmid. Cells were grown to saturation (~24 h) at 30 °C in either polypropylene 250 ml Nalgene centrifuge bottles or 21 polypropylene Erlenmeyer flasks coated with Sigmacote to reduce adsorption of peptide to the growth vessel walls.

Purification of Mas-DH + Gly

Solid-phase extraction

The cells were removed by centrifugation (4100 g, 10 min) from 300 ml or 2 l of culture. Soluble proteins from the supernatant were adsorbed on Vyday C1 bulk protein support (20–30 μm particles) packed in a 75 ml polypropylene syringe barrel and equilibrated with 0.1% trifluoroacetic acid (TFA). The proteins were then eluted with 15 and 60% CH3CN–0.1% TFA. Fifty μl of a 1 mg/ml stock BSA (Sigma) solution was added to the 60% CH3CN–0.1% TFA fraction. This fraction was further purified by one of the following two RPLC-based methods.

A Spectra Physics SP8700 pump modified so that large sample volumes could be loaded onto the column [6], a Rheodyne loop injector, and a Spectra Chrom 100 variable-wavelength detector set at 220 nm were used to purify large quantities of Mas-DH + Gly and the α-amidated Mas-DH. The sample from solid-phase extraction of 2 l of culture medium was in two aliquots of 60% CH3CN–0.1% TFA; these were diluted to 10% CH3CN–0.1% TFA with 200 ml 0.1% TFA each and loaded into a 10 μm, 10 × 250 mm Vyday C1 semipreparative column equilibrated with 10% CH3CN–0.1% TFA. Proteins were eluted at 5 ml/min using a linear gradient of 10–60% CH3CN–0.1% TFA in 50 min. Peaks were collected manually and 50 μl of 1 mg/ml BSA was added to each fraction. Fractions were analysed with ESI–MS. Fractions corresponding to Mas-DH + Gly from the two RPLC runs were combined and dried to less than 1 ml final volume. This sample was then α-amidated to create Mas-DH.

RPLC Purification Method 2

The purified peptide from RPLC Method 2 was α-amidated using the enzyme peptide tyldiglycine α-amidating mono-oxygenase (PAM, Unigene Laboratories). The reaction conditions were based on the recommendations of Unigene Laboratories with modifications provided by Dr. B. Eipper (personal communication); these consisted of a buffer containing 0.03 M Mes/NaOH pH 6.5, 5 μM CuSO4, 0.01% Surfact-Amps X-100 (Pierce), 0.2 mM peptide, 0.1 mg/ml catalase (bovine liver; Sigma), 1.5 mM ascorbate, and 8000 units/ml PAM added in order, then incubated at 37 °C for 1 h. The α-amidated peptide was purified using RPLC Method 2. The 2 ml sample was diluted to 4 ml with 0.1% TFA and loaded into the column by loop injection. The peptide was eluted using a linear gradient of 10–60% CH3CN–0.1% TFA in 25 min. The identity and purity of the peak was confirmed by ESI–MS.

Site-directed Mutagenesis

An oligonucleotide was synthesized that changed Lys98 to Gln (5’TCAAGCAGATCGACTTTCTCGTGTCCAAAAGATA3′, complementary to the coding sequence). This primer plus the 5′ primer described above (see preparation of Mas-DH + Gly expression vector) were used to amplify the mutant 5′ portion of the gene using pKSC2 plasmid DNA (digested with HindIII) as the template (PCR conditions as stated above). The product was purified by agarose gel electrophoresis (PCR fragment 1). To generate the full-length gene, PCR fragment 1 was used as a
primer along with the Mas-DH 3′ primer and pKSC2 (digested with HindIII) as the template for a second PCR reaction using the same conditions as reported above. The product from this reaction (PCR 2 product) was the full-length Mas-DH + Gly gene containing the Lys to Gln mutation and was purified using agarose gel electrophoresis. Because of the low yield of the PCR 2 product, it was amplified by another round to PCR using the 5′ and 3′ primers. The reamplified product containing the site-directed mutation was purified by agarose gel electrophoresis. The mutant gene was digested with EcoR1 and Bsp106 I and ligated into the YEp-1K vector creating the plasmid pKSC3, which was transformed into E. coli (XL1-Blue, Stratagene).

Transformants carrying the mutant gene were preferred over those that may have contained the wild-type gene that was used as the template because the template DNA was Dam-methylated and not cut with the Bsp106 I enzyme. To verify the construction of the mutant Mas-DH[K22Q]+Gly, plasmids from three bacterial transformants were transformed into the BF25 yeast strain. These isolates were screened for expression of Mas-DH[K22Q]+Gly by RPLC purification and ESI-MS analysis (solid-phase extraction and RPLC Method 1). The identity of Mas-DH[K22Q]+Gly was confirmed by amino acid analysis and Edman sequencing.

Bioassay based on cAMP production

Five newly emerged adult male M. sexta moths less than 8 h old were used. Malpighian tubules were dissected with care taken to cut tubules into 1 cm lengths and to use only the white portion of the tubules proximal to the midgut for assays. The tubules were placed in a 96-well microtiter plate containing 100 µl MS-saline [prepared as described [8], containing IBMX (0.5 mM) and BSA (1 mg/ml)]. The tubules were preincubated for 1 h at 30 °C, then transferred to a polypolymer 96-well microtiter plate (Costar no. 3790) containing 100 µl serial dilutions of the peptide. EC50 values for the recombinant peptides were calculated from the production of cAMP by Malpighian tubules. To confirm the synthesis and secretion of intact diuretic hormone by yeast, Yap3 [13] and Mkc7 [14]. Two proteases are known to cleave proteins at basic residues in the yeast secretory pathway, Yap3 [13] and Mkc7 [14]. Accordingly we examined Mas-DH[K22Q]+Gly expression in a series of isogenic strains that carried disruptions of the MKC7 and YAP3 genes (Table 1).

Expression of Mas-DH + Gly in S. cerevisiae

The pKSC2 plasmid was transformed into several different wild-type S. cerevisiae strains. Purification of the culture medium using solid-phase extraction followed by RPLC Method 1 showed that there was substantial degradation of the Mas-DH + Gly product. The primary products isolated were analysed by ESI-MS and sequenced. The results showed that an endoproteolytic cleavage occurred between residues Lys22 and Glu25. Two proteases are known to cleave proteins at basic residues in the yeast secretory pathway, Yap3 [13] and Mkc7 [14]. Accordingly we examined Mas-DH + Gly expression in a series of isogenic strains that carried disruptions of the MKC7 and YAP3 genes (Table 1).

The plasmid pKSC2 was transformed into the single mutant strains HKY20 (yap3Δ) and HKY21 (mkc7Δ), the double-mutant strain HKY24 (yap3Δ, mkc7Δ), along with the wild-type parent strain CRY2 (YAP3Δ, MKC7Δ). The plasmid was also published amino acid sequence of the mature hormone [12], and included a codon for a C-terminal Gly for post-translational modification by z-amidation. Adhering to the yeast codon bias, a unique 168 base oligonucleotide was synthesized encoding information for a peptide linker followed by the Mas-DH + Gly coding region (Figure 1B). The peptide linker encodes a Kex2 protease cleavage site and includes Lys–Arg residues preceded by four amino acids normally found prior to the α-factor Kex2 cleavage site. Kex2 cleavage at this site would produce Mas-DH + Gly with the mature N-terminus. Two primers complementary to the 20 bases at each end of the 168mer were used to amplify the entire 168 base oligonucleotide. The resulting double-stranded DNA fragment containing the peptide linker plus the Mas-DH + Gly coding region was then inserted into the yeast expression vector, YEp-1K, to create the plasmid pKSC2. The fragment was ligated into the vector so that Mas-DH + Gly expression was controlled by the constitutive GAPDH promoter for high-level expression of Mas-DH + Gly. After construction, the synthetic Mas-DH + Gly gene contained the GAPDH promoter, the α-factor pre-pro region (to direct secretion of Mas-DH + Gly from the cell), the peptide linker, the Mas-DH + Gly coding region, and finally the PGK transcription termination region (Figure 1A).
transformed into the pep4Δ strain BFY25, which is deficient in protease A. Transformants were grown overnight under normal conditions to high cell density. The products in the culture medium were purified by solid-phase extraction followed by RPLC Method 1. Mas-DH + Gly and peptide fragments were identified by RPLC retention time and ESI–MS (Table 2). The CYR2[pKSC2] wild-type strain produced the fragments Mas-DH[1–22], Mas-DH[23–42], and Mas-DH[25–42] (Figure 2A). Full-length Mas-DH + Gly was not detected in this strain. The same fragments were identified in the BFY25[pKSC2] (pep4Δ) strain even when the cells were grown in the presence of a mixture of protease inhibitors.

Prior to seeking to express Mas-DH in protease-deficient cells, we grew cells in the presence of 10 mM DTT, 1 M sorbitol and a variety of protease inhibitors to see if the proteolytic activity at Lys could be inhibited. The protease inhibitors used included three serine protease inhibitors [100 µM 4-(2-aminoethyl)benzenesulphonyl fluoride (AEBSF), 1 ± 5 µM aprotinin, and 100 µM Nα-p-tosyl-L-lysine chloromethyl ketone (TLCK)], 100 µM of the aspartic protease inhibitor pepstatin, 42 µM leupeptin (an inhibitor of Ser and Cys proteinases), and 400 µM of the metalloprotease inhibitor 1,10-phenanthroline. Cells were grown using numerous combinations of the protease inhibitors listed above (data not shown). Full-length Mas-DH + Gly was not found in the supernatant of any of these growth conditions; of these protease inhibitors, only 1,10-phenanthroline had distinctly adverse effects on growth of the yeast.

In contrast, deletion of proteases from the secretory pathway was useful in diminishing proteolysis of the expressed peptide. HKY20[pKSC2] (yap3Δ) produced a small amount of full-length Mas-DH + Gly (~ 9 % of recovered Mas-DH + Gly fragments). The major peptide components isolated from the media were the fragments Mas-DH[1–21], Mas-DH[1–22], and Mas-DH + Gly[23–42] (Figure 2B). HKY21[pKSC2] (mkc7Δ) produced a relatively larger amount of full-length Mas-DH + Gly (~ 27 % of recovered Mas-DH + Gly fragments) and yielded fragments Mas-DH[1–22], Mas-DH + Gly[23–42], and Mas-DH + Gly[16–42] (Figure 2C). In all cases the fragment resulting from cleavage after Lys22 was the most abundant one found, suggesting that the Lys22 site is the primary cleavage site. In the mkc7Δ deletion strain (HKY21[pKSC2]) there was evidence for a secondary cleavage site occurring after Arg15 (Table 2). The double-mutant strain HKY24[pKSC2] (yap3Δ, mkc7Δ) produced ~ 90 % full-length Mas-DH + Gly and ~ 10 % ‘linker’-Mas-

### Table 1

<table>
<thead>
<tr>
<th>Strain*</th>
<th>Genotype</th>
<th>Plasmid†</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRY2</td>
<td>MATα, can1-100, ade2-101, his3–11,15, leu2-3,112, trp1-1, ura1-1</td>
<td>pKSC2</td>
</tr>
<tr>
<td>HKY20</td>
<td>CRY2 with yap3Δ::LEU2</td>
<td>pKSC2</td>
</tr>
<tr>
<td>HKY21</td>
<td>CRY2 with mkc7Δ::HIS3</td>
<td>pKSC2</td>
</tr>
<tr>
<td>HKY24</td>
<td>CRY2 with yap3Δ::LEU2, mkc7Δ::HIS3</td>
<td>pKSC2</td>
</tr>
<tr>
<td>BFY106-4D</td>
<td>CRY2 with kon2-Δ2::HIS3</td>
<td>pKSC2</td>
</tr>
<tr>
<td>BFY25</td>
<td>MATα, his3–Δ200, leu2-3,112, trp1-Δ901, ura3-52, ade5, pep4::LEU2</td>
<td>pKSC2</td>
</tr>
</tbody>
</table>

* All strains were generously provided by R. Fuller.
† Indicates which plasmids were transformed into given strain.

### Table 2

<table>
<thead>
<tr>
<th>Strains</th>
<th>Mutant Mas-DH[K22Q] + Gly fragments</th>
<th>Mas-DH + Gly fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MKC7</td>
<td>RLPQLQLRNLIPPVPLAVRQNLRLPLQKLY</td>
<td>RMPSLSIDLPMSVLRQKLSLEKRKVHALRAAANRNFLNDIG; Mas-DH[K22Q] + Gly = RMPSLSIDLPMSVLRQKLSLEQERKVHALRAAANRNFLNDIG.</td>
</tr>
<tr>
<td>MKC7 yap3Δ</td>
<td>RLPQLQLRNLIPPVPLAVRQNLRLPLQKLY</td>
<td>RMPSLSIDLPMSVPLAVRQNLRLPLQKLY</td>
</tr>
</tbody>
</table>

* Fragments joined with a brace on the right eluted together.
† Fragments joined with a brace on the right eluted together.
DH+Gly (containing a nonapeptide-linker, not removed by Kex2, but without the α-factor pre-pro sequence) (Figure 2D). Thus both Yap3 and Mkc7 are capable of contributing to the cleavage of Mas-DH+Gly at the Lys85 residue, as well as at other sites.

Expression of Mas-DH[K22Q]+Gly

In an attempt to eliminate cleavage at Lys82 by the Yap3 and Mkc7 proteases, this Lys was mutated to Gln. Substituting Gln for Lys at residue 22 should not affect the activity of the peptide because the change is conservative and in a location of the sequence that varies between the different diuretic hormones (alignment not shown). Plasmid pKSC3 carrying the K22Q sequence that varies between the different diuretic hormones for Lys at residue 22 should not affect the activity of the peptide.

Production of biologically active Mas-DH and Mas-DH[K22Q]

Mas-DH+Gly and Mas-DH[K22Q]+Gly expressed in the HKY24 strain were purified by solid-phase extraction and RPLC Method 2, then α-aminated using the PAM enzyme. The α-aminated Mas-DH and Mas-DH[K22Q] were then repurified.
The α-amidated products were identified using ESI-MS. Before α-amination an aliquot of the Mas-DH + Gly peptide was taken for amino acid analysis. The results indicated a yield of peptide of 0.8 mg/l (167 nM). After α-amination and the final purification, the yield of Mas-DH was 0.4 mg/l (84 nM), while the final yield for the Mas-DH[K22Q] peptide was 0.13 mg/l (27 nM).

The purified recombinant peptides were assayed for biological activity by measuring the production of cAMP by Malpighian tubules of adult male M. sexta. Malpighian tubules, cut to 1 cm lengths, were dissected from insects and incubated in M. sexta saline (MS-saline) for one h to equilibrate the tubules. The yeast-expressed Mas-DH or Mas-DH[K22Q] was then added. After 1 h, aliquots were removed, cAMP production quantified using the Gilman assay [9], and EC50 values calculated. The recombinant peptides Mas-DH and Mas-DH[K22Q] had EC50 values of ~2 nM, equivalent to that of synthetic Mas-DH [15], demonstrating that the Mas-DH synthesized in yeast is biologically active in an in vitro assay.

**DISCUSSION**

We created an *S. cerevisiae* expression system for production of *M. sexta* diuretic hormone (Mas-DH). Authentic Mas-DH purified from insect tissue is a 41 amino acid peptide with a C-terminal amide group; synthetic Mas-DH with a C-terminal carboxylate has 1000-fold reduced biological activity [12,15]. α-Amination of the C-terminus is carried out by the bifunctional enzyme PAM found in the secretory pathway of insects and other higher eukaryotes, but not in *S. cerevisiae*. Thus, we produced Mas-DH in yeast as the 42 amino acid Mas-DH + Gly form. This unmodified peptide was secreted into the growth medium from which it was easily purified in two steps. Our procedure included treatment of containers (flanks, etc.) with Sigmacote and inclusion of BSA in the partially purified fractions to reduce adsorption of Mas-DH + Gly to the containers; this increased yields by 33%. The purified peptide was examined by ESI-MS and amino acid sequencing to verify its identity. After two purifications our yield of Mas-DH + Gly was 0.8 mg/l of yeast culture (167 nM). This value can be compared with yields of other heterologous peptides expressed in yeast. The reported yields for bovine pancreatic trypsin inhibitor and glucagon are 3 mg/l (453 nM) and 177 nM respectively [16,18]. However, these values are for unpurified recombinant protein; thus, our production of Mas-DH + Gly is approximately the same as that for production of other heterologous peptides in yeast. Our Mas-DH + Gly expression system is practical and desirable because the purification and α-amination steps are quick and easy. The amount of peptide that can be generated, coupled with the rapidity and ease of making and expressing site-directed mutations of our synthetic Mas-DH gene, will allow us to thoroughly examine the contribution of specific amino acids to ligand–receptor interactions and to study the three-dimensional conformation of Mas-DH.

Because of our difficulties with proteolysis, we examined the literature for peptides less than 10 kDa in size which have been expressed and secreted from *S. cerevisiae*. A list of such peptides is shown in Table 3; not listed are peptides expressed as part of a fusion protein requiring further processing. Of the 17 peptides listed in Table 3, 12 are constrained by disulphide bonds; of the remaining five, three were partially or completely cleaved by endoproteases, while the fourth, the antifreeze peptide AFP6, was only expressed as a multicopy form. The AFP6 peptide is also unusual in that it is 70% alanine and contains only two basic residues. Glucagon was the only non-disulphide-containing peptide, to our knowledge, that was expressed at significant levels (177 nM) [18]. Nonetheless, glucagon appears to have suffered some uncharacterized proteolysis; during their identification of full-length glucagon, Moody et al. [18] did not characterize two immunoreactive zones with a different RPLC retention time to that of full-length glucagon, but they did suggest that these fractions could be products of a cleavage at R15-R18.

Although our initial aim was simply to express Mas-DH in yeast, we have gleaned information on the properties of proteases of the secretory system. We recovered and characterized proteolytic products of Mas-DH + Gly that resulted from two recently identified proteases (Yap3 and Mkc7) found in the yeast secretory pathway. Analysis of these proteolytic products provides insight into the *in vivo* proteolytic activities of the Yap3 and Mkc7 enzymes. In previous studies, Yap3 and Mkc7 were overexpressed and their cleavage-site preferences studied by changing residues around a Lys-Arg cleavage site. Studies done using anglerfish pro-somatostatin II showed that Yap3 cleaved after a monobasic site, Arg22 [19]. Another study [13] done using partially purified Yap3 and a synthetic substrate showed that Yap3 cleaved after dibasic sites, and Yap3 activity was enhanced by placing an Arg in the +2 position (e.g. the second residue after the scissile bond). Our studies confirm that Yap3 efficiently cleaves after a Lys that has an Arg at the +2 position. Ledgerwood et al. [13] observed the highest cleavage rate using

---

**Table 3** Heterologous expression of secreted peptides < 10 kDa in *S. cerevisiae*

<table>
<thead>
<tr>
<th>Peptide name</th>
<th>No. of amino acids</th>
<th>No. of disulphide bonds</th>
<th>No. of Lys and Arg</th>
<th>Cleavage sites</th>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somatostatin [26]</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td></td>
<td>Full-length†</td>
</tr>
<tr>
<td>Glucagon [18]</td>
<td>29</td>
<td>0</td>
<td>3</td>
<td>R18</td>
<td>Evidence of cleavage; products uncharacterized</td>
</tr>
<tr>
<td>β-Endorphin [27]</td>
<td>31</td>
<td>0</td>
<td>5</td>
<td>K9, K19</td>
<td>No full-length product found</td>
</tr>
<tr>
<td>Calcitonin [28]</td>
<td>32</td>
<td>1</td>
<td>1</td>
<td></td>
<td>Detection by RIA</td>
</tr>
<tr>
<td>Scorpion insect-selective toxin</td>
<td>55</td>
<td>4</td>
<td>5</td>
<td></td>
<td>Not active</td>
</tr>
<tr>
<td>ISA [29]</td>
<td>37</td>
<td>0</td>
<td>2</td>
<td></td>
<td>Required multiplicity gene for expression</td>
</tr>
<tr>
<td>Antifreeze peptide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AFP6 [29a]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mas-DH + Gly</td>
<td>42</td>
<td>0</td>
<td>8</td>
<td></td>
<td>Full-length‡, ~ 93% full-length, ~ 7% cleavage</td>
</tr>
<tr>
<td>Echistatin [30]</td>
<td>49</td>
<td>4</td>
<td>7</td>
<td>R22</td>
<td></td>
</tr>
<tr>
<td>hEGF (ungastrone)</td>
<td>53</td>
<td>3</td>
<td>5</td>
<td></td>
<td>Full-length‡</td>
</tr>
<tr>
<td>h pancreatic secretory trypsin inhibitor</td>
<td>55</td>
<td>3</td>
<td>7</td>
<td></td>
<td>Full-length‡</td>
</tr>
<tr>
<td>Insulin</td>
<td>55</td>
<td>3</td>
<td>2</td>
<td></td>
<td>Full-length‡</td>
</tr>
<tr>
<td>Aprotinin ((\approx) Bovine pancreatic trypsin inhibitor)</td>
<td>58</td>
<td>3</td>
<td>10</td>
<td>R42</td>
<td>Some cleavage, K41S mutation increased yield</td>
</tr>
<tr>
<td>Bovine pancreatic trypsin inhibitor</td>
<td>58</td>
<td>3</td>
<td>10</td>
<td></td>
<td>Full-length‡</td>
</tr>
<tr>
<td>Hirudin [35]</td>
<td>65</td>
<td>3</td>
<td>3</td>
<td></td>
<td>C-terminal degradation</td>
</tr>
<tr>
<td>A. alastrim insect-selective toxin 1 [36]</td>
<td>69</td>
<td>4</td>
<td>6</td>
<td></td>
<td>Low yield, perhaps due to cleavage</td>
</tr>
<tr>
<td>Insulin-like growth factor 1 [37]</td>
<td>70</td>
<td>3</td>
<td>6</td>
<td></td>
<td>Full-length‡</td>
</tr>
<tr>
<td>h parathryoid hormone* [38]</td>
<td>84</td>
<td>0</td>
<td>14</td>
<td>K26, F35, R44</td>
<td>Yield very low due to cleavage</td>
</tr>
</tbody>
</table>

* h, human. † No endoproteolytic products reported. ‡ No proteolytic products found in Yap3, Mkc7 protease defective strain.
a synthetic peptide that contained the sequence RR↓DR (position numbers −2 −1 +1 +2 +3) [13]. The primary cleavage site in Mas-DH+Gly was EK↓ER. The two sequences are similar; both contain a basic residue in the −1 position followed by an acidic, then basic, residue in the +1 and +2 positions. One secondary cleavage site seen in the Mas-DH[K22Q]+Gly mutant expressed in HKY21 (contains Yap3 but no Mkc7) also has a basic residue (K) at the +2 position, but contains a Gln at the +1 position. The other secondary cleavage site occurs between Arg and Lys in the sequence ER↓KV, in which the +2 residue is neutral. Interestingly, the −2 residue is acidic as it is in the primary cleavage site. Ledgerwood et al. [13] did not examine the effects of an acidic residue in the −2 position.

The cleavage-site preference of Mkc7 was unknown prior to this work. Komano and Fuller [14] used partially purified Mkc7 and a synthetic peptide containing Lys-Arg as a cleavage site to assay for Mkc7 activity. We found that the HKY20 strain (contains Mkc7 but not Yap3) producing Mas-DH+Gly exhibits only one cleavage site, EK↓ER, while in the strain producing mutant Mas-DH[K22Q]+Gly, where the EK↓ER site is mutated to EQER, a second cleavage site, RK↓VH, appears. Our data suggest that Mkc7 prefers to cleave after Lys residues and that it has preferences for the −2, +1, and +2 residues very similar to those of Yap3.

Thus Mkc7 and Yap3 may have similar but not identical preferences for the residues used in substrate recognition. The preferred residues appear to be either Asp or Glu at positions −2 and +1 and either Lys or Arg, or to a lesser extent His, at position +2. Mkc7 may prefer Lys at position −1 and Yap3 appears to prefer Lys or Arg. Limited data on the secondary cleavage sites suggest that the basic residue at −1 and one of the other three determinants discussed are required for even minimal cleavage at that site. Each secondary cleavage site contains a basic residue at position −1 (the prime determinant for whether the protease will cleave the sequence) and a preferred but not absolutely required residue at either −2 or +2, demonstrating some flexibility at the cleavage site.

Many researchers have suggested that the conformation of a protein substrate is as important as the identity of particular residues for substrate recognition. Molecular modeling and circular dichroism studies done here (Copley et al., unpublished data) show that Mas-DH has a helix–loop–helix structure. Loop regions are often the targets for proteolysis. The primary cleavage site and all but one of the secondary cleavage sites of Mas-DH+Gly and Mas-DH[K22Q]+Gly are in fact located in the predicted loop region. An unusual feature of the CRF-related insect DH family is the gap region that must be inserted into the sequence alignments of the seven published DH, which vary in length from 30 to 46 amino acids [20–25], in order to get the best sequence alignment. The requirement of adding a gap region for best alignment is unusual in a peptide family, but is not unusual for protein families. These gaps in sequence alignment often correspond to loop regions in a protein structure; which are known to be more susceptible to endoproteolytic cleavage.

Expression of Mas-DH and its analogues provides an in vivo system that should be very useful for studying the proteolytic specificity of yeast secretory pathway processing enzymes, and complement in vitro studies on purified proteins. The advantages of this system include the ability to examine the activities of the Yap3 and Mkc7 enzymes (either independently or together) in their native habitats, expressed at normal levels, and acting on an authentic peptide hormone. The peptide Mas-DH+Gly is a good test substrate because it is a natural peptide that normally survives an insect cell secretory pathway, and the degradation products are readily isolated and easily characterized using RPLC and ESI–MS. Our Mas-DH+Gly expression vector is constructed so that site-directed mutagenesis of virtually every residue is easily and quickly accomplished. Thus, it is possible to examine the contribution of any residue to the substrate cleavage in vivo. Production of mature, unproteolysed Mas-DH+Gly depended on the availability of HKY24 (yap3Δ and mkc7Δ). This strain may prove of general utility for the expression of other peptides and proteins that are improperly proteolysed at Lys and Arg residues in wild-type yeast cells or other expression systems.

We thank Dr. I. V. Karpichev, Centre of Bioengineering, Russian Academy of Sciences, Moscow, 117984, Russia for providing the YEp-IK vector and Dr. Robert Fuller (U. of Michigan School of Medicine) for the isogeneric protease-deficient yeast strains. We also thank Dr. B. Elipper (Johns Hopkins U. School of Medicine) for her advice on using PAM to amide the Mas-DH+Gly peptide. The ESI–MS characterizations were done with the help of Dr. Houle Wang (currently at U. of Washington) and Dr. David Quilici (U. of Nevada, Reno). The Edman sequencing of the peptides was done by Dr. Kathleen M. Schegg (U. of Nevada, Reno). We acknowledge the Nevada Agriculture Experiment Station and NIH GM 49172 for partial financial support (K.S.C.).

REFERENCES

7 Reference deleted
18 Reference deleted

Received 26 August 1997/24 October 1997; accepted 6 November 1997