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^{*1}, Sheri M. ALM^{*}, David A. SCHOOLEY^{*2} 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'-AGCGAA-TTCTCAACCAATGT-3' (5' primer) and 5'-GGCATCGAT-CATTAACCAAT-3' (3' primer)] and used with the 168 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, 9373 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 *Eco*R1 and *Bsp*106 I restriction enzymes (Stratagene), and purified by agarose gel electrophoresis. The gene fragment was then ligated into the YEp-IK expression plasmid (a 2 μ plasmid-based *E. coli*–yeast shuttle vector provided by Dr. I. V. Karpichev), which contains the *TRP1* gene for selection in yeast. The YEp-IK 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 in-frame 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 2 l 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 Vydac C₄ bulk protein support (20–30 μ m 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 μ l 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.

RPLC Purification Method 1

A Perkin–Elmer model 410 Bio pump and a Perkin–Elmer model 235 dual wavelength detector set at 220 and 280 nm were used. The pump was modified so that large sample volumes could be pumped into the column [6]. The 60 % CH₃CN–0.1 % TFA fraction from solid-phase extraction (300 ml starting culture) was diluted to a final concentration of 10 % CH₃CN–0.1 % TFA and loaded on to a 4.6 $\times$ 150 mm, 8 μ m PLRP-S column (Polymer Labs) equilibrated with 2 % CH₃CN–0.1 % TFA. Proteins were eluted at 1 ml/min using a linear gradient of 2–60 % CH₃CN–0.1 % TFA in 60 min. Peaks were collected manually. The relative percentages of fragments given in the text are approximate and based on the assumption that the nanomolar amount of peptide is directly proportional to the peak area at 220 nm. Extinction

coefficients at 220 nm are sequence-dependent and profoundly affected by the presence of aromatic amino acid residues; thus, peak heights (or areas) of various peptide fragments cannot be accurately used for relative quantification.

Fractions were analysed by electrospray ionization mass spectrometry (ESI–MS) using a Finnigan MAT SSQ 710 mass spectrometer interfaced with an Analytica ESI ion source. Those fractions containing multiple fragments (see Table 2) were purified with an additional RPLC step using a Hewlett–Packard Model 1050 liquid chromatograph with a Zorbax 2.1 $\times$ 150 mm C₈ column equilibrated with 0.1 % TFA. Fragments were eluted at 0.2 ml/min using a linear gradient of 0–40 % CH₃CN–0.1 % TFA in 60 min and re-analysed by ESI–MS. The fragments listed in the upper left-hand panel of Table 2 (below), full-length Mas-DH + Gly, and full-length Mas-DH[K22Q] + Gly were sequenced with a Porton Instruments PI 2090 gas-phase sequencer with an integrated phenylthiohydantoin amino acid analyser.

RPLC Purification Method 2

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 μ m, 10 $\times$ 250 mm Vydac C₄ 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 μ l 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.

Peptide α -amidation

The purified peptide from RPLC Method 2 was α -amidated using the enzyme peptidylglycine 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 μ M 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 Lys²² to Gln (5'-TCAAAGCATGCACTTTCCTTTCTTGTTCCTCAAAGATA-3', 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 *Hind*III) 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 *Hind*III) 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 *Eco*R1 and *Bsp*106 I and ligated into the YEp-IK 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 *Bsp*106 I enzyme. To verify the construction of the mutant Mas-DH[K22Q] + Gly, plasmids from three bacterial transformants were transformed into the BFY25 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 microtitre plate containing 100 μ l 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 polypropylene 96-well microtitre plate (Costar no. 3790) containing 100 μ l serial dilutions of the peptide. EC_{50} values for the recombinant peptides were calculated from the production of cAMP by Malpighian tubules. To calculate the EC_{50} value of synthetic Mas-DH and the two recombinant peptides, production of cAMP was measured for concentration of peptide between 10^{-5} and 10^{-11} M. Dried peptide was weighed and dissolved at an initial concentration of 10^{-5} M in MS-saline. This stock concentration was used to generate serial dilutions. Seven replicates were assayed for each concentration. The plate was incubated 1 h at 30 °C. After incubation, 50 μ l aliquots from each well were transferred to a microcentrifuge tube for quantification of cAMP using the Gilman competitive binding assay [9,10]. The cAMP standard curve is linear in the range of 1–16 pmol. Samples incubated with high concentrations of peptide had to be diluted (10^{-5} , 10^{-6} M, 1:10 dilution; 10^{-7} , 10^{-8} M; 1:5 dilution) to remain within the linear region of the standard curve.

RESULTS

Construction of the Mas-DH + Gly expression vector

Digan et al. [11] isolated a cDNA encoding Mas-DH from *M. sexta* pharate adult heads. The cDNA clone encoded a 138 amino acid precursor of Mas-DH that included a putative signal sequence (amino acids 1–19) and a 40 amino acid pro-region followed by the Mas-DH coding region. There is a dibasic Kex2-like endoproteolytic site present on both ends of the coding region that would generate a precursor of natural Mas-DH, Mas-DH + Gly. For expression of Mas-DH in *S. cerevisiae*, we constructed a gene of organization similar to the natural gene. A DNA fragment encoding Mas-DH was made based upon the

Figure 1 Diagram of the pKSC2 expression vector

(A) View of the Mas-DH gene. The elements in the diagram are not drawn to scale. (B) Synthetic 168 base oligonucleotide coding for the peptide linker and Mas-DH + Gly coding region. The Gly is required for post-translational amidation.

published amino acid sequence of the mature hormone [12], and included a codon for a C-terminal Gly for post-translational modification by α -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-IK, 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).

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 Glu²³. 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 (*yap3* Δ) and HKY21 (*mkc7* Δ), the double-mutant strain HKY24 (*yap3* Δ , *mkc7* Δ), along with the wild-type parent strain CRY2 (*YAP3*, *MKC7*). The plasmid was also

Table 1 Mas-DH + Gly expression in isogenic strains carrying disruptions of several protease genes

Strain*	Genotype	Plasmid†
CRY2	<i>MATα</i> , <i>can</i> 1-100, <i>ade</i> 2-101, <i>his</i> 3-11, 15, <i>leu</i> 2-3, 112, <i>trp</i> 1-1, <i>ura</i> 3-1	pKSC2
HKY20	CRY2 with <i>yap3Δ::LEU2</i>	pKSC3
HKY21	CRY2 with <i>mkc7Δ::HIS3</i>	pKSC2
HKY24	CRY2 with <i>yap3Δ::LEU2</i> , <i>mkc7Δ::HIS3</i>	pKSC3
BFY106-4D	CRY2 with <i>kex2-Δ2::HIS3</i>	pKSC2
BFY25	<i>MATα</i> , <i>his</i> 3-Δ200, <i>leu</i> 2-3, 112, <i>trp</i> 1-Δ901, <i>ura</i> 3-52, <i>ade</i> 5, <i>pep</i> 4::LEU2	pKSC2
		pKSC3

* All strains were generously provided by R. Fuller.

† Indicates which plasmids were transformed into given strain.

transformed into the *pep4Δ* strain BFY25, which is deficient in proteinase 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 CRY2[pKSC2] wild-type strain produced the fragments Mas-DH[1-22], Mas-DH + Gly[23-42], and Mas-DH + Gly[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 Lys²² 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 *N*^α-*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 2 Proteolytic fragments of Mas-DH + Gly and Mas-DH[K22Q] + Gly recovered after expression in *S. cerevisiae*

Strains	Mas-DH + Gly fragments	Mutant Mas-DH[K22Q] + Gly* fragments
<i>MKC7 YAP3</i>	<p>RMPSLSIDLPMSVLRQKLSLEK</p> <p>ERKVKHALRAAAANRNFLNDIG }† KVHALRAAAANRNFLNDIG</p>	<p>RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG RMPSLSIDLPMSVLRQKLSLEQERK KVHALRAAAANRNFLNDIG QKLSLEQERKVKHALRAAAANRNFLNDIG</p>
<i>MKC7 yap3Δ</i>	<p>RMPSLSIDLPMSVLRQKLSLEKVKHALRAAAANRNFLNDIG RMPSLSIDLPMSVLRQKLSLE RMPSLSIDLPMSVLRQKLSLEK</p>	<p>RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG KVHALRAAAANRNFLNDIG</p>
<i>mkc7Δ YAP3</i>	<p>RMPSLSIDLPMSVLRQKLSLEK ERKVKHALRAAAANRNFLNDIG } KVHALRAAAANRNFLNDIG QKLSLEKVKHALRAAAANRNFLNDIG</p>	<p>RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG RMPSLSIDLPMSVLRQKLSLEQERK KVHALRAAAANRNFLNDIG QKLSLEQERKVKHALRAAAANRNFLNDIG</p>
<i>mkc7Δ yap3Δ</i>	<p>RMPSLSIDLPMSVLRQKLSLEKVKHALRAAAANRNFLNDIG linker-RMPSLSIDLPMSVLRQKLSLEKVKHALRAAAANRNFLNDIG linker-RMPSLSIDLPMSVLRQKLSLEKVKHALRAAAANRNFLNDIG</p>	<p>RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG linker-RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG linker-RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG</p>

* Mas-DH + Gly: RMPSLSIDLPMSVLRQKLSLEKVKHALRAAAANRNFLNDIG; Mas-DH[K22Q] + Gly = RMPSLSIDLPMSVLRQKLSLEQERKVKHALRAAAANRNFLNDIG.

† Fragments joined with a brace on the right eluted together.

Figure 2 RPLC chromatograms of Mas-DH + Gly fragments recovered from a series of isogenic strains transformed with pKSC2

RPLC conditions were similar for (A), (B), and (C) (RPLC purification Method 1). Proteins were eluted at 1 ml/min using a linear gradient of 2–60% CH₃CN–0.1% TFA in 60 min. Peaks were collected by hand and analysed by ESI–MS. The preparative separation shown in (D) was performed with RPLC purification Method 2. Proteins were eluted at 5 ml/min using a linear gradient of 10–60% CH₃CN–0.1% TFA in 50 min. Peaks were collected and analysed by ESI–MS. For all figures absorbance was measured at 220 nm. In (A–D) the Mas-DH + Gly fragments are labelled as follows: [1–22] corresponds to the fragment Mas-DH[1–22]; [1–21], Mas-DH[1–21]; [15–42], Mas-DH + Gly[15–42]; [23–42], Mas-DH + Gly[23–42]; [25–42], Mas-DH + Gly[25–42]. The peak marked by the asterisk (*) was present in cells not expressing Mas-DH + Gly. The M_r of unlabelled peaks do not correspond to the calculated M_r of fragments of Mas-DH + Gly.

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 Lys²² residue, as well as at other sites.

Expression of Mas-DH[K22Q] + Gly

In an attempt to eliminate cleavage at Lys²² 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 mutant gene was transformed into the same set of strains used to express Mas-DH + Gly (see Experimental procedures). Each strain was grown overnight to saturation and components of the culture medium were purified by solid-phase extraction and RPLC Method 1. Mas-DH[K22Q] + Gly and its fragments were identified by RPLC retention time and ESI–MS (Table 2). BFY25[pKSC3] (*pep4Δ*) produced primarily fragments of Mas-DH[K22Q] + Gly generated by three cleavage sites; after Arg¹⁶, Arg²⁴ and Lys²⁵, while only ~2% of the recovered Mas-

Figure 3 RPLC chromatograms of Mas-DH[K22Q] + Gly fragments recovered from a series of isogenic strains transformed with pKSC3

RPLC conditions were similar for (A–D) (RPLC purification Method 1). All peaks were collected and analysed by ESI–MS. For all figures absorbance was measured at 220 nm. In (A–D) the Mas-DH[K22Q] + Gly fragments are labelled as follows: [1–24] corresponds to the fragment Mas-DH[K22Q] [1–24]; [15–42], Mas-DH[K22Q] + Gly[15–42]; [25–42], Mas-DH[K22Q] + Gly[25–42]; [26–42], Mas-DH[K22Q] + Gly[26–42]. Full-length Mas-DH[K22Q] + Gly is labelled as DH. The peak marked by the asterisk (*) was present in cells not expressing Mas-DH[K22Q] + Gly. The M_r of unlabelled peaks do not correspond to the calculated M_r of fragments of Mas-DH[K22Q] + Gly.

DH[K22Q] + Gly was full-length peptide (Figure 3A). Thus, elimination of the Lys²² cleavage by substitution did not prevent proteolysis, rather it revealed secondary cleavage sites. Using HKY20 and HKY21, it was possible to identify the protease responsible for each of the cleavages. Strain HKY20[pKSC3] (*yap3Δ*) produced the full-length Mas-DH[K22Q] + Gly (~56% of total) and the fragment Mas-DH[K22Q] + Gly[25–42] (Figure 3B), suggesting that Mkc7 cleaved after Lys²⁵ in the context RK↓VH. The *mkc7Δ* strain HKY21[pKSC3] generated full-length Mas-DH[K22Q] + Gly (~25% of total) and the fragments Mas-DH[K22Q] + Gly[16–42] and Mas-DH[K22Q] + Gly[24–42] (Figure 3C), suggesting that Yap3 cleaved after Arg at the two sites ER↓KV and LR↓QK. This suggests that Mkc7 prefers to cleave after Lys, and Yap3 after either Lys or Arg. Finally, the double-mutant strain HKY24[pKSC3] (*yap3Δ*, *mkc7Δ*) produced full-length Mas-DH[K22Q] + Gly (~90%) and 'linker'-Mas-DH[K22Q] + Gly (~10%) (Figure 3D), demonstrating that proteolysis of Mas-DH[K22Q] + Gly was dependent on Yap3 and Mkc7.

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 -amidated using the PAM enzyme. The α -amidated Mas-DH and Mas-DH[K22Q] were then repurified.

The α -amidated products were identified using ESI-MS. Before α -amidation 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 -amidation 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 EC_{50} values calculated. The recombinant peptides Mas-DH and Mas-DH[K22Q] had EC_{50} 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 -Amidation 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 (flasks, 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 -amidation 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 R¹⁷-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 over-expressed 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, Arg⁷³ [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*

Peptide name	No. of amino acids	No. of disulphide bonds	No. of Lys and Arg	Cleavage sites	Production
Somatostatin [26]	14	1	2		Full-length†
Glucagon [18]	29	0	3	R18	Evidence of cleavage; products uncharacterized
β -Endorphin [27]	31	0	5	K9, K19	No full-length product found
Calcitonin [28]	32	1	1		Detection by RIA
Scorpion	35	4	5		Not active
insect-selective toxin ISA [29]					
Antifreeze peptide AFP6 [29a]	37	0	2		Required multicopy gene for expression
Mas-DH + Gly	42	0	8		Full-length‡
Echistatin [30]	49	4	7	R22	$\sim 93\%$ full-length, $\sim 7\%$ cleavage
hEGF (urogastrone) [31]*	53	3	5		Full-length†
h pancreatic secretory trypsin inhibitor* [32]	55	3	7		Full-length†
Insulin [33]	55	3	2		Full-length†
Aprotinin (= Bovine pancreatic trypsin inhibitor) [34]	58	3	10	R42	Some cleavage, K41S mutation increased yield
Bovine pancreatic trypsin inhibitor [16]	58	3	10		Full-length†
Hirudin [35]	65	3	3		C-terminal degradation
<i>A. australis</i> insect-selective toxin 1 [36]	69	4	6		Low yield, perhaps due to cleavage
Insulin-like growth factor 1 [37]	70	3	6		Full-length†
h parathyroid hormone* [38]	84	0	14	K26, F35, R44	Yield very low due to cleavage

* 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) [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 modelling 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 isogenic protease-deficient yeast strains. We also thank Dr. B. Elipper (Johns Hopkins U. School of Medicine) for her advice on using PAM to amidate 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 48172 for partial financial support (K.S.C.).

REFERENCES

- Maeda, S. (1989) *Biochem. Biophys. Res. Commun.* **165**, 1177–1183
- Chang, S. P. (1994) *Am. J. Trop. Med. Hyg.* **50**, 11–19
- Bernasconi, P., Li, J. P., Reagan, J. D. and Kramer, S. J. (1992) *Arch. Insect Biochem. Physiol.* **21**, 145–153
- Elipper, B. A., Stoffers, D. A. and Mains, R. E. (1992) *Annu. Rev. Neurosci.* **15**, 57–85
- Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- Schooley, D. A., Kataoka, H., Kramer, S. J. and Toschi, A. (1990) in *Insect Neurochemistry and Neurophysiology*, 1989 (Borkovec, A. B. and Masler, E. P., eds.), pp. 39–62, The Humana Press, Clifton, NJ
- Reference deleted
- Audsley, N., Coast, G. M. and Schooley, D. A. (1993) *J. Exp. Biol.* **178**, 231–243
- Gilman, A. G. (1972) *Adv. Cyclic Nucleotide Res.* **2**, 9–23
- Farndale, R. W., Allan, L. M. and Martin, B. R. (1992) in *Signal Transduction: A Practical Approach* (Milligan, G., ed.), pp. 75–103, Oxford University Press/IRL Press, Oxford
- Digan, M. E., Roberts, D. N., Enderlin, F. E., Woodworth, A. R. and Kramer, S. J. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 11074–11078
- Kataoka, H., Troetschler, R. G., Li, J. P., Kramer, S. J., Carney, R. L. and Schooley, D. A. (1989) *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2976–2980
- Ledgerwood, E. C., Brennan, S. O., Cawley, N. X., Loh, Y. P. and George, P. M. (1996) *FEBS Lett.* **383**, 67–71
- Komano, H. and Fuller, R. S. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 10752–10756
- Audsley, N., Kay, I., Hayes, T. K. and Coast, G. M. (1995) *Comp. Biochem. Physiol. [A]* **110A**, 87–93
- Parekh, R., Forrester, K. and Wittrup, D. (1995) *Protein Express. Purif.* **6**, 537–545
- Reference deleted
- Moody, A. J., Norris, F., Norris, K., Hansen, M. T. and Thim, L. (1987) *FEBS Lett.* **212**, 302–306
- Cawley, N. X., Noe, B. D. and Loh, Y. P. (1993) *FEBS Lett.* **332**, 273–276
- Blackburn, M. B., Kingan, T. G., Bodnar, W., Shabanowitz, J., Hunt, D. F., Kempe, T., Wagner, R. M., Raina, A. K., Schnee, M. E. and Ma, M. C. (1991) *Biochem. Biophys. Res. Commun.* **181**, 927–932
- Kay, I., Coast, G. M., Cusinato, O., Wheeler, C. H., Totty, N. F. and Goldsworthy, G. J. (1991) *Biol. Chem. Hoppe-Seyler* **372**, 505–512
- Lehmborg, E., Ota, R. B., Furuya, K., King, D. S., Applebaum, S. W., Ferenz, H.-J. and Schooley, D. A. (1991) *Biochem. Biophys. Res. Commun.* **179**, 1036–1041
- Kay, I., Patel, M., Coast, G. M., Totty, N. F., Mallet, A. I. and Goldsworthy, G. J. (1992) *Regul. Pept.* **42**, 111–122
- Clottens, F. L., Holman, G. M., Coast, G. M., Totty, N. F., Hayes, T. K., Kay, I., Mallet, A. I., Wright, M. S., Chung, J.-S., Truong, O. and Bull, D. L. (1994) *Peptides* **15**, 971–979
- Furuya, K., Schegg, K. M., Wang, H., King, D. S. and Schooley, D. A. (1995) *Proc. Natl. Acad. Sci. U.S.A.* **92**, 12323–12327
- Bourbonnais, Y., Bolin, D. and Shields, D. (1988) *J. Biol. Chem.* **263**, 15342–15347
- Bitter, G. A., Chen, K. K., Banks, A. R. and Lai, P. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5330–5334

-
- 28 Zsebo, K. M., Lu, H. S., Fieschko, J. C., Goldstein, L., Davis, J., Duker, K., Suggs, S. V., Lai, P. H. and Bitter, G. A. (1986) *J. Biol. Chem.* **261**, 5858–5865
- 29 Pang, S. Z., Oberhaus, S. M., Rasmussen, J. L., Knipple, D. C., Bloomquist, J. R., Dean, D. H., Bowman, K. D. and Sanford, J. C. (1992) *Gene* **116**, 165–172
- 29a Driedonks, R. A., Toschka, H. Y., van Almkerk, J. W., Schaffers, I. M. and Verbakel, J. M. (1995) *Yeast* **11**, 849–864
- 30 Jacobson, M. A., Forma, F. M., Buenaga, R. F., Hofmann, K. J., Schultz, L. D., Gould, R. J. and Friedman, P. A. (1989) *Gene* **85**, 511–516
- 31 Urdea, M. S., Merryweather, J. P., Mullenbach, G. T., Coit, D., Heberlein, U., Valenzuela, P. and Barr, P. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 7461–7465
- 32 Izumoto, Y., Sato, T., Yamamoto, T., Yoshida, N., Kikuchi, N., Ogawa, M. and Matsubara, K. (1987) *Gene* **59**, 151–159
- 33 Thim, L., Hansen, M. T. and Sorensen, A. R. (1987) *FEBS Lett.* **212**, 307–312
- 34 Norris, K., Norris, F., Bjorn, S. E., Diers, I. and Peterson, L. C. (1990) *Biol. Chem. Hoppe-Seyler* **371**, suppl. 37–42
- 35 Heim, J., Takabayashi, K., Meyhack, B., Marki, W. and Pohlig, G. (1994) *Eur. J. Biochem.* **226**, 341–353
- 36 Martin-Eaucclair, M. F., Sogaard, M., Ramos, C., Cestele, S., Bougis, P. E. and Svensson, B. (1994) *Eur. J. Biochem.* **223**, 637–645
- 37 Bayne, M. L., Applebaum, J., Chicchi, G. G., Hayes, N. S., Green, B. G. and Cascieri, M. A. (1988) *Gene* **66**, 235–244
- 38 Gabrielson, O. S., Reppe, S., Saether, O., Blingsmo, O. R., Sletten, K., Gordeladze, J. O., Høgset, A., Gautvik, V. T., Alestrøm, P., Øyen, T. B. and Gautvik, K. M. (1990) *Gene* **90**, 255–262
-

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