N-terminal methionine processing and N-α-acetylation are key determinants in intracellular protein sorting

A thesis submitted to The University of Manchester for the degree of Master of Philosophy in the Faculty of Life Sciences

2010

Gabriella M. A. Forte

Contents

List of Figures **5**
List of Tables **7**
Abstract **8**
Declaration **9**
Copyright statement **9**
Acknowledgements **10**
Abbreviations **11**
1. Introduction **15**
1.1 The secretory pathway and the Endoplasmic reticulum 15
1.2 Protein translocation across the ER membrane 17
1.2.1 Signal sequences & targeting to the ER membrane 17
1.2.2 The translocation complexes 21
1.2.3 Co-translational translocation and SRP-dependent targeting 22
1.2.4 Post-translational translocation and SRP-independent targeting 24
1.2.5 CPY & OPY model secretory substrates 27
1.3 N-terminal modification of proteins 29
1.3.1 N-terminal initiator methionine removal 29
Methionine aminopeptidases (MetAPs) 30
MetAP substrate specificity 33
1.3.2 N-α-Acetylation 36
N-α-Acetyl transferase complexes (NATs) and substrate specificity 37
1.3.3 N-Myristoylation 40
1.3.4 The N-end Rule Hypothesis 42
1.4 Coordinated binding of protein biogenesis factors at the ribosome 43
1.5 MPhil hypothesis and aims 45
2. Materials and Methods **50**
2.0 Reagents and Chemicals 50
2.1 *S. cerevisiae* strain construction 51
2.2 Plasmid constructions 55
2.3 *Escherichia coli* techniques 56
2.3.1 Growth and maintenance of *Escherichia coli* 56
2.3.2 Preparation and transformation of electro-competent *E. coli* 56

2.3.3 Preparation and transformation of $CaCl_2$ competent *E. coli* 57
2.3.4 Preparation of plasmid DNA from *E. coli* 57
2.4 *Saccharomyces cerevisiae* techniques 57
2.4.1 Growth and maintenance of *Saccharomyces cerevisiae* 57
2.4.2 Sporulation and tetrad dissection of *S. cerevisiae* 58
2.4.3 One-step transformation of *S. cerevisiae* 59
2.4.4 High efficiency transformation of *S. cerevisiae* 59
2.4.5 Isolation of microsomes from *S. cerevisiae* 60
2.4.6 Extraction of cytosol from *S. cerevisiae* 60
2.4.7 Rapid isolation of genomic DNA from *S. cerevisiae* 61
2.2.8 Total proteins extracts from *S. cerevisiae* for SDS-PAGE 62
2.4.9 Pulse labelling of *S. cerevisiae* proteins *in vivo* 62
2.5 *In vitro* DNA manipulations 62
2.5.1 Agarose Gel Electrophoresis of DNA 62
2.5.2 Extraction of DNA fragments from Agarose gels 63
2.5.3 Restriction endonuclease digestion of plasmid DNA 63
2.5.4 Alkaline phosphatase treatment of DNA fragments 63
2.5.5 Filling in DNA cohesive ends 64
2.5.6 Polymerase Chain Reaction (PCR) 64
2.5.7 Ligations of DNA fragments 64
2.5.8 Site directed mutagenesis (SDM) 65
2.5.9 DNA sequencing 65
2.6 Protein analysis techniques 66
2.6.1 SDS-PAGE analysis of proteins 66
2.6.2 Western blotting analysis 66
2.6.3 Immunoprecipitation of proteins 67
2.7 *In vitro* translation and translocation techniques 68
2.7.1 *In vitro* transcription of mRNA 68
2.7.2 *In vitro* translation/translocation assay in rabbit reticulocyte lysate 68
2.7.3 Isolation and analysis of ribosome nascent chain complex (RNC) 69
2.7.4 Translations/translocation assay in yeast cytosol 69
2.7.5 N-terminal biotinylation 70
2.8 Bioinformatics data processing 71
3. Results 73
3.1 Background 73

3.2 Mutation of position 2 of CPY signal sequence ... 73
3.3 Translocation of ppCPY position 2 signal sequence mutants *in vivo* ... 74
3.4 The affect of MetAP activity on signal sequences ... 77
3.5 The affect of N-α-Acetylation on signal sequences ... 80
3.6 An *in vitro* analysis of N-terminal modification effects on signal sequences ... 82
3.6.1 Translocation of ppαF and position 2 mutants ... 83
3.6.2 Assessing N-α-Acetylation of the ppαF position 2 serine mutant ... 85
3.7 Effect of N-terminal modification on a SRP dependent signal sequence ... 88
3.7.1 Translocation of position 2 mutants of ppOPY *in vivo* ... 88
3.7.2 Assessing N-terminal modification of the Ost1p signal sequence *in vitro* ... 89
4. Discussion ... 99
4.1 Further experimentation and studies ... 107
4.2 Concluding remarks ... 108
References ... 109
Appendix ... 120

List of Figures

1. Introduction

Figure 1.1: A schematic cartoon of the secretory pathway of *Saccharomyces cerevisiae*. 16

Figure 1.2: Common features of the signal sequence 18

Figure 1.3: Signal sequence processing and membrane protein integration 20

Figure 1.4: The translocation complexes 21

Figure 1.5: Overview of Co-translational translocation 24

Figure 1.6: Overview of Post-translational translocation 26

Figure 1.7: Schematic diagram representing CPY processing. 28

Figure 1.8: Schematic diagram showing features of type I & II MetAPs 32

Figure 1.9: Ribbon structures representing MetAPs. 32

Figure 1.10: Summary of N-terminal modification. 42

Figure 1.11: Schematic diagram of the N-end rule pathway. 43

Figure 1.12: Comparison of amino acid frequency at position 2 of proteins with signal sequences and cytosolic proteins. 47

3. Results

Figure 3.1: ppCPY signal sequence and position 2mutations. 74

Figure 3.2: Carboxypeptidase Y processing and pulse-chase in a *Δpep4* yeast strain 75

Figure 3.3: N-terminal modification leads to defective ER translocation of CPY 76

Figure 3.4 Schematic diagram of MetAP activity in the absence & presence of Fumagillin. 78

Figure 3.5: N-terminal Methionine removal affects CPY translocation 79

Figure 3.6: N-α-Acetylation affects translocation of CPY mutants to the ER 82

Figure 3.7: ppαF signal sequence mutation at position 2. 83

Figure 3.8: Translocation assay for *in vitro* expressed ppαF 84

Figure 3.9: Translocation assay for *in vitro* expressed lysine-less ppαF 86

Figure 3.10: Assessing N-α-Acetylation of ppαF 87

Figure 3.11: ppOPY signal sequence mutation at position 2 88

Figure 3.12: Analysis of ppOPY translocation *in vivo* 89

Figure 3.13: Translocation assay for *in vitro* expressed lysine-less ppOαF 90

Figure 3.14: Assessing N-α-Acetylation of ppOαF 92

Figure 3.15: Immunoprecipitation of truncated ppOαF 94

Figure 3.16: Assessing N-α-Acetylation of truncated ppOαF 95

Figure 3.17: Biotinylation free N-termini of ppOαF 97

Discussion

Figure 4.1: Combined analysis of signal sequence peak hydrophobicity and predicted N-terminal processing 104

Figure 4.2: Bias for retention of an unprocessed methionine appears to be conserved in the signal sequences of higher eukaryotes 105

Figure 4.3: Prediction retention of N-terminal methionine in the subcellular localisations of *S. cerevisiae* 106

List of Tables

Table 1.1: N-Met removal form substrates in Prokaryotes and Eukaryotes 33

Table 1.2: Physical properties of Amino Acid side chains 33

Table 1.3: Subunits of the NAT complexes 38

Table 2.1: *Escherichia coli* strains used in this study 50

Table 2.2: *S. cerevisiae* strains used in this study 50

Table 2.3: DNA primers used in this study 52

Table 2.4: Plasmids used in this study 54

Table 3.1: N-terminal modification prediction of CPY and P2 mutants by Termi*N*ator 81

Abstract

N-terminal initiator methionine removal and N-α-acetylation represent two of the most common protein modification events in eukaryotes. N-terminal methionine removal occurs on two thirds of a given proteome and N-α-acetylation between 50-80% of cytosolic proteins. These modifications occur co-translationally as the nascent protein is being synthesised and therefore are amongst the earliest protein modification events. The primary determinant for these modifications is the N-terminal amino acid residue at position 2. Many secretory proteins are targeted via N-terminal signal sequences. Bioinformatics data has revealed that the signal sequences of secretory proteins have a strong observed bias against certain amino acids at position 2, specifically residues which are predicted to promote the most common N-terminal modifications (manuscript submitted Forte *et al*). In this study we focused upon secretory proteins which display this bias against N-terminal modification. It was found that mutation of these secretory signal sequences in a fashion that may promote N-terminal modification results in defective sorting of these proteins to the cytosol. This phenomenon was further investigated *in vitro* demonstrating the mis-sorted proteins are N-α-acetylated and *in vivo* showing that inhibition of N-α-acetylation rescued the defect. These findings indicate that N-α-acetylation can affect sorting of secretory proteins. Further bioinformatics data presented in this study show that proteins targeted to other subcellular locations via their N-termini may also exhibit a bias against N-terminal processing similar to signal sequences. Taken together these observations may indicate that N-terminal modification represents a novel sorting step in protein biogenesis.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trademarks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (seehttp://www.campus.manchester.ac.uk/medialibrary/policies/intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester.ac.uk/library/aboutus/regulations) and in The University's policy on presentation of Theses.

Acknowledgements

I would like to thank my supervisor Professor Colin Stirling for his support, guidance and encouragement throughout my MPhil. I would also like to acknowledge Dr M Pool for his contributions to this work.

Many thank also to all the members of the Stirling lab both past and present for their support and discussion, in particular Dr Barrie Wilkinson for his very useful and knowledgeable guidance. I also owe many thanks to Dr Timothy Tavender for his advice, support and generally "putting up" with me throughout my studies and beyond! It goes without saying (but I should mention it anyway) that I thank my parents and family for providing support despite not being entirely clear what I do or why I do it!

This work is supported by the Wellcome Trust and also the unfortunately abbreviated STDU (Staff Training Development Unit).

Abbreviations

ADP	Adenosine-5’-diphosphate
APS	Ammonium persulphate
ATP	Adenosine-5’-triphosphate
BiP	B cell immunoglobulin heavy chain binding protein
bp	Base pair
C-	Carboxy
ddH_2O	double distilled water
DNA	Deoxyribo nucleic acid
dNTPs	Deoxynucleoside triphosphates
DTT	Dithiothreitol
ECL	Enhanced chemiluminescence
EDTA	Ethylenediaminetetra-acetic acid
EGTA	Ethylene glycol-bis(2-aminoethylether)-*N,N,N′,N′*-tetraacetic acid
ER	Endoplasmic reticulum
GTP	Guanosine-5’-triphosphate
HEPES	N-92-hydroxylethyl piperazine-*N’*-(2-ethanesulphonic acid)
HRP	Horse radish peroxidase
Hsp	Heat shock protein
IP	Immunoprecipitation
IPTG	Isopropyl β-D- thiogalactopyranoside
Kb	Kilo base pair
kDa	Kilo Dalton
LB	Luria-Bertani media
MetAP	Methionine aminopeptidase

mRNA	Messenger ribonucleic acid
N-	Amino
NAT/Nat	N-Acetylase
NP40	Nonidet P40
OAc	Acetate
OD_{xxxnm}	Optical density at xxx nano meters
ORF	open reading frame
OST	Oligosaccharyl transferase
P2	Position 2
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
ppCPY	pre-pro-Carboxypeptidase Y
ppOPY	pre-pro-Ost1-Carboxypeptidase Y
ppαF	pre-pro-α-factor
RAMP	Ribosome-associated membrane protein
RNA	Ribonucleic acid
RNase	Ribonuclease
rNTPs	Ribonucleoside triphosphates
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	Secretion
SP	Signal peptide
SR	Signal recognition particle receptor
SRP	Signal recognition particle

TEMED	*N,N,N',N'*-tetramethylethylenediamine
TMD	Transmembrane domain
Tris	Tris-(hydroxymethyl) aminomethane
YNB	Yeast nitrogen base media
YP	Yeast peptone media
YPD	Yeast peptone dextrose media

The three letter and single letter code for amino acids was used throughout this thesis:

Amino Acid	**Three letter code**	**Single letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate / Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamate / Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Iso	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Try	Y
Valine	Val	V

Introduction

1. Introduction

1.1 The secretory pathway and the Endoplasmic reticulum

Eukaryotic cells are internally divided into membrane bounded compartments commonly known as organelles. Each organelle is a distinct micro-environment containing a defined complement of proteins.

The majority of proteins are synthesised on cytosolic ribosomes, despite a large proportion of these proteins requiring transport to the various cellular organelles to carry out their functions. In many cases e.g. the mitochondria, peroxisomes and nucleus, the synthesised proteins are directly targeted from the cytosol to their final destination. However, for proteins destined to be secreted out of the cell, integrated into the plasma membrane or transported to organelles such as vacuoles and endosomes, this feat is mainly achieved by protein trafficking through the secretory pathway (Figure 1.1). The secretory pathway is a major route by which proteins are targeted to their ultimate cellular destination (Palade, 1975). The entry point of secretory pathway is via the endoplasmic reticulum (ER) an organelle which is an extensive network of interconnected tubules and cisternae formed from a continuous membrane. Proteins trafficked via the secretory pathway cross the ER membrane via an integrated proteinaceous channel, in a process termed translocation (Blobel and Dobberstein, 1975) the channel is commonly known as the translocon.

Proteins translocated into the ER can pass through the compartments of the secretory pathway and may be subjected to a number of post-translational modifications. The modifications that can occur in ER lumen include disulphide bond formation, *N*-linked glycosylation and subsequent trimming of certain glucose and mannose residues. If the proteins are transported to the Golgi further protein modifications may take place including *O*-linked glycosylation, phosphorylation and addition of mannose residues to *N*-linked glycans. The modifications which occur are usually essential for the protein to reach its final conformation. Proteins that do not have a motif for retrieval to the ER or retention in the Golgi may be trafficked via secretory

vesicles to their destination which may be any of the other compartments of the secretory pathway, plasma membranes or the exterior/surface of cell.

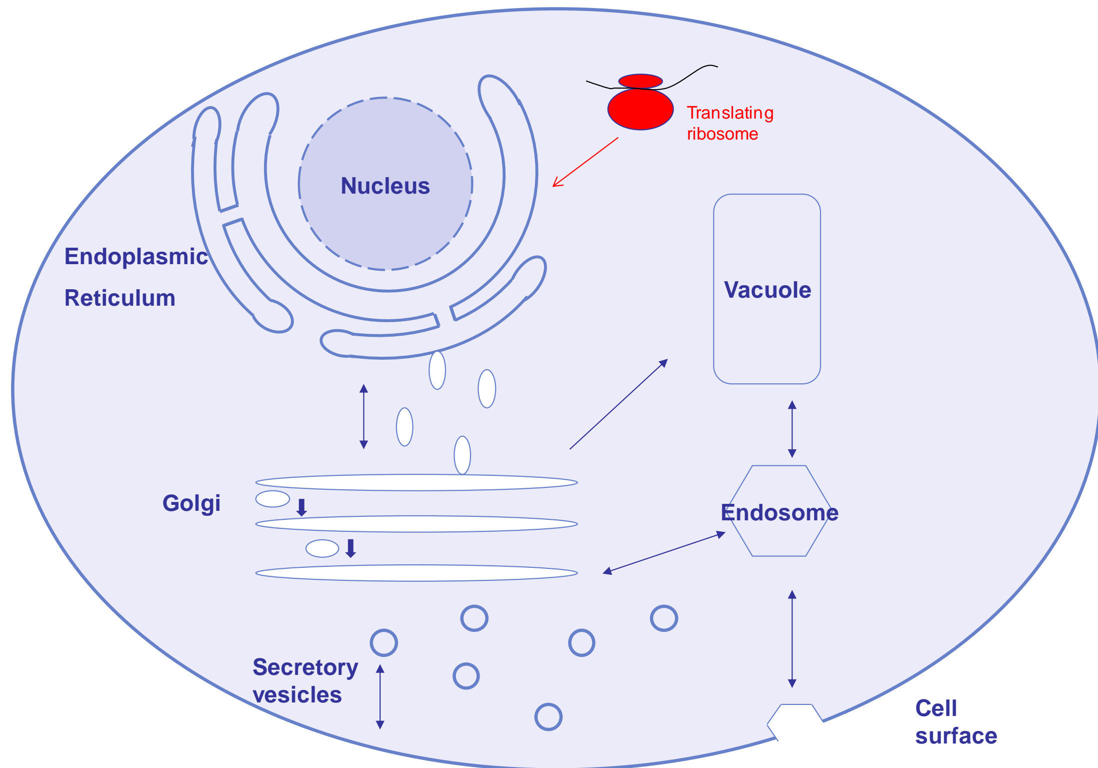


Figure 1.1 A schematic cartoon of the secretory pathway of *Saccharomyces cerevisiae*. The translation of a protein on a cytosolic ribosome indicates the initial step of the secretory pathway. Secretory polypeptides are targeted to the ER were they can enter the secretory pathway via a process termed translocation. Once in the ER, folding and modification, such as glycosylation may take place. If the protein is trafficked to the Golgi further modification may occur there. From the Golgi, secretory vesicles may traffic polypeptides to other subcellular compartments, the cell surface or the extracellular environment. Proteins may also be retained within or integrated into membranes of the secretory components in this pathway.

1.2 Protein translocation across the ER membrane

Proteins that enter the secretory pathway may be destined for either secretion out of the cell, internalisation into another cellular compartment (shown in figure 1.1) or for integration into membranes. Those proteins which are secreted or internalised are fully translocated across the ER membrane into the lumen, whereas membrane proteins are integrated into the ER membrane as they are translocated and this process will be described in further detail in section 1.2.1.

Prior to translocation proteins must first be targeted to the ER membrane. Targeting to the ER occurs via an amino-terminal (N-terminal) signal sequence or in the case of some integrated membrane proteins via a transmembrane domain (TMD), a region of the protein that integrates into and spans the ER membrane. These targeting sequences are described further in section 1.2.1. Targeting can occur either as the protein is being synthesised (co-translationally) or once translation is complete (post-translationally), as will be described below in sections 1.2.3 and 1.2.4 respectively. These fundamental mechanisms of targeting and translocation have comparable pathways in prokaryotes. However, this report will focus principally upon eukaryotic systems, in particular the yeast *Saccharomyces cerevisiae* as a model organism.

1.2.1 Signal sequences & targeting to the ER membrane

As indicated above, Proteins destined for the secretory pathway often have an N-terminal signal sequence which targets them to the ER for translocation. This sequence has a key role in determining whether a protein is targeted for translocation via the co- or the post-translational pathway. Proteins targeted co-translationally typically have a more hydrophobic signal sequence compared with those targeted post-translationally (Ng et al., 1996). Proteins commonly have cleavable signal sequences, meaning that once the protein is translocated or integrated into the ER membrane the targeting signal is enzymatically removed. However, some integrated membrane proteins lack a cleavable signal sequence and instead are targeted via a transmembrane domain (TMD) usually known as a signal anchor, this is a

hydrophobic region of the protein which inserts into the membrane and remains a structural or functional part of the translocated protein.

Signal sequences (also referred to as signal peptides or leader sequences) have little consensus in terms of actual amino acid sequence but characteristically have three distinct regions (Figure 1.2). The N-region at the amino terminus can vary considerably in length and has a net positive charge usually due to the inclusion of lysine or arginine residues. The hydrophobic H-region is typically rich in leucine, with 25-40% leucine across a mean overall length of 12-15 residues (von Heijne, 1985). The final C-region includes a cleavage site for signal peptidase. This is characterised by small uncharged amino acids such as alanine in positions -3 and -1 relative to the cleavage site (von Heijne, 1984). Evidence suggests that hydrophobicity of a signal sequence is key to whether a polypeptide is recognised by the signal recognition particle (SRP), a complex required for co-translational targeting, and subsequently is a substrate for co- rather than post-translational translocation pathway (Ng et al., 1996). However, it has also been suggested that the structural orientation/conformation of a signal sequence as it exits the ribosome may also be important for SRP interaction (Matoba and Ogrydziak, 1998).

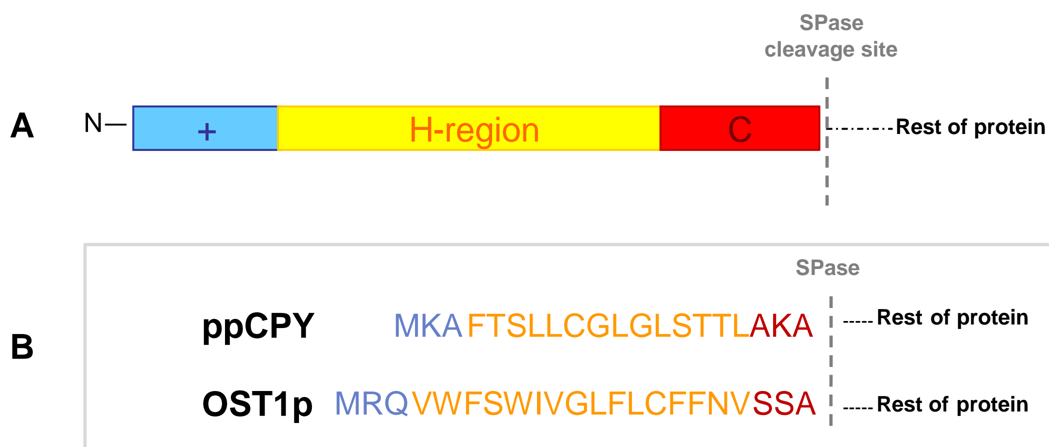


Figure 1.2 Common features of the signal sequence
A) Regions of a typical signal sequence, + positive N-region, Hydrophobic H-region and C region containing the SPase cleavage site. **B)** Signal sequences of ppCPYp which is targeted post-translationally (independently of SRP) and Ost1p which is targeted co-translationally via SRP.

When a cleavable signal sequence enters the channel of the translocon, it is inserted into the membrane creating a loop (as depicted in Figure 1.3A). As more of the polypeptide emerges into the lumen of ER, signal peptidase cleaves the signal sequence from the polypeptide (von Heijne, 1984). In mammals it has been shown that the cleaved signal sequence is then further processed by a second enzyme, signal peptide peptidase (SPPase), and is released into the cytosol or the ER lumen (Lyko et al., 1995). To date there is no evidence that this further processing step occurs in yeast.

Membrane proteins are integrated into the membrane via their hydrophobic TMDs. In some cases the first TMD to be integrated will be the one which targets the protein to the ER membrane, but for many others it may be a TMD that is translocated following signal sequence cleavage. In either case TMDs insert into the membrane in a similar way to a cleavable signal sequence, as the protein is translocated.
Depending on the sequence that follows a TMD, a protein may insert in one of two orientations. Type I signal anchored proteins have their C-terminus in the cytosol and N-terminus in the ER lumen (Figure 1.3B) while Type II signal anchored proteins are in the reverse orientation with their C-terminus in the ER lumen (Figure 1.3C). If the protein is polytopic (containing a number of TMDs) the orientation of the second TMD will differ from that of the first and so on. Thus correct orientation of each TMD depends on proper orientation of the first (Figures 1.3B&C right panel after dotted arrows) (Rothman et al., 1988, Yost et al., 1990)

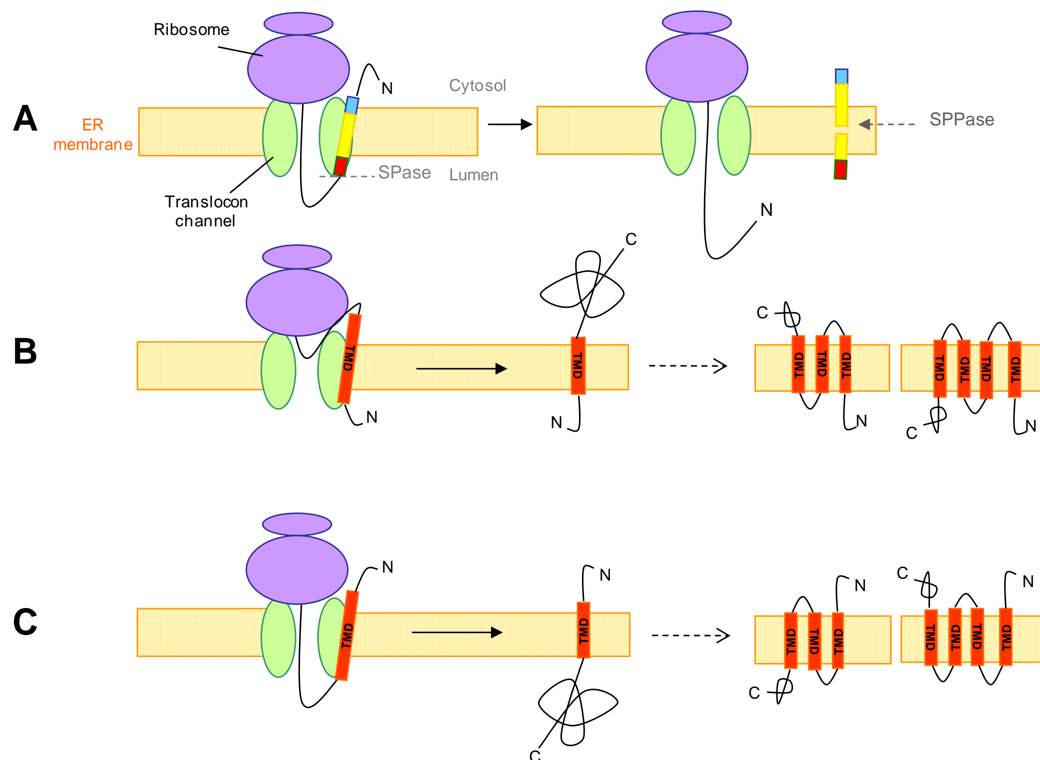


Figure 1.3 Signal sequence processing and membrane protein integration
Panel **A** depicts cleavage of the membrane integrated signal sequence by SPase and the subsequent processing by SPPase cleaving the signal sequence into smaller fragments. **B** type I membrane anchor integration and **C** integration of a type II membrane anchor. Right panels show the topologies of polytopic membrane proteins which may follow.

1.2.2 The translocation complexes

Although targeting of proteins to the ER membrane can occur via different mechanisms, many of the components utilised to translocate the protein across or into the ER membrane are shared (Figure 1.4). These components are mainly integrated membrane proteins in a complex, often referred to as the translocon. The translocon is formed from the components of two smaller complexes, the Sec61 complex (Sec61p, Sbh1p, and Sss1p) and the Sec63 complex (Sec63p, Sec62p, Sec71p, and Sec72p) (Deshaies et al., 1991, Panzner et al., 1995, Finke et al., 1996) of these proteins Sbh1p, Sec71p and Sec72p are not essential for viability (Deshaies and Schekman, 1990, Feldheim et al., 1992, Rapoport et al., 1999). These complexes are those of the yeast system nevertheless the Sec61 complex is highly conserved in higher eukaryotes (Hartmann et al., 1994, Van den Berg et al., 2004).
The ER luminal Hsp70 chaperone Kar2p has also been shown to have a role in both co- and post-translational translocon complexes (Panzner et al., 1995, Young et al., 2001) but its role in the co-translational pathway is not well understood.

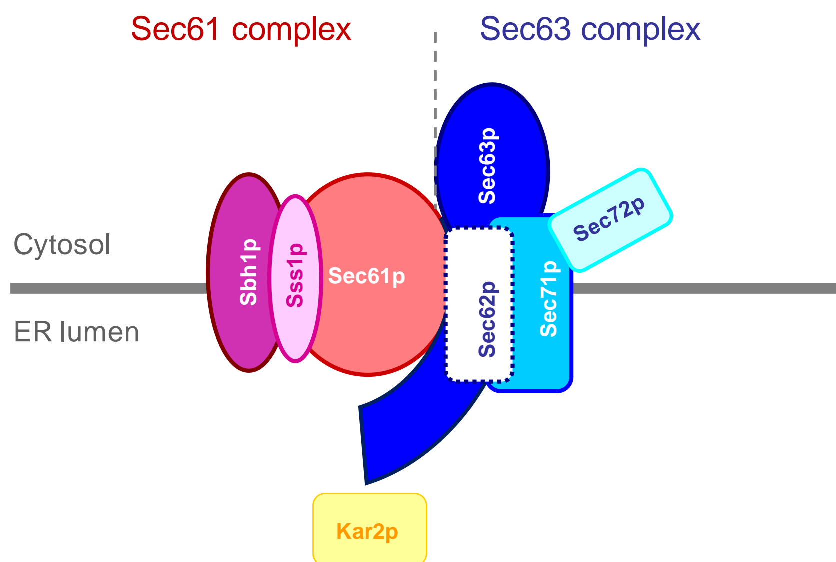


Figure 1.4 The translocation complexes
The translocon complexes used for co- and post- translational translocation, SEC’ & SEC respectively, share many components of the Sec61 and Sec63 complexes. The SEC translocon is a heptameric complex which includes all of the protein components illustrated above, the SEC’ translocon is a hexameric complex differing from SEC complex by the exclusion of Sec62p (Jermy et al., 2006). Kar2p is not part of the complex but is required for both SEC & SEC’ complexes for translocation.

The pore through which the polypeptide is translocated is commonly believed to be formed by the components of the Sec61 complex (Gorlich and Rapoport, 1993, Panzner et al., 1995, Hanein et al., 1996, Beckmann et al., 1997, Hamman et al., 1997). There has been much debate as to whether the translocon channel is formed by a monomer, dimer or oligomers of the Sec61 translocon complex. This issue is difficult to resolve based on the data available, although most authors favour models including either a monomer or dimer (Mitra et al., 2006, Rapoport, 2007). The major issue is whether a monomer can form a large enough channel and can mechanistically cope with protein insertion into and across a membrane as effectively as a dimer. Those authors who choose the dimer model are also divided as to orientations of the two complexes forming the dimer (reviewed in Mitra et al., 2006).

The Sec63 complex is essential for post-translationally targeting. However, Sec63p is also essential for co-translational translocation (Young et al., 2001). It is not entirely clear how it functions but there is evidence it stabilises the translocon complex (Jermy et al., 2006), and also recruits and activates chaperones of the ER lumen (McClellan et al., 1998, Misselwitz et al., 1998). Homologues of Sec63p and Sec62p are present in mammals but their function is unknown.

1.2.3 Co-translational translocation and SRP-dependent targeting

Ribosomes translating secretory polypeptides can be targeted to the ER membrane where translocation of these polypeptides can be coupled with their translation on ribosomes (co-translational). Co-translational translocation occurs in all eukaryotes and is the route by which the majority of identified ER proteins are translocated (Rapoport, 2007).

The process of targeting a translating ribosome to the ER membrane is dependent upon a ribonucleoprotein complex known as the signal recognition particle (SRP) (Walter et al., 1981). The SRP complex is composed of six proteins (Walter and

Blobel, 1980) and one RNA (Srp14p, Srp21p, Srp54p, Srp65p, Srp68p, Srp72p & ScR1 in yeast) (Keenan et al., 2001).

During synthesis of polypeptides destined for ER translocation, the first residues synthesised are those of the N-terminal signal sequence or signal anchor. As translation proceeds, these residues enter the ribosome exit tunnel of the translating ribosome and this induces SRP to be recruited to the ribosome (Berndt et al., 2009). As translation continues the signal sequence emerges and in turn SRP binds to this region of the ribosome nascent chain complex (RNC). When SRP binds the RNC further translation of the nascent protein is stalled (Figure 1.5) (Walter and Blobel, 1981). SRP binds to the L25a and L35 ribosomal proteins close to the exit site, the Srp54 subunit of SRP is bound to a GTP molecule (Pool et al., 2002, Walter et al., 1981, Bacher et al., 1996). SRP directs its associated stalled ribosome to the ER membrane by interacting with the ER membrane integrated SRP-receptor (SR) which is also bound to a GTP molecule. Both SRP and SR hydrolyse GTP causing a conformational change in SRP and its subsequent dissociation from the RNC and SR exposing the ribosome's binding site for the ER translocon complex (Rapiejko and Gilmore, 1997, Connolly and Gilmore, 1993, Miller et al., 1993, Halic et al., 2006). The released SRP can go on to assist in further cycles of targeting. It may be the case that at this point in the pathway that the SRP-receptor also interacts with components of the translocon (Jiang et al., 2008). When the ribosome binds the translocon translation resumes, as the polypeptide elongates from the ribosome it is directed through the channel of the translocon into the ER lumen (Rapoport, 2007). It has also been shown that Sec63p and Kar2 are required for co-translational translocation (Young et al., 2001, Hamman et al., 1998), and it has been suggested that they act together to form a plug on the luminal side of the translocon channel and therefore play a role in translocon gating to maintain segregation of the cytosol and ER lumen (Hamman et al., 1998, Haigh and Johnson, 2002).

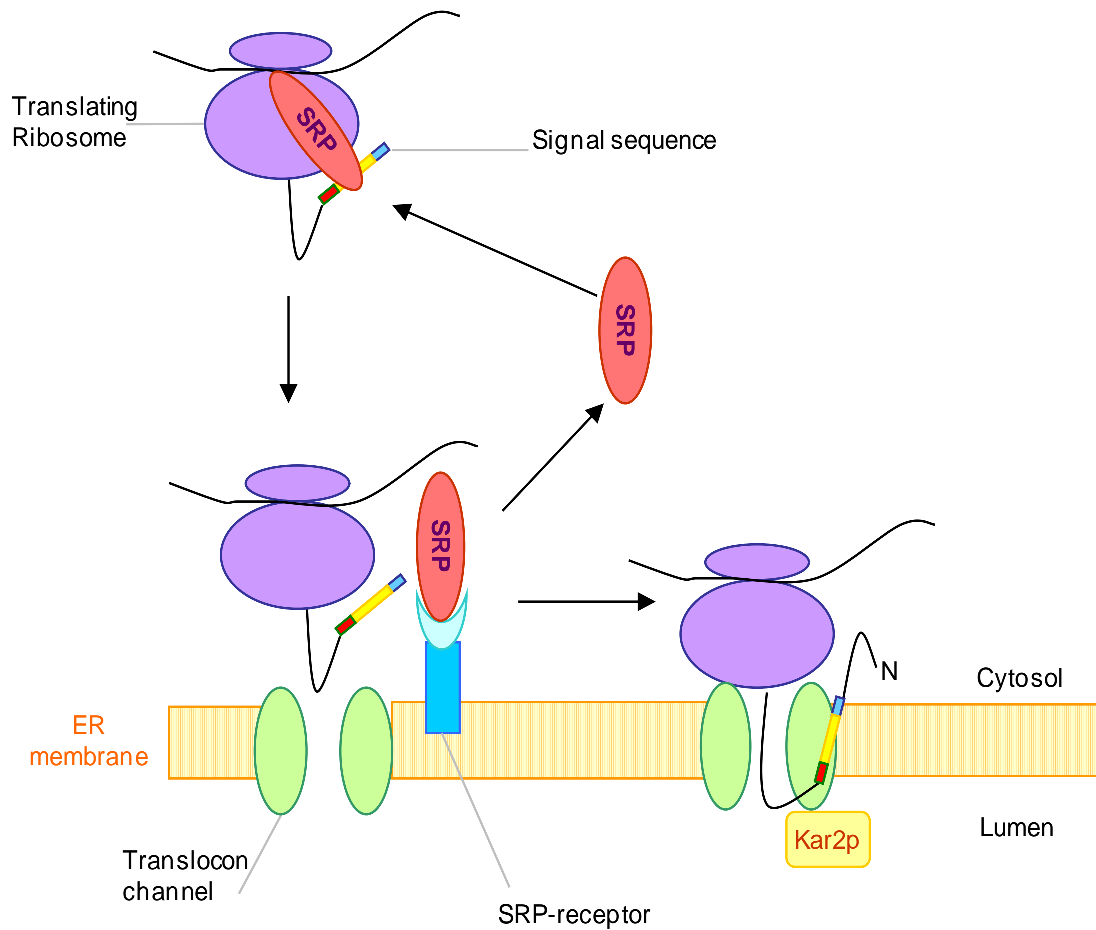


Figure 1.5 Overview of Co-translational translocation
As a signal sequence emerges from a translating ribosome, the RNC-SRP complex associates with it and directs it to the ER membrane by binding to the SRP-receptor. SRP dissociate from the ribosome, which is then transferred to the translocon and thus translocation of the translating protein commences.

1.2.4 Post-translational translocation and SRP-independent targeting

The post-translational targeting pathway is utilised by secretory polypeptides that are completely synthesised prior to their transport to the ER. Targeting of these proteins occurs independently of SRP (Figure 1.6) (Stirling and Hewitt, 1992) and again these substrates are also targeted to the ER membrane with the aid of a signal sequence or signal anchor. It is commonly believed that the signal sequences of the substrates that use this post-translational pathway are likely to be less hydrophobic than those targeted via SRP (Ng et al., 1996).

The mechanisms of SRP-independent targeting are not as well characterised as those of the SRP-dependant pathway, there is a requirement for cytosolic chaperones but neither ribosomes nor SRP are required. For these reasons it is understood that the mechanism of targeting and translocation must be significantly different.

As the polypeptide substrates which use this pathway are completely translated, cytosolic chaperones are required to maintain the polypeptide in an incompletely folded and translocation-competent conformation. These cytosolic chaperones include those belonging to the Hsp70 family, Ssa1p and Ssa2p (Chirico et al., 1988, Deshaies et al., 1988). These Hsp70 chaperones interact with their co-chaperone Ydj1p of the Hsp40 family, which is able to stimulate their ATPase activity (Cyr et al., 1992). The resulting chaperone-bound nascent polypeptide complex is targeted to the ER membrane via an as yet uncharacterised mechanism where it is suggested that Ssa1p/Ssa2p undergo nucleotide exchange and the polypeptide is released (Plath and Rapoport, 2000). This polypeptide associates with components of the Sec63 complex (Sec62p, Sec71p and Sec72p) which has been shown to have signal sequence binding activity and it has been suggested that the Sec63p complex brings polypeptide substrate to be translocated close to the translocon channel (Lyman and Schekman, 1997). As no ribosome is present to direct the substrate polypeptide away from the cytosol and towards the ER lumen, it is believed that the polypeptide is pulled through the channel into the ER lumen by the action of an ATP bound lumenal chaperone (Kar2p in *S. cerevisiae*) associated with the J-domain of Sec63p (Misselwitz et al., 1998). The J-domain of Sec63p promotes Kar2p hydrolysis of bound ATP to ADP causing a conformational change in Kar2p and allowing it to trap the hydrophobic portions of the translocating polypeptide (McClellan et al., 1998, Panzner et al., 1995). As the polypeptide emerges into the ER lumen multiple Kar2p molecules can bind the polypeptide in the same manner, while other Kar2p molecules release the substrate due to a conformational change following nucleotide exchange and go on to bind to other portions of the translocating substrate. This continuous cycle of Kar2p molecules binding and releasing the substrate as ATP is hydrolysed means Kar2p molecules can act as a molecular ratchet pulling the substrate into the ER lumen (Matlack et al., 1999, Rapoport et al., 1999). Kar2p is

believed to use the ER luminal proteins Sil1p, Lhs1p as nucleotide exchange factors (Kabani et al., 2000, Steel et al., 2004).

Whilst post-translational translocation may occur in mammalian cells it is not as well characterised as in *S. cerevisiae* and there are few suggested substrates. In yeast prepro-Carboxypeptidase Y (ppCPY) is a well characterised substrate for this targeting pathway and will be discussed further below.

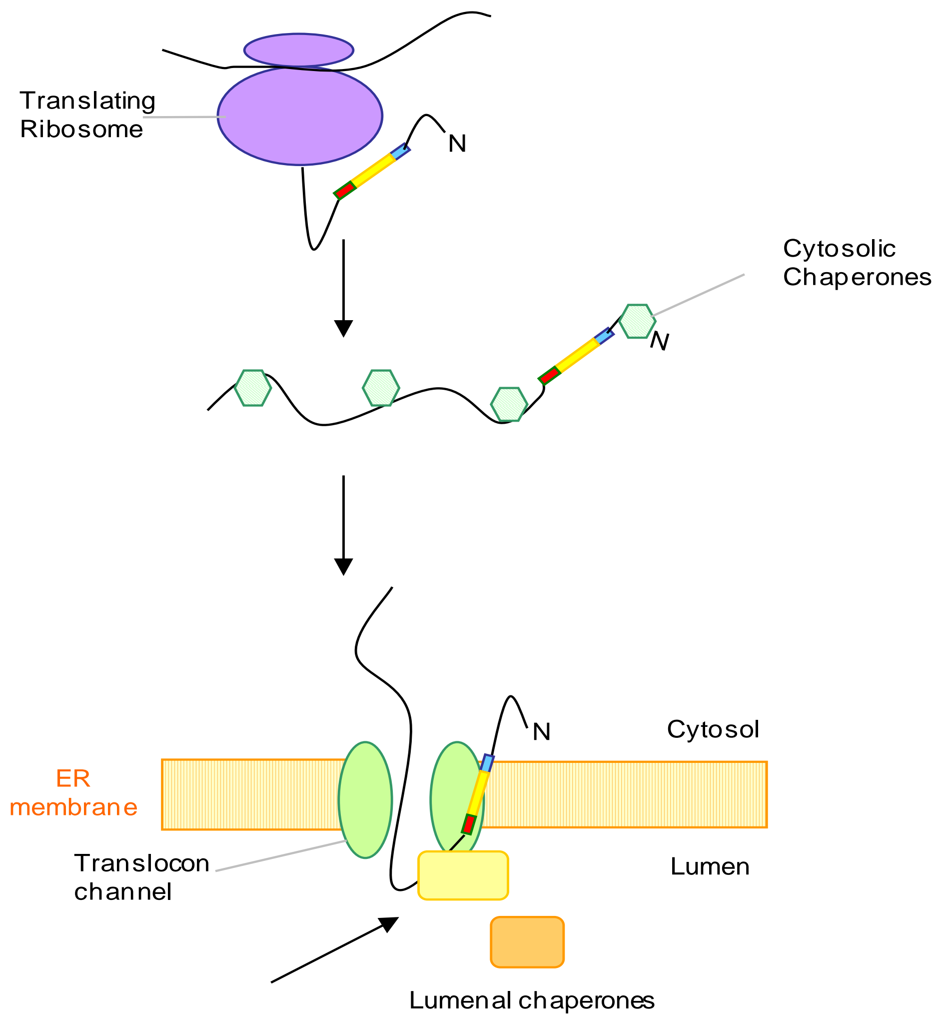


Figure 1.6 Overview of Post-translational translocation
Polypeptides which are completely translated in the cytosol are maintained in a translocation competent state by cytosolic chaperones prior to targeting to the ER. These polypeptides are translocated through the translocon, with the assistance of ER lumenal chaperones pulling the polypeptide though the translocon.

1.2.5 CPY & OPY model secretory substrates

The protein prepro-Carboxypeptidase Y (ppCPY) of *S. cerevisiae* is a well characterised post-translationally translocated secretory protein. ppCPY undergoes extensive processing and modification (see Figure 1.7) as it is trafficked through the secretory pathway to the vacuole, where its role is to hydrolyse peptides, esters, amides and anilides. (reviewed by Jung et al., 1999)
Prepro-CPY is 532 amino acid residues in length (~60kDa) and encoded by the gene *PRC1*. This protein is targeted to the ER via a signal sequence (20 residues) which is removed to give pro-CPY (p-CPY (~58kDa)) this form is transient Figure 1.7) as four *N*-linked glycans are added to the protein as it enters the ER. The glycosylation acceptor sites are located in the mature region of the protein (asparagines 13, 87, 168, 368) generating the intermediate form termed p1-CPY (~67kDa) (Stevens et al., 1982, Holst et al., 1996). Correctly folded p1-CPY is trafficked to the Golgi where the glycans added in the ER are modified by addition of mannose residues giving rise to p2-CPY (~69kDa) (Ballou et al., 1990). Finally p2-CPY is trafficked to the vacuole, via an internal sequence, where multi-step removal of the pro-peptide (91 residues) by hydrolases forms mature CPY (~61kDa). This pro-peptide is essential for *in vivo* folding in the ER (Ramos et al., 1994) and also has a role in maintaining CPY in its inactive form. To activate CPY the pro-peptide is cleaved firstly by proteinase A, (product of the *S. cerevisiae* gene *PEP4*), which removes 56 residues, then a further 30 residues are removed by proteinase B and the remaining 5 residues by aminopeptidases (Sorensen et al., 1994, Van Den Hazel et al., 1996). Consequently, if the gene *PEP4* is deleted from *S. cerevisiae*, p2-CPY accumulates within the cell as it cannot be processed to the mature CPY (Rupp and Wolf, 1995).

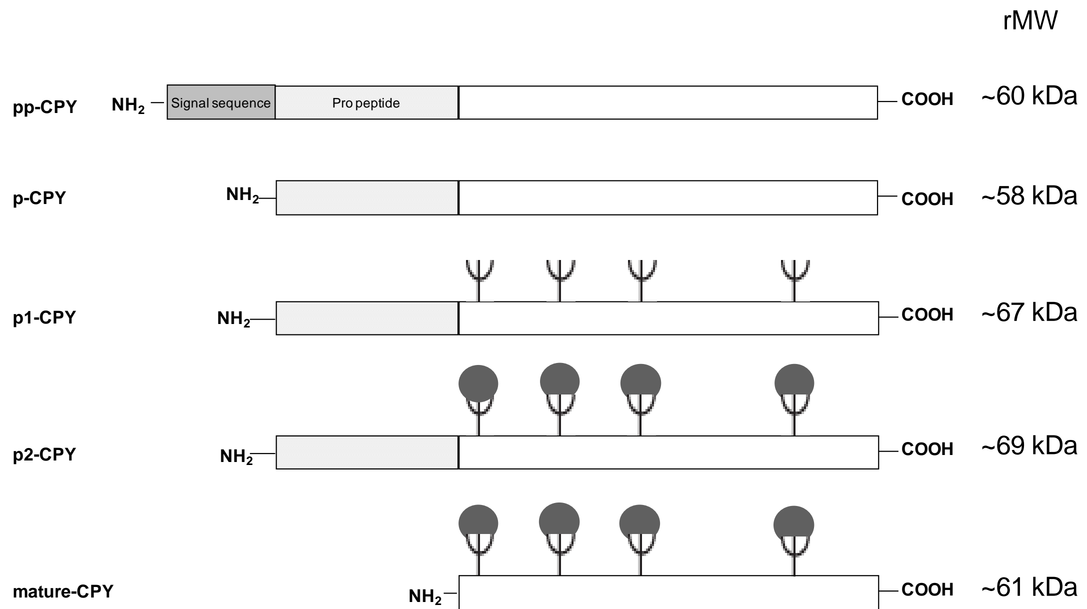


Figure 1.7 Schematic diagram representing CPY processing.
Prepro-CPY, a model secretory pathway substrate, enters the secretory pathway post-translationally via translocation to the ER. Here it is processed to the p1-CPY by signal sequence removal and glycosylation. The p1-CPY is then trafficked to the Golgi where the glycans are modified to p2-CPY. The pro peptide contains a sequence (QPRL) which target p2-CPY to the vacuole, were it is processed to mature-CPY by pro-peptide removal. p-CPY is generated by treatment of cell with the drug Tunicamycin which inhibits glycosylation. Also indicated is the relative molecular weight (rMW) of each species in kDa.

Prepro-OPY (ppOPY) is a laboratory construct which expresses a variant of ppCPY in which the signal sequence of CPY has been replaced with that of Ost1p. The Ost1p signal sequence is more hydrophobic than that of ppCPY and therefore allows recognition by SRP, making ppOPY a substrate for co-translational rather than post-translational translocation (Ng et al., 1996, Willer et al., 2008). Aside from its alternative targeting, further processing of ppOPY is unaltered with the pro-CPY region of the construct proceeding through its maturation pathway as normal after signal sequence cleavage (Willer et al., 2008). Therefore ppOPY and ppCPY provide a useful pair of tools when analysing factors influencing entry into, and passage through, the yeast secretory pathway.

1.3 N-terminal modification of proteins

Using ppCPY as an example it is evident that proteins can undergo extensive co- and post-translational modification such as glycosylation and proteolytic cleavage. Other modifications observed within cells include formation of disulphide bonds between cysteine residues (also observed for CPY), acetylation and phoshorylation. These modifications often occur due to specific amino acid motifs encoded in the protein amino acid sequence. Many of these modifications are important as they are believed to influence activity, turnover/stability and cellular localisation of proteins and may be crucial for biological function.

There are a number of modifications which occur specifically at the N-terminus of a protein and are determined by the protein's N-terminal amino acid sequence.
N-terminal modification is of particular interest due to the large proportion of proteins which may be affected and also because these events often occur at a very early stage of protein biogenesis, usually occurring co-translationally as the first few N-terminal amino acids emerge from the ribosome. The most common of these N-terminal modifications events are described below.

1.3.1 N-terminal initiator methionine removal

All eukaryotic proteins are synthesised with an N-terminal initiator methionine as the first amino acid. In many cases this methionine can be removed co-translationally by methionine aminopeptidases (MetAPs) located in the cytosol (Adams and Capecchi, 1966, Huang et al., 1987). It is believed that this methionine removal only occurs if the adjacent second residue is small or uncharged (see section MetAP substrate specificity). In prokaryotes and plastids proteins are synthesised with an initial formylmethionine, a methionine which has an N-formyl group attached. Polypeptides synthesised with a formylmethionine must first be deformylated by peptide deformylases (PDFs) before MetAPs can remove the initiator methionine (Adams and Capecchi, 1966, Solbiati et al., 1999).

Initiator methionine removal is a potentially widespread event and it has been suggested to occur for two thirds of protein species in any given proteome (Giglione et al., 2000, Meinnel et al., 1993, Serero et al., 2003). It will become apparent in later sections that initial methionine removal is an important prerequisite for other N-terminal modifications, although it can also occur independently of other modifications. Therefore, N-terminal methionine removal may itself play an important role in protein biogenesis.

Methionine aminopeptidases (MetAPs)

The Methionine aminopeptidases (MetAPs) enzymes responsible for initiator methionine removal are present in all organisms, with Eukaryotes possessing two types encoded by the genes *MAP1* (Type I; MetAP1) and *MAP2* (Type II; MetAP2) in *S. cerevisiae*. In contrast Eubacteria have only one MetAP corresponding to the eukaryotic type I, and Archae only type II. These enzymes act co-translationally and thus are located in the cytosol of eukaryotes. However, isoforms similar to bacterial MetAPs have been found to localise to plastids of *Arabidopsis thaliana* via N-terminal pre-sequences (Giglione et al., 2000). This indicates that some proteins synthesised in these organelles or following localisation to these compartments require initiator methionine removal.

The process of initiator methionine removal appears to be an essential process in prokaryotes as deletion of the *E.coli MAP* gene is lethal (Chang et al., 1989). Similarly Li *et al* showed that while a deletion of either *MAP1* or *MAP2* alone in *S. cerevisiae* is viable (Li and Chang, 1995), albeit having slow growth phenotypes, a deletion of both genes results in non viable cells. This indicates a degree of functional redundancy between type I and type II MetAPs but demonstrates that the process of initiator methionine removal is essential for eukaryotes also. The fact that distinct MetAPs have been conserved throughout evolution and remain essential across species clearly indicates a fundamentally important role for their activity within cells.

Type I and type II MetAPs can be distinguished initially on the basis of structural differences in the C-terminal region which forms the active site (Figure 1.8). Both types of MetAPs have five conserved amino acid residues in this region (D219, D230, H294, E327and Q358 in MetAP1 of *S. cerevisiae*) and these residues have two closely associated metal ions (Roderick and Matthews, 1993, Arfin et al., 1995). The identity of these metal ions is unclear, although candidates include Zn^{2+}, Co^{2+}, Mn^{2+} and Fe^{2+} (Leopoldini et al., 2007, Li et al., 2003, Walker and Bradshaw, 1998). A major difference between type I and type II MetAPs in this region is that type II enzymes also have a 60 amino acid insertion between the fourth and fifth conserved residues, which is not present in type I MetAPs. In addition, whilst the active sites of both types of MetAPs are thought to be able to accommodate up to two amino acids of a substrate protein, MetAP1 possesses a slightly smaller binding pocket due to different amino acid side chains orientated in this region (Liu et al., 1998, Roderick and Matthews, 1993). As will be discussed this difference potentially influences the specificity of activity of the two types of MetAP. In addition this difference also allows us to make use of inhibitors which differentially bind the active sites of the two types of MetAPs. Fumagillin and its derivatives specifically and irreversibly inhibit MetAP2, only having a weak interaction with MetAP1. On the other hand Barbituric acid derivatives specifically inhibit MetAP1 (Addlagatta and Matthews, 2006, Chen et al., 2004, Haldar et al., 2008, Liu et al., 1998).

As well as differences existing between type I and type II MetAPs, there are also structural differences between each eukaryotic MetAP and their corresponding prokaryotic forms (Figure 1.8). Eukaryotic MetAPs have an additional N-terminal extension not present in the prokaryotic enzymes. In the type I MetAPs these extensions include two zinc finger motifs (Li and Chang, 1995), which are important in yeast for normal MetAP1 function (Zuo et al., 1995) and for functional alignment on the ribosome (Vetro and Chang, 2002). These extensions in the type II MetAPs differ between yeast and mammals. In yeast this includes one polylysine block and mammals two polylysine blocks which flank a polyaspartate block. These polylysine blocks have been proposed to mediate protein-protein or protein-nucleic acid interactions. In mammalian cells it has also been suggested these regions might mediate MetAP2 interaction with eIF2α (Wu et al., 1993).

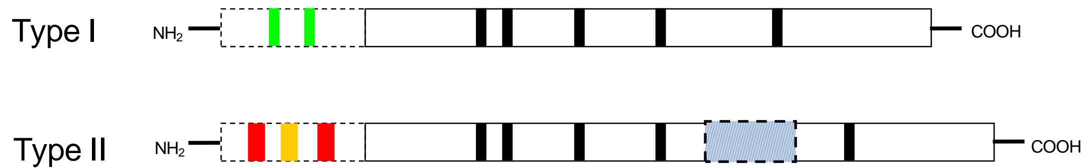


Figure 1.8 Schematic diagram showing features of type I & II MetAPs

Black bars indicate the conserved residues D, D, H, E and E (Q in yeast MetAP1), the blue striped box represents the extra C-terminal insertion of Type II.The N-terminal extensions present only in eukaryotic MetAPs are represented by the dashed boxes, green bars indicate zinc finger motifs, red polylysine blocks and orange polyaspartate block. (Type II is represented as mammalian, yeast does not have a polyaspartate block and only one polylysine block)

Despite some differences in sequence, the crystal structures that have been solved for type I and II MetAPs show great similarity (Figure 1.9). These enzymes have a pseudo 2-fold symmetry which is also common to other aminopeptidases and has led to them be termed the ‘pitta bread’ family of enzymes (Lowther and Matthews, 2000). This suggests that MetAPs structure is functionally important, conserved between organisms and also between enzymes with a similar function.

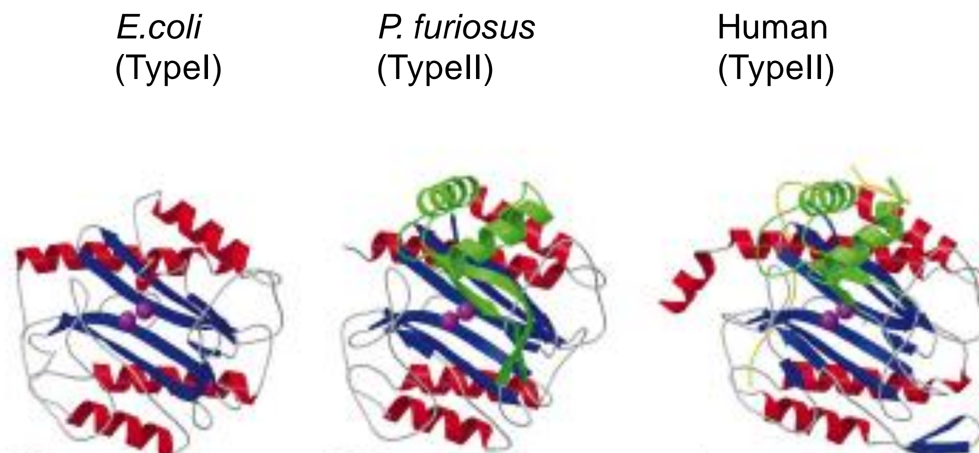


Figure 1.9 Ribbon structures representing MetAPs. Adapted from (Lowther and Matthews, 2000) Ribbon structures of MetAPs termed “pitta bread” due to the 2-fold symmetry in their structures. Red represents α-helices and the Blue β-sheets. Purple sphere indicate the position of metal ions, Green represents C-terminal the insertion present in type II MetAPs and Yellow depicts the N-terminal extension present in the eukaryotic enzymes.

MetAP substrate specificity

Despite the slight structural differences, both types of MetAP (from both eukaryotes and prokaryotes) display activity towards essentially the same substrates. In general both remove an N-terminal methionine from polypeptides only where the adjacent (second) residue is small or uncharged. This has been studied extensively *in vitro* and in *vivo* for both prokaryotes (Hirel et al., 1989, Tsunasawa et al., 1997) and eukaryotes (Tsunasawa et al., 1985, Chang et al., 1990, Kendall and Bradshaw, 1992, Moerschell et al., 1990) and this specificity relative to the second amino acid residue has been summarised in Table 1.1. It can been seen that position 2 residues that allow initiator methionine removal are Gly, Ala, Ser, Cys, Thr, Pro and Val. The conservation of MetAP specificity indicates initiator methionine removal must have an important function in cells.

Table 1.1 N-Met removal form substrates in Prokaryotes and Eukaryotes

	Second Amino Acid (Met-Xaa)																			
	Gly	Ala	Ser	Cys	Thr	Pro	Val	Asp	Asn	Leu	Ile	Gln	Glu	His	Met	Phe	Lys	Tyr	Trp	Arg
Prokaryotes	+	+	+	+*	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
Eukaryotes	+	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-

Compiled from *in vivo* studies
+ indicates N-Met is removed,
- indicates N-Met is retained
* N-Met removal for Met-Cys substrate has not been shown in *Pyrococcus furiosus*

Table 1.2 Physical properties of Amino Acid side chains

	Amino Acid																			
	Gly	Ala	Ser	Cys	Thr	Pro	Val	Asp	Asn	Leu	Ile	Gln	Glu	His	Met	Phe	Lys	Tyr	Trp	Arg
Radius of gyration Å	0	0.77	1.08	1.22	1.24	1.25	1.29	1.43	1.45	1.54	1.56	1.75	1.77	1.78	1.8	1.9	2.08	2.13	2.21	2.38
Max. side chain length Å	0	1.51	2.41	2.83	2.54	2.4	2.55	3.74	3.68	3.9	3.91	4.93	4.97	4.64	5.46	5.1	6.37	6.43	6.64	7.4

It is commonly believed that MetAP specificity is simply due to size of the side chain on the second residue. The initiator methionine only being removed if position 2 has radius of gyration of 1.29 Å or less, anything larger causes steric hindrance between MetAP and any potential substrates. This is evident when comparing the data for initiator methionine removal in Table 1.1 with that for amino acid physical properties in Table 1.2. This is not surprising considering the aforementioned structure of the active site only being able to accommodate up to two amino acids, one being the N-terminal methionine which is of comparatively intermediate size (Table 1.2). As indicated above there are a distinct group of amino acids residues which allow initiator methionine removal when present at position 2. However, there is some evidence that they may not all be equal with respect of the efficiency with which they allow this removal. For example Val and Thr in position 2 lead to a less predictable removal of methionine. This may be due to the size of these residues, both being the larger of the small residues which allow initiator methionine removal (Frottin et al., 2006).

In addition to the specificity conferred by the second amino acid residue, studies have also shown that the amino acid at the third and forth position can influence MetAP activity resulting in inefficient initiator methionine removal for some substrates (Ben-Bassat et al., 1987, Hirel et al., 1989, Moerschell et al., 1990, Frottin et al., 2006). This may depend upon the preceding amino acids e.g. Met-Thr-Pro-Leu, Met-Thr-Pro-Phe and Met-Val-Pro-Leu are not efficiently cleaved, whereas for Met-Ser-Pro-Leu, Met-Cys-Pro-Leu, Met-Pro-Pro-Leu and Met-Val-Pro-Glu the initiator methionine is more successfully removed. A recent study by Xiao *et al* supports these early finding by examining MetAP specificity in greater detail (Xiao et al., 2010). This study revealed that the N-terminal residues present up to position 6 of a sequence can influence removal of the initiator methionine.

The work of Li *et al* mentioned previously, show that *S. cerevisiae* is viable in the absence of either of its two MetAPs. However, is clear from the slow growth phenotypes observed that when lacking one MetAP, the other is not able to fully compensate (Li and Chang, 1995). This suggests that while they appear to have similar substrate specificity *in vitro* they may not have the same functional activity *in*

vivo. Works by Chen *et al*., (2002) also indicate this may be the case. In this study the second residue of glutathione-*S*-transferase is mutated from Ser to either Ala, Cys, Gly, Pro, Thr, Val, Gln, Glu or Leu. These constructs were transformed into wild-type, *Δmap1/MAP2* and *MAP1/Δmap2* strains of *S. cerevisiae* and the degree of initiator methionine removal was determined. These data show that in *MAP1/Δmap2* the removal of the initiator was similar to wild-type for all the substrates. However, in the *Δmap1/MAP2* strain initiator methionine removal was markedly less efficient for all normally cleaved substrates; furthermore overexpression of *MAP2* in this strain did increase overall efficiency but was still inefficient at cleaving methionine in substrates where Cys, Thr or Gly was the second residue. Overall, these findings indicated that MetAP1 is primarily responsible for initiator removal in *S. cerevisiae*.

It is evident from work presented by Chen *et al.* that the two yeast MetAPs differ with regard to their cleavage activity *in vivo*. A major factor of this difference may be the structure of the active sites which as previously mentioned is slightly smaller in type I molecules and may have different amino acid side chains exposed in a comparison of *E. coli* MetAP1 and human MetAP2 (Liu et al., 1998).

In addition to initiator methionine removal, it is believed that both type I and type II MetAPs may have other role in the cell. In yeast MetAP1 has a role in methionine metabolism/ salvage pathways, preventing activation of *MET* genes and also having a synthetic interaction with *MET4*. A *Δmap1 Δmet4* strain exhibits a severe synthetic growth phenotype (Dummitt et al., 2003). Furthermore, in *S. cerevisiae Δmap1* cells, growth is inhibited by excess methionine in the culture medium, an effect caused by product inhibition of MetAP2 rendering it unable to compensate for the loss of MetAP1 (Dummitt et al., 2003).

Taking into account the differences in the yeast MetAP activities and their behaviours it is conceivable that MetAP2 may have another role in yeast. This is certainly true for mammalian cells in which MetAP2 (rat p67) has been shown to associate with the eIF-2α subunit preventing phosphorylation by eIF-2 kinase, and therefore promoting protein synthesis/ cell proliferation even in the presence of eIF-2 kinase (Ray et al., 1992). MetAP2 expression may therefore have significant

effects on fundamental processes independent of its aminopeptidase activity. MetAP2 has also been shown to be increased in tumour cells (Tucker et al., 2008, Wu et al., 1993) implicating it in progression to cancer. In all, this indicates that MetAPs may have multiple roles in the cell, and together with their selective inhibition this makes them feasible targets for anti-bacterial /fungal and cancer agents.

The function of removing the N-terminal methionine of a protein alone is not well understood but in some instances it is known to be required for other N-terminal modifications of proteins, some of which are discussed below.

1.3.2 N-α-Acetylation

N-terminal acetylation is a co-translational event where specific N-terminal residues of elongating polypeptides accept an acetyl group from the donor substrate acetyl Co-enzyme A (acetyl-CoA). This process involves one of four N-acetyl transferase complexes (NATs) to catalyze this reaction, each NAT complex acting on a specific set of protein N-termini.

The frequency of N-α-Acetylation in yeast is generally accepted as occurring for 40-50% of yeast proteins (Lee et al., 1989, Polevoda and Sherman, 2003b). In contrast, N-α-Acetylation is a rare event in prokaryotes and with many modifications affecting only a few ribosomal subunits, this is likely to be due to a functional requirement for acetylation of these proteins. This rarity of N-α-Acetylation is further highlighted in *E.coli* in which 807 out of 810 verified N-terminal sequences were not acetylated (Waller, 1963, Polevoda and Sherman, 2003b). For mammalian proteins the extent of N-α-Acetylation is less clear. Some time ago it was estimated that 80-90% of cytosolic proteins are N-terminally acetylated in mammals (Driessen et al., 1985, Persson et al., 1985). Recent studies by Polevoda *et al* (Polevoda and Sherman, 2003b) and another by Meinnel *et al* (Meinnel et al., 2005) aimed to accurately evaluate the authenticity of such a high proportion of N-α-acetylated proteins. Both studies compiled data-sets of experimentally determined N-termini sequences but used slightly different approaches, and obtained substantially different findings.

Polevoda's figure was around 90% and Meinnel's around 30%. The major differences between the studies was that Polevoda's study (Polevoda and Sherman, 2003b) analysed experimentally determined N-termini of mammalian (bovine, mouse, human) proteins (discarding orthologs) and then calculated the percentage of these that were determined to be N-α-acetylated. In contrast Meinnel's study (Meinnel et al., 2005) takes experimentally determined N-α-acetylated proteins and uses these sequences to look for similar sequences and estimate N-α-acetylated proteins in the open reading frames of the human genome. Polevoda's figure may be inaccurate due to the data set used having a strong bias for N-α-acetylated proteins, Whereas Meinnel's approach is predictive and does not use such a restricted data set. As both studies are predictive and potentially have an inherent skew due to the data sets used the real value for N-α-acetylated in mammals is likely to be between 30% and 90% and therefore closer to the recent prediction in yeast (Polevoda and Sherman, 2003b). Therefore N-α-acetylation is a fairly common event in both yeast and mammals.

N-α-Acetyl transferase complexes (NATs) and substrate specificity

To date four N-α-Acetyl transferase complexes have been identified in the cytosol of eukaryotes Nat A, B, C and D (Polevoda and Sherman, 2003a, Polevoda et al., 1999, Arnold et al., 1999, Song et al., 2003). Each complex has specificity for different N-terminal sequences with the major determinant of this specificity being the amino acid residues at position 1 and 2 of the primary sequence. Each of the NAT complexes are composed of different subunits detailed in Table 1.3 (Polevoda et al., 2009a, Polevoda and Sherman, 2003b)

Table 1.3 Subunits of the NAT complexes

	NAT complex			
	NatA	NatB	NatC	NatD
Catalytic subunit	Ard1p	Nat3p	Mak3p	Nat4p
Auxiliary subunit	Nat1p ?	Mdm20p ?	Mak10p Mak31p	?

? indicates other subunits may be in the complex

The NatA complex acetylates N-terminal residues exposed following N-terminal methionine removal by MetAPs. It is generally accepted that these exposed residues include serine, alanine, glycine, and threonine (Polevoda et al., 1999, Tsunasawa et al., 1985, Arnesen et al., 2009). There is also some evidence that valine and cysteine can be acetylated although these events are rarer. In some part this may due to the lower frequency of these residues appearing N-terminally, cysteine in particular being rare at position 2. Also these residues rely on MetAP activity for their exposure and as mentioned previously valine is a residue which may not always efficiently promote initiator methionine removal (Polevoda and Sherman, 2003b, Martinez et al., 2008). In contrast to NatA, the NatB and NatC complexes acetylate the initiator methionine itself, with the difference in specificity between these complexes incurred predominantly by the residue at position 2. NatB preferentially acetylates the initiator methionine when the position 2 residue is either Glu, Asp, Asn and potentially Met. NatC acetylates the initiator methionine when Ile, Leu, Trp, or Phe is a residue at position 2 (Polevoda et al., 1999). The NatD complex is unusual as it requires a sequence of 23-51 residues to efficiently acetylate a substrate. To date only two substrates have been identified which consist of the histones H2A and H4 which have the N-terminal sequences of Ser-Gly-Gly-Lys-Gly and Ser-Gly-Arg-Gly-Lys respectively (Polevoda et al., 2009b, Song et al., 2003).

Most of the knowledge which has been gained regarding specificity of NAT complexes has been deciphered from genetic studies in *S. cerevisiae*. These studies used deletions of the genes encoding the various subunits of the NAT complexes in particular the catalytic subunits. Substrates with various N-termini were then assayed

to determine the degree of N-α-acetylation in these deletion strains (Polevoda et al., 2003, Polevoda et al., 2009b, Polevoda et al., 1999, Polevoda and Sherman, 2001).

The NAT complexes have been widely shown to act on specific N-termini of substrates. However, the subsequent sequences of these substrates are degenerate. It has been suggested that residues other than those which occupy positions 1 and 2 in a proteins N-terminal sequence may also contribute towards determining whether the protein is a substrate for a specific NAT complex. These residues may be required in addition to those discussed above for efficient acetylation of an N-terminus. In contrast, suboptimal residues in the N-terminal sequence may diminish or inhibit acetylation of the substrate (Moerschell et al., 1990, Polevoda and Sherman, 2000). Examples of residues which diminish the acetylation of a sequence include proline in the sequence Met-Asp-Pro-Leu-Ala in which acetylation is reduced by 33% compared to the sequence Met-Asp-Phe-Leu-Ala and also the presence of lysine in the sequence Met-Glu-Phe-Lys-Ala which is 55 % acetylated compare to the sequence Met-Glu-Phe-Leu-Ala which is 100% acetylated (Moerschell et al., 1990). Not all of the amino acid residues which may be required or can inhibit N-α-acetylation are known. It is also not clear at which position in a sequence these required/inhibitory residues may be present to have their effect on acetylation. For this reason definitive prediction of N-α-acetylation from amino acid sequence is difficult to achieve with complete confidence. Consequently recent studies have focused on predicting N-α-acetylation by comparing sequence with those that have been experimentally determined as being N-α-acetylated (Martinez et al., 2008, Meinnel et al., 2005, Polevoda and Sherman, 2003b).

Despite being predicted to occur extensively the biological significance of N-α-acetylation is not well understood. It seems that the role of this modification varies for particular proteins. It has been suggested that N-α-acetylation may affect protein stability (Arfin and Bradshaw, 1988) and has recently been shown to function as a signal for degradation of certain cytosolic proteins via the N-end rule, discussed further in section 1.3.4 (Hwang et al., 2010). In some cases N-α-acetylation has been shown to be important for protein functionality, one example of this is the N-α-acetylation of the peroxidase Peroxiredoxin II in which N-terminal modification is

suggested to protect this protein from irreversible hyperoxidation by inducing a structural change (Seo et al., 2009).

The NAT complexes discussed above refer to those which are encoded by the nuclear genome and reside in the cytoplasm. To date there is no definitive evidence to suggest that there are NAT complexes which reside in organelles were protein can be synthesised *de novo* like there is for the MetAPs. However, there are three photosystem II proteins synthesised in the chloroplasts of spinach which are believed to be N-α-acetylated and phoshorylated on the N-terminal threonine (Michel et al., 1988). How this acetylation occurs remains to be determined.

1.3.3 N-Myristoylation

Of the N-terminal modifications discussed here, N-myristoylation generally occurs with the lowest frequency. In mammals and yeast it is estimated to affect 0.5% of total proteins (Maurer-Stroh et al., 2002a). N-Myristoylation is a co-translational modification by which myristic acid, a C14 saturated fatty acid, is transferred to an N-terminal glycine of a polypeptide. This process is catalyzed by the enzyme N-Myristoyltransferase (NMT) (Rajala et al., 2000, Boutin, 1997).

N-Myristoylated substrate proteins have a glycine in position 2 of their primary sequence. For this glycine to become the N-terminal residue which is myristoylated, the initiator methionine must be removed by MetAPs. An N-terminal glycine is not the only determinant for N-Myristoylation. Maurer-Stroh's extensive analysis of NMT substrates found the 15 residues following the glycine are also potentially important (Maurer-Stroh et al., 2002b). These studies predicted three distinct regions that might influence NMT interaction;

- Residues 1-6 fit the binding pocket of NMT. Glycine is absolutely required at position 1, small residues are favoured at positions 2 and 5 (with a bias for polarity at 5) while position 4 tends to be hydrophobic and 6 is biased towards lysine.

- Residues 7-10 interact with the surface NMT, close to the catalytic cavity. Hydrophobic residues are disfavoured in this region, particularly at positions 7, 9 and 10, but can be tolerated if compensated for by other proximal residues. A bias for basic residues is observed at position 8 and a flexible side chain at 9, while positions 10 and 11 are frequently small and/or polar residues.
- The residues 11-17 form a hydrophilic linker region with a tendency for inclusion of polar residues. (Maurer-Stroh et al., 2002b)

One well characterised function of N-Myristoylation is to assist targeting of proteins to membranes. An example of this is NADH-cytochrome b(5) reductase, a protein that is differentially targeted to both the ER and also the outer mitochondrial membranes. The N-myristoylation of NADH-cytochrome b(5) reductase is required only for mitochondrial targeting , as the non myristoylated protein only targets to the ER. It is believed that N-myristoylation interferes with ER targeting (Colombo et al., 2005).

N-Myristoylation also has a role in virus protein maturation, such as the structural *gag* proteins of HIV retrovirus and also in oncogenesis. The N-Myristoylation of the proto-oncogene *c-src* is essential for its cell transforming capability and is required for this proteins translocation to the plasma membrane where it becomes activated. For these reasons NMT has been suggested as a target for antiviral and anticancer applications (Boutin, 1997).

It is evident from the N-terminal modifications discussed above that these processes can lead to a variety of N-termini with vastly different properties on newly synthesised proteins. While these modifications depend primarily upon the amino acid sequence of the polypeptide both N-Myristoylation and some N-Acetylation events are also dependent upon initiator methionine removal by MetAPs. These events and the possible N-termini which they can lead to are summarised in Figure 1.10.

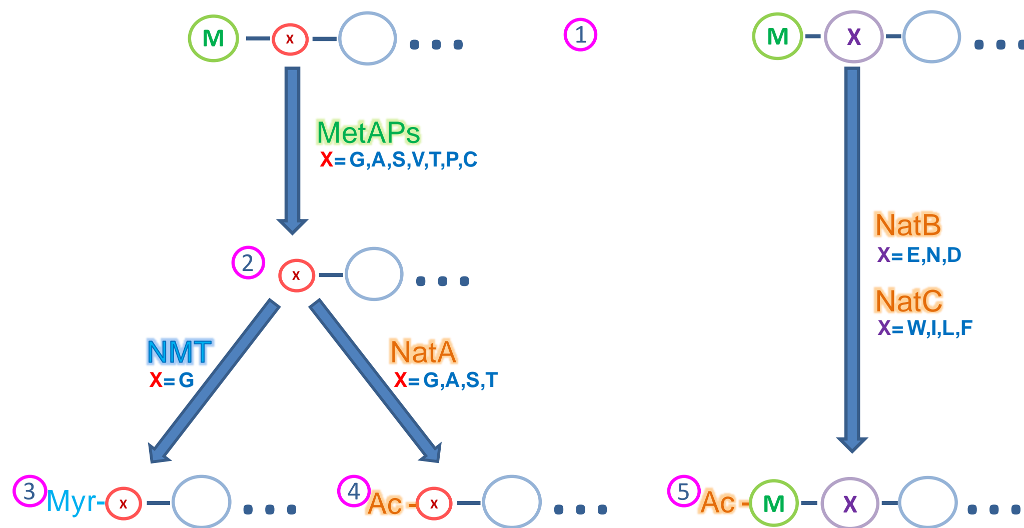


Figure 1.10 Summary of N-terminal modification. A schematic diagram of the N-terminal modification events discussed above and the N-termini they give rise to. M being the initiator methionine and X the amino acid residue at P2. The different N-termini which can arise are indicated by pink circles; 1) unmodified initiator methionine remains in place, 2) Methionine removed by MetAPs and P2 residue remains unmodified 3) If methionine is removed and Glycine is revealed as N-terminus it can be N-myristoylated 4) If methionine is removed by MetAPs and glycine, alanine, serine or threonine are revealed they can be acetylated by NatA 5) If the initiator methionine remains intact it can be Acetylated by NatB or NatC depending on the residue at P2

1.3.4 The N-end Rule Hypothesis

It is evident that the N-terminal modification events may affect a vast proportion of the proteins in the cytosol. For a majority of these modification events it is largely unknown what the biological function is, and those proteins where the roles of the modification have been identified are on a case by case basis. However, a theory termed the N-end rule hypothesis attempts to attribute a global role for the N-terminus of proteins. This hypothesis states that the *in vivo* half life of a protein relates to the residue at the N-terminus (Bachmair et al., 1986, Gonda et al., 1989). The N-terminal amino acids being categorised as either stabilising residues or destabilising residues the latter acting as a signal for degradation. A recent study also encompasses N-α-acetylated residues as being destabilising (Hwang et al., 2010). The main features of the N-end rule hypothesis are outlined in Figure 1.11. Although this hypothesis could potentially affect a vast number of proteins, only a hand full have identified *in vivo*.

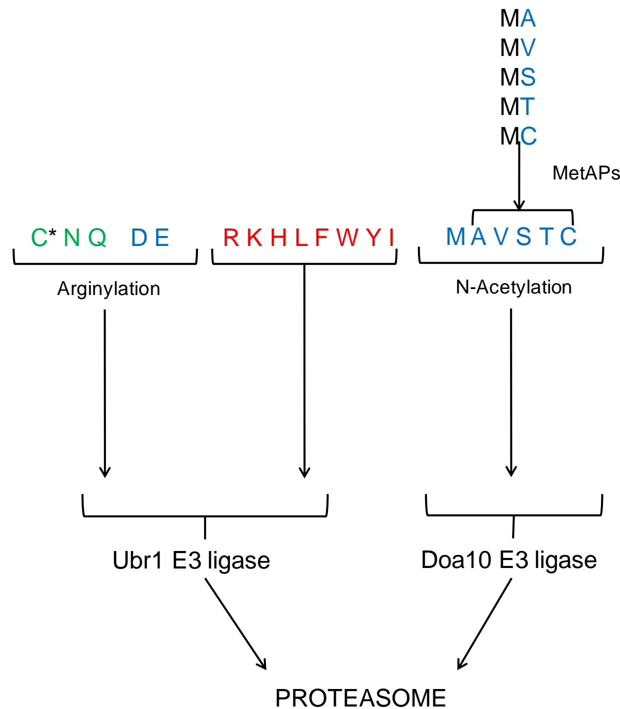


Figure 1.11 Schematic diagram of the N-end rule pathway. Adapted from Hwang *et al*., (2010). This diagram illustrates the destabilising N-terminal amino acids of the N-end rule, primary, secondary and tertiary. tertiary residues are deaminated to aspartate (D) and glutamate (E), these secondary destabilising residues can then be arginylated by arginyl-transferase (Kwon et al., 2000, Kwon et al., 2002) to become a primary destabilising residue. primary destabilising residues can be targeted to the proteasome for degradation via Ubr1 ubiquitin ligase (Meinnel et al., 2006, Mogk et al., 2007). other secondary destabilising residues are N-Acetylated to become primary destabilising residues targeted these acetylated via an alternate ubiquitin ligase Doa10 to the proteasome for degradation (Hwang et al., 2010).

1.4 Coordinated binding of protein biogenesis factors at the ribosome

A number of the processes and modifications discussed so far which occur during protein biogenesis are co-translational, taking place as the nascent protein is emerging from the exit site of a translating ribosome. These events include N-terminal methionine removal, N-α-acetylation and SRP-mediated targeting to the ER. It has been shown that a number of these biogenesis factors associate with the various parts of ribosome close to the site where the nascent protein emerges (reviewed in Kramer et al., 2009). These include MetAP1 (Vetro and Chang, 2002) , MetAP2 (Raue et al., 2007) , SRP (Hann and Walter, 1991), NatA (Gautschi et al., 2003), NatB and C (Polevoda et al., 2008) and also NatD (Polevoda et al., 2009b).

There are also a number of chaperones and chaperone complexes; nascent polypeptide associated complex (NAC), the Hsp70s Ssb1/2p, ribosome-associated complex (RAC) consisting of the Hsp40 zuotin and the Hsp70 Ssz1p (Rospert et al., 2002, Yan et al., 1998, Nelson et al., 1992, Gautschi et al., 2001) which can also associate in the same region. It is unlikely that all of these factors can associate with the ribosome and the nascent protein simultaneously. A study by Raue *et al.* (Raue et al., 2007) aimed to reveal how some of these factors coordinate their association. This study showed that MetAP1/2, SRP and NatA associate preferentially with translating ribosomes regardless of the polypeptide being translated. However, additional experiments also suggest that if the emerging nascent protein is a substrate for SRP, Dap2p in this case, increasing amounts of SRP molecules are recruited to the ribosome. Interestingly this study further showed that although SRP does not affect association of NatA to ribosomes translating Dap2p, cross-linking experiments suggest that NatA was unable to interact with this SRP substrate protein. This suggests some ribosome associated factors can occupy differing position depending on the emerging N-terminus of the nascent protein (Raue et al., 2007). This is interesting as work by Gautschi *et al.* have shown that NatA can simultaneously bind to ribosome in conjunction with NAC and Ssb1/2p (Gautschi et al., 2003). This seems possible as NAC and Ssb1/2p associate with emerging nascent proteins at a length of 45-50 amino acids and NatA interacting with longer nascent proteins regardless of whether it is a substrate of NatA (Gautschi et al., 2003). SRP interacts specifically with secretory proteins of 50-70 amino acids in length as they emerge from the ribosome (Kurzchalia et al., 1986). Therefore it may be that NatA also interacts with proteins emerging from the ribosome around this length but somehow SRP stericly hinders NatA’s interaction.

1.5 MPhil hypothesis and aims

It is clear from the topics discussed so far that the N-terminal amino acid sequence of a protein can significantly influence the proteins final fate. This may involve a relatively long N-terminal sequence which targets a protein to a specific sub-cellular location. Alternatively it could be that the presence of a single residue at the second position leads to specific chemical modification of this part of the protein. Investigations of N-terminal modification have focused upon cytosolic proteins. Presumably this is because targeted proteins often contain signal sequences which may be removed prior to the protein attaining its functional status. Any such modification would be assumed to be transient and unlikely to affect function of the mature protein. Alternatively, there is the possibility that N-terminal modification might actually compromise targeting of proteins changing the properties of signal sequences and/or masking recognition by targeting factors. If this were true, cells may possess mechanisms to prevent N-terminal modification of targeted proteins. Co-ordination of modifying and targeting factors at the ribosome exit site (as described above) would be one process with potential for control of such events.

Interestingly, recently generated data (Forte *et al*. manuscript submitted) suggests that eukaryotic cells may actually not simply limit exposure of targeted proteins to modifying factors. Bioinformatics analyses suggest that yeast has a clear bias against the inclusion of amino acid residues that may lead to N-terminal modification of signal peptides. Initial analyses focused upon the frequency with which each amino acid appeared at position 2 for a large set of yeast cytosolic and secretory proteins. As detailed previously, position 2 is a major determinant for a number N-terminal modifications, especially MetAP activity which may itself be a prelude to further modification. Intriguingly the data showed some clear difference between the two groups of proteins, alanine and serine appear at high frequency in the cytosolic group whereas in the secretory group leucine and lysine are relatively frequent (Figure 1.12a). These and other differences in amino acid frequency for the two groups are more apparent when the frequency of amino acids observed at position 2 for the two groups of proteins are expressed as a ratio (Figure 1.12b). These data indicate the cytosolic group of proteins have a bias towards small and acidic amino acids at

position 2, while proteins with signal sequences have a bias towards large and basic residues in this position. Proteins with small residues at position 2 are favoured as substrates of MetAPs. Therefore when these data are expressed in terms of whether cytosolic or signal sequence proteins are predicted to be substrates for initiator methionine removal the cytosolic group show a strong bias (~70%) for N-terminal methionine removal to occur (Figure 1.12c) a good agreement with data from other studies (Giglione et al., 2000, Meinnel et al., 1993, Serero et al., 2003). In contrast in the group of signal sequence proteins only ~20% would be predicted to be MetAP substrates. This reveals a strong bias against removal of the initiator methionine for signal sequences. This bias against initiator methionine removal in signal sequences has been noted in a study by Flinta et al. (Flinta et al., 1986). The conclusion to this study suggests the reason for this is a factor of the cytosol not present in the environment of secretory proteins.

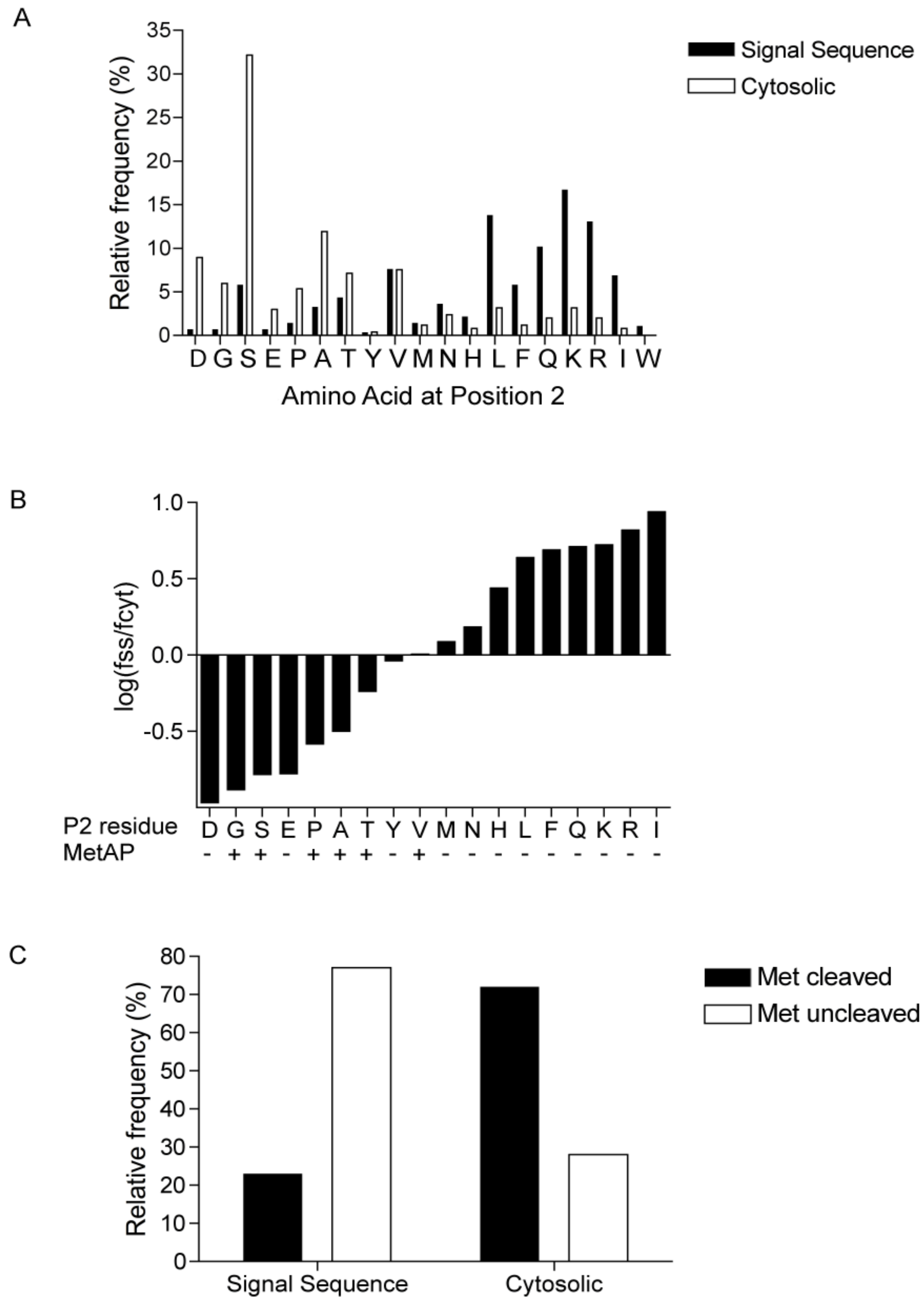


Figure 1.12 Comparison of amino acid frequency at position 2 of proteins with signal sequences and cytosolic proteins. (provided by Dr M Pool) **A.** Relative frequency of amino acids at P2 of a filtered set of 275 predicted signal sequence-containing proteins from *S. cerevisiae* was compared to a similar size group (n=252) of randomly selected cytosolic proteins. Frequency distribution between the groups differed significantly (P<0.0001, χ^2=207.3 18 d.f.). **B.** Ratio of relative frequency of amino acids at P2 between signal sequence (f_{ss}) and cytosolic (f_{cyt}) proteins. Data in panels A and B are ordered according to increasing value of f_{ss}/f_{cyt}. **C.** Predicted methionine cleavage of signal sequence and cytosolic N-termini based on relative P2 frequency. For complete data sets see appendix

Finally; all this suggests that there is a clear bias against the inclusion of position 2 residues which may promote N-terminal modification. The aim of this MPhil study is to investigate if the observed bias against removal of the initiator methionine of signal sequences has biological significance. This will involve the following approaches;

- This study will use the model organism *Saccharomyces cerevisiae* as the previous studies of both N-terminal modification and ER translocation which forms the basis of study have been extensively investigated in this organism. Also in *S. cerevisiae* is relatively easy to perform the genetic manipulation required in this study.

- In order to investigate the affects of N-terminal modification on secretory proteins with signal sequences, the amino acid residue at position 2 will be altered to residues which may promote the occurrence of modifications by MetAPs and NATs.

- As the signal sequence of secretory proteins is required for targeting to the ER it is important to explore if targeting of these proteins is affect by the alteration of the position 2 residue. This can be investigated by looking for the presence of precursor protein in cells.

- A comparison will be made of how these alterations of signal sequences affects proteins targeted via either the SRP-independent or SRP-dependent pathway. Studies by Raue *et al.* (Raue et al., 2007) have shown that NatA cannot interact with an SRP dependant secretory substrate. In yeast secretory substrates can target independently of SRP, therefore the NATs may be able to act on these substrates.

- The nature of any effects seen will be investigated using genetic manipulations in *Saccharomyces cerevisiae* and biochemical techniques.

Materials and Methods

2. Materials and Methods

2.0 Reagents and Chemicals

DNA restriction enzymes were from Roche diagnostic and New England Biolabs. *In vitro* transcription and rabbit reticulocyte lysate reagents were from Promega. Media components were from Becton Dickinson & Co. All radio-chemicals were from NEN Perkin Elmer. All other reagents were of analytical grade from Sigma-Aldrich, Fisher scientific or Melford laboratories except where stated. Antiserum raised against CPY (residues 155–525) and α-factor (residues 10–120) have been described previously (Young et al., 2001, Tyson and Stirling, 2000).

Table 2.1 *Escherichia coli* strains used in this study

Name	*Genotype*	Source
DH5α	F^- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17*(r_K^- m_K^+), *λ–*	(Hanahan, 1983)
SCS110	*rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) [F´ traD36 proAB lacIqZΔM15].*	Stratagene

Table 2.2 *S. cerevisiae* strains used in this study

Name	***Genotype***	**Source**
MWY63	*MATa***,** *prc1::Kan*MX, *leu2, his3, ura3, ade2, pep4-3, sec61-3*	(Willer et al., 2008)
Δprc1α	*MATα, prc1::KanMX4, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*	(Winzeler et al., 1999)
Δpep4a	*MATa, pep4::KanMX4, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0*	(Winzeler et al., 1999)
GFY3	*MATa, prc1::KanMX4, pep4::KanMX4, his3Δ, leu2Δ, ura3Δ0*	This study
GFY7	*MATa, prc1::KanMX4, pep4::KanMX4, map1::HIS3MX6, his3Δ1, leu2Δ0, ura3Δ0*	This study
GFY11	*MATα, prc1::KanMX4, ard1::hphMX3, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*	This study
GFY12	*MATα, prc1::KanMX4, nat3::hphMX3, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0*	This study

2.1 *S. cerevisiae* strain construction

The yeast strains in this study are listed in Table 2.2.

GFY3 was constructed by mating the Δpep4a and Δprc1α strains, the resulting diploid cells were sporulated and the tetrads dissected. Tetrads were selected which had three G418 resistant spores. The spores were scored for null mutations; PCRs from genomic DNA was used to assess the size of the *PEP4* locus (primers "pep4 for A" & "pep4 rev" A), diminished CPY was assessed by western blotting yeast total protein extracts.

GFY7 was made by disruption of the *MAP1* locus with the product of a PCR amplification of PFA6a-His3MX6 module (Longtine et al., 1998). PCR product was prepared with primers Map1 KO F and Map1 KO R (Table 2.3) and was used to transform GFY3 strain conferring ability to grow without the requirement of histidine. Disruption was assessed by sensitivity to the drug fumagillin.

GFY11 and GFY12 were made by disruption or the *ARD1* and *NAT3* loci with the product of PCR amplification of the HphMX3 module from pAG26 (Goldstein and McCusker, 1999), primers used were either Ard1/A KO F2 and Ard1/A KO R or Nat3/B KO F2 and Nat3/B KO R (Table 2.3) respectively. PCR products were then used to transform Δprc1α strain and confer Hygromycin B resistance. Disruption was assessed by diagnostic PCR using primers Hph1 int-F and Ard1 ex-R or Nat3 ex-R.

Table 2.3 DNA primers used in this study

No.	Name	Sequence	Description
1	Map1 KO F	TCCTAGCAAGAAAAAAATAAGCAAAAAAATTGTATAATC GGATCCCCGGGTTAATTAA	HIS3MX6disruption of MAP1
2	Map1 KO R	GTACAAGTTCAAGTTTTTATTGGTTTCTATATGTATAAG ATAAGAATTCGAGCTCGTTTAAAC	HIS3MX6 disruption of MAP1
3	Ard1/A KO F2	AAATACATACGATCAAGCTCCAAAATAAACTTCGTCAAC CCGGATCCCCGGGTTAATTAA	HphMX3 disruption of ARD1
4	Ard1/A KO R	CCTGGATGAAAATATACTACGTTTATATAGGTTGATTTAA AATACGACTCACTATAGGGAG	HphMX3 disruption of ARD1
5	Nat3/B KO F2	ATTGAGAATATTCAAGGAAAGAGACAGGAGGATTCGAG AACGGATCCCCGGGTTAATTAA	HphMX3 disruption of NAT3
6	Nat3/B KO R	ATTATTATGTTCTGAGTATGAGGACGAGGTAATACATACC AATACGACTCACTATAGGGAG	HphMX3 disruption NAT3
7	CPY-A-F	CAACTTAAAGTATACATACGCTATGGCTAAAGCATTCAC CAGTTTACTATG	SDM of pMW346
8	CPY-A-R	CATAGTAAACTGGTGAATGCTTTAGCCATAGCGTATGTAT ACTTTAAGTTG	"
9	CPY-C-F	CAACTTAAAGTATACATACGCTATGTGTAAAGCATTCACC AGTTTACTATG	"
10	CPY-C-R	CATAGTAAACTGGTGAATGCTTTACACATAGCGTATGTAT ACTTTAAGTTG	"
11	CPY-E-F	CAACTTAAAGTATACATACGCTATGGAAAAAGCATTCAC CAGTTTACTATG	"
12	CPY-E-R	CATAGTAAACTGGTGAATGCTTTTTCCATAGCGTATGTAT ACTTTAAGTTG	"
13	CPY-G-F	CAACTTAAAGTATACATACGCTATGGGTAAAGCATTCAC CAGTTTACTATG	"
14	CPY-G-R	CATAGTAAACTGGTGAATGCTTTACCCATAGCGTATGTAT ACTTTAAGTTG	"
15	CPY-R-F	CAACTTAAAGTATACATACGCTATGAGAAAAGCATTCAC CAGTTTACTATG	"
16	CPY-R-R	CATAGTAAACTGGTGAATGCTTTTCTCATAGCGTATGTAT ACTTTAAGTTG	"
17	CPY-S-F	CAACTTAAAGTATACATACGCTATGTCCAAAGCATTCACC AGTTTACTATG	"
18	CPY-S-R	CATAGTAAACTGGTGAATGCTTTGGACATAGCGTATGTAT ACTTTAAGTTG	"
19	CPY-V-F	CAACTTAAAGTATACATACGCTATGGTCAAAGCATTCAC CAGTTTACTATG	"
20	CPY-V-R	CATAGTAAACTGGTGAATGCTTTGACCATAGCGTATGTAT ACTTTAAGTTG	"
21	OPY-A-F	GGTGCTGAAAAAATGGCTAGGCAGGTTTGGTTC	SDM of pOPY
22	OPY-A-R	GAACCAAACCTGCCTAGCCATTTTTTCAGCACC	"
23	OPY-C-F	GGTGCTGAAAAAATGTGTAGGCAGGTTTGGTTC	"
24	OPY-C-R	GAACCAAACCTGCCTACACATTTTTTCAGCACC	"
25	OPY-E-F	GGTGCTGAAAAAATGGAAAGGCAGGTTTGGTTC	"
26	OPY-E-R	GAACCAAACCTGCCTTTCCATTTTTTCAGCACC	"
27	OPY-G-F	GGTGCTGAAAAAATGGGTAGGCAGGTTTGGTTC	"

28	OPY-G-R	GAACCAAACCTGCCTACCCATTTTTTCAGCACC	"
29	OPY-S-F	TTGGTGCTGAAAAAATGTCTAGGCAGGTTTGGTTCTCTTG G	"
30	OPY-S-R	CCAAGAGAACCAAACCTGCCTAGACATTTTTTCAGCACC AA	"
31	ppαF wt sp6 F	ATTTAGGTGACACTATAGACGATTAAAAGAATGAGATTT C	MR mRNA template
32	ppαF S sp6 F	ATTTAGGTGACACTATAGACGATTAAAAGAATGTCCAGA TTTCCTTCAAT	MSR mRNA template
33	pEH3 R	GCCTGCAGGTCGACTCTAGAGTCG	pEH3 mRNA template
34	ppalpha k5 R	GCATGCCTGCAGGTCGACTTTGTTAC	pGF22 mRNA template
35	Ost1ss EcoRI-F	GCAAGCTGAATTCCTTCTTTGACAAGTACCCGATTGC	PCR Ost1p signal sequence
36	Ost1ssHincII-R	GGACAGGATGTTGACCGGCTCGTATTGGGCAGCAG	PCR Ost1p signal sequence
37	T7 promoter	TAATACGACTCACTATAGGG	ppOαF 92aa mRNA
38	OαF metHA-R	AGCGTAATCTGGAACATCGTATGGGTACATTCTAGCAGC AATGCTGGCAATAG	ppOαF 92aa mRNA
39	Ard1 ex-R	CCTATCTGCCGGATATTAC	Diagnostic PCR *ard1* disruption
40	Nat3 ex-R	CAACCTCCTTCCAACGTTG	Diagnostic PCR *nat3* disruption
41	Hph1 int-F	GGATTCCCAATACGAGGTCGCC	Diagnostic PCR HphMX3 insert
42	Pep4 -F (A)	CGTCTTATGCCTTCCGGG	External to *PEP4* locus
43	Pep4 -R (A)	CTGTGAATCTCTGCAGCC	External to *PEP4* locus
44	Map1-F	GGGAGAACTGCTGCCCATGGC	External to *MAP1* locus
45	Map1-R	CACAATGGGCAGAAGGACAAG	External to *MAP1* locus

F= forward primer complementary to non-coding strand.

R= reverse primer complementary to the coding strand.

Table 2.4. Plasmids used in this study

Name	Description	Source
PFA6a-His3MX6	*his5+* gene from *S.pombe* used as a gene deletion module	(Longtine et al., 1998)
pAG26	*hph* gene from *Klebsiella pneumoniae* encoding hygromycin B phosphotransferase used as a gene deletion module	(Goldstein and McCusker, 1999)
pMW342	ppOPY* LEU2 PRS315	(Willer et al., 2008)
pMW346	ppCPY LEU2 PRS315	(Willer et al., 2008)
pGEM3-CPY	ppCPY in pGEM3	This study
pGEM3-OPY	ppOPY in pGEM3	This study
pCPY-A	SDM of pMW346 alanine inserted at position 2 of signal sequence	This study
pCPY-C	SDM of pMW346 cysteine inserted at position 2 of signal sequence	This study
pCPY-E	SDM of pMW346 glutamate inserted at position 2 of signal sequence	This study
pCPY-G	SDM of pMW346 glycine inserted at position 2 of signal sequence	This study
pCPY-H	SDM of pMW346 histidine inserted at position 2 of signal sequence	This study
pCPY-Q	SDM of pMW346 glutamine inserted at position 2 of signal sequence	This study
pCPY-R	SDM of pMW346 arginine inserted at position 2 of signal sequence	This study
pCPY-S	SDM of pMW346 serine inserted at position 2 of signal sequence	This study
pCPY-V	SDM of pMW346 valine inserted at position 2 of signal sequence	This study
pOPY	pMW346 modified so that Ost1p signal sequence replaced that of CPY	This study
pOPY-A	SDM of pOPY Alanine inserted at position 2 of signal sequence	This study
pOPY-C	SDM of pOPY Cysteine inserted at position 2 of signal sequence	This study
pOPY-E	SDM of pOPY Glutamate inserted at position 2 of signal sequence	This study
pOPY-G	SDM of pOPY Glycine inserted at position 2 of signal sequence	This study
pOPY-S	SDM of pOPY Serine inserted at position 2 of signal sequence	This study
pA11-k5	K5 ppαF in pAlter, all lysines codons in *wild-type* ppαF altered to arginine and single lysine introduced at position 5 of signal sequence	(Plath et al., 1998)
pEH3	pGEM3z-ppαF	(Steel et al., 2002)
pGF22	pGEM3z-ppαF with all lysines codons in *wild-type* ppαF altered to arginine	This study
pGF23	pGF22 cut with SalI klenow filled and re-ligated to remove one HincII site	This study
pGF24	O-αFactor signal sequence of ppαF (pGF23) replaced with that of Ost1p	This study
pGF25	As pGF24 except contains serine codon insertion at position 2 of Ost1p signal sequence	This study

2.2 Plasmid constructions

Plasmid used in this study are listed in table 2.4.
SCS110 strain of E. coli was used for cloning to produce un-methylated plasmid DNA which can then be cut with certain restriction enzymes (SexAI).
Construction of pOPY and pGEM3-OPY; pMW342 was digested with SacI/SexAI and the 1185 kb fragment containing Ost1p signal sequence was ligated to the SacI/SexAI backbone fragment of pMW346 (7837kb) or pGEM3-CPY(4707kb) respectively to replace the CPY signal sequence

pGEM3-CPY was made by restriction digest of pMW346 with EcoRI/HindIII and ligation of CPY/OPY containing fragments in the EcoRI/HindIII site of pGEM3.

The constructs which express ppCPY and ppOPY with position 2 insertion mutations of the signal sequence (Table 2.4) were made via site directed mutagenesis (SDM) of pGEM3-CPY and pGEM3-OPY using the respective pairs of SDM primers (Table 2.3). Mutated constructs were checked for the appropriate mutation by automated DNA sequencing. For the pCPY constructs SacII /Bsu36I fragment containing the mutation were then ligated into pMW346 to replace the *wild-type* region, resulting in ppCPY position 2 insertion mutant in a yeast expression vector. For the pOPY constructs SacI/SexAI fragment containing the mutation were then ligated into pOPY to replace the *wild-type* Ost1p signal sequence, resulting in ppOPY position 2 insertion mutant in a yeast expression vector.

pGF22, the PsiI/Sph1 fragment of pA11-k5 was cloned into pEH3 to replace this portion of *wild-type* ppαF and thus making a lysine free ppαF.
pGF24 and pGF25 were constructed by PCR (primers table 2.3) of the Ost1p signal sequence from pOPY and pOPY-S respectively. The EcoRI/HincII fragment generated by digestion of these PCRs were cloned into pGF23 (Table 2.4) to replace the ppαF signal sequence with that of Ost1p or the serine mutant version, resulting in constructs encoding ppOαF and ppOαF-S respectively.

2.3 *Escherichia coli* techniques

2.3.1 Growth and maintenance of *Escherichia coli*

E. coli cells were grown in Luria-Bertani (LB) medium (1% (w/v) Bacto tryptone, 0.5% (w/v) Bacto yeast extract, 1% (w/v) NaCl, pH 7.5). To make solid medium 2% (w/v) agar was added prior to autoclaving. Where plasmid maintenance/selection was required, LB medium was supplemented filter sterilised with carbenicillin (100µg/ml final). Glycerol stocks of *E. coli* for long term storage were made by adding glycerol to a final concentration of 15% (w/v) to log phase cultures, aliquots placed in cryo-vials were snap frozen in liquid nitrogen then stored at -80°C.

2.3.2 Preparation and transformation of electro-competent *E. coli*

E. coli cells were grown to an OD_{600nm} 0.4 in 200ml LB medium at 37ºC, with shaking. Cells were incubated on ice for 30 minutes then harvested by centrifugation at 1500x*g* and 4ºC for 10 minutes. The cell pellet was then washed three times with 50ml ice cold sterile ddH_2O before being resuspended in 5ml 10% glycerol. The cells were harvested at 1500x*g* and 4ºC for 10 minutes and the cell pellet resuspended in 200µl 10% glycerol. The cells were snap frozen as 40µl aliquots and stored at –80ºC if not used immediately for electroporation.

For transformation reactions 1µl plasmid DNA or 2µl of a ligation/SDM was added to 40µl of electro-competent cells. The cells were mixed and transferred to cooled electroporation cuvettes. The cells were pulsed at 2.1kV using 25µF and 200Ω settings and 1ml of recovery medium (0.5% glucose, 10mM $MgCl_2$ in LB medium) was immediately added to the cells. The cells were allowed to recover for 1 hour at 37ºC, with shaking. The reactions were plated onto selective LB medium and incubated overnight at 37ºC.

2.3.3 Preparation and transformation of $CaCl_2$ competent *E. coli*

E. coli cells were grown in 200ml of LB medium with shaking at 37°C to an OD_{600nm} of 0.4. The cells were cooled on ice for 30 minutes, and then harvested by centrifugation at 1500xg and 4°C for 10 minutes. The pellet was gently resuspended in 10ml of ice cold 50mM $CaCl_2$ and the cells harvested again. Then cells were resuspended in 2.5ml of ice cold 50mM $CaCl_2$ left on ice overnight at 4°C.
For the transformation reaction 100µl of cells were mixed gently with DNA and incubated on ice for 30 minutes. The cells were then heat shocked by incubation at 42°C for 90 seconds and immediately incubated on ice for 10 minutes. Following this, 900µl of recovery Medium (0.5% glucose, 10mM $MgCl_2$ in LB medium) was added and the cells were allowed to recover for 1 hour at 37ºC, with shaking. The cells were then plated out onto selective LB medium and incubated overnight at 37°C.

2.3.4 Preparation of plasmid DNA from *E. coli*

A single colony of E. coli strain transformed with plasmid DNA was inoculated into 5ml of LB medium containing appropriate selective antibiotics, and the culture grown at 37ºC with shaking overnight. The cells were harvested by centrifugation at 2300x*g* for 5 minutes. The GenElute™ Plasmid mini-prep Kit (Sigma) was used to separate and purify plasmid DNA from the *E. coli* cells following manufacturer's guidelines. The plasmid DNA in each case was eluted in 80µl of ddH_2O.

2.4 *Saccharomyces cerevisiae* techniques

2.4.1 Growth and maintenance of *Saccharomyces cerevisiae*

S. cerevisiae was grown and maintained on YP medium (1% (w/v) Bacto-yeast extract, 2% (w/v) Bacto-peptone) supplemented with 2% (w/v) glucose (YPD). Plasmid selection was maintained on YNB minimal media (0.67% (w/v) yeast nitrogen base w/o amino acids) containing 2% (w/v) glucose appropriate amino acids

and supplements which were autoclaved separately prior to addition. Amino acids, adenine and uracil were added to a final concentration of 0.002%. To make solid medium 2% (w/v) agar was added prior to autoclaving. Strains containing the KanMX4 or hphMX3 marker were selected on YPD solid medium with the addition filter sterilised G418 (200µg/ml final) and/or Hygromycin B (300µg/ml final) respectively. Glycerol stocks of *S. cerevisiae* for long term storage were made by addition of glycerol to a final concentration of 15% (w/v) to log phase cultures, aliquots placed in cryo-vials were snap frozen in liquid nitrogen and then stored at -80°C.

2.4.2 Sporulation and tetrad dissection of *S. cerevisiae*

Diploid yeast strains were constructed by mixing haploid strains of opposite mating types on YPD solid medium. After incubating over night, diploids were positively selected on YNB solid medium using appropriate auxotrophic markers. Diploid *S. cerevisiae* strains were grown to stationary phase at 30°C, in pre-sporulation medium (0.8% Bacto-yeast extract, 0.3% Bacto-peptone) supplemented with 10% (w/v) glucose after autoclaving. Cells were washed with sterile ddH_2O and resuspended in 10 ml of sporulation medium (1% (w/v) potassium acetate, 0.1% (w/v) Bacto-Yeast extract) containing appropriate amino acid and supplements at a final concentration of 0.002% (w/v) and 0.05% (w/v) glucose. Sporulation was allowed to proceed by growing at 24°C for 3-4 days, after which the efficiency of sporulation was monitored microscopically. Once sufficient sporulation had occurred, 10µl of cells were diluted with 190µl of ddH_2O and treated with 5 units of Zymolyase T100 (AMS Biotechnology Ltd) for 5 minutes. Then 10µl of the treated spores were run vertically down a YPD solid medium plate and allowed to dry. A Narishige micro-manipulator microscope was used to separate the four spores from each ascus which were arranged 5mm apart. This was typically carried out upon a minimum of 10-20 complete tetrads. Spores were allowed to grow on YPD solid medium at 30°C for 2-3 days.

2.4.3 One-step transformation of *S. cerevisiae*

Transformation of plasmid DNA into *S. cerevisiae* was achieved by using the one-step protocol (Chen et al., 1992). Yeast cells were grown overnight in appropriate medium and temperature with shaking to stationary phase. For each transformation reaction 0.1-1ml of these cells were harvested by centrifugation at 16,000x*g* for 5 minutes. The cells were washed with 1 ml sterile ddH_2O and harvested as previously. These cells were resuspended in 100µl of one-step buffer (0.2M Lithium acetate pH5.5, 40% (w/v) PEG 3350, 100mM DTT) to this 50µg of salmon sperm DNA (denatured 95°C, then placed on ice) was added and finally 1µl of plasmid DNA was added. Reactions were mixed vigorously and heat-shocked at 42ºC for 30 minutes before plating onto selective YNB solid media. The transformations were incubated at permissive temperature for 3-5days.

2.4.4 High efficiency transformation of *S. cerevisiae*

For transformation of linear PCR products for gene disruption (*MAP1, ARD1& NAT3)* of *S. cerevisiae*, a high efficiency transformation protocol was used (Gietz and Schiestl, 2007). A culture of yeast cells was grown overnight in YPD medium at 30°C with shaking, was diluted to an OD_{600nm} of 0.5 in 50 ml YPD medium and incubated at 30 °C with shaking. When the cells reached an OD_{600nm} of 2.0, they were harvested by centrifugation at 3000x*g*. Cells were washed twice in 25 ml ddH_2O then cells were resuspended in 1ml of ddH_2O. For each transformation cells were harvested from 100µl of suspension by centrifugation at 13,000x*g* for 30 second in micro-centrifuge tubes. To the pellet 360µl of transformation mix (33% PEG-3350, 100mM lithium acetate, 100µg denatured salmon sperm DNA and 34µl PCR product) was added and mixed vigorously. The reactions were heat shocked at 42°C for 30 minutes. The cells were harvested by centrifugation at 13,000x*g* for 30 seconds then either; A) for eukaryotic antibiotic gene selection, cells were resuspended in 1 ml YPD medium and incubated at 30°C for 2-4 hours to recover prior to plating out on selective medium or, B) for prototrophic gene selection, cells were mixed with ddH_2O and directly plated out on selective medium. The plates were incubated at 30°C for 3-5 days.

2.4.5 Isolation of microsomes from *S. cerevisiae*

Preparation of yeast microsomes was performed using strain Δpep4 and was carried out as previously described (Wilkinson et al., 1996). Yeast cells were grown to an OD_{600nm} of 1.0-2.0 in YPD medium and then harvested by centrifugation at 3000 g for 5 minutes. Cells were resuspended in 100 mM Tris-SO_4, pH9.4, 10 mM DTT at a concentration of 50 OD_{600} units per ml and incubated at room temperature for 10 minutes. Cells were harvested as before and then resuspended in spheroplast buffer (0.75xYP medium, 0.7 M sorbitol, 0.5% (w/v) glucose, 10 M Tris.HCl, pH7.4) at a concentration of 100 OD_{600nm} units per ml. 1.5u of lyticase was added per OD_{600nm} unit of cells and the mixture incubated at 30˚C for 30 minutes. Spheroplasts were harvested by centrifugation at 2,500 g for 5 minutes and then resuspended in ice cold lysis buffer (0.1 M sorbitol, 50 mM KOAc, 20 mM HEPES, pH7.4, 2 mM EDTA, 1 mM DTT) at 200 OD_{600} units per ml. Acid-washed glass beads (0.5mm, Biospec) were added to 75% of the final volume and the mixture was vortexed for 30 seconds and then incubated on ice for 30 seconds. This was repeated a further 4 times and the cell breakage assessed microscopically. Once the spheroplasts were fully broken, a further 1 ml of lysis buffer was added and the mixture vortexed for 5 sec before centrifugation at 1000x*g* at 4°C for 10 minutes. The upper soluble fraction was separated and centrifuged again at 1000x*g* and 4°C for a further 10 minutes. The microsomes were isolated from the supernatant by centrifugation at 15,000x*g* The microsome pellet was washed by resuspension in membrane storage buffer (250 mM sorbitol, 20 mM HEPES, pH7.4, 50 mM KOAc, 1 mM DTT, 2 mM MgOAc). The microsome pellet was resuspended in membrane storage buffer at a concentration of 50 OD_{280nm} units per ml and snap frozen in small aliquots and stored at -80°C.

2.4.6 Extraction of cytosol from *S. cerevisiae*

Yeast cells were grown to an OD_{600nm} of 1.0-2.0 in 2 litres of YPD medium and then harvested by centrifugation at 3000x*g* for 5 minutes. The cells were washed twice in ice cold DEPC-ddH_2O and harvested by centrifugation at 3000x*g* for 5 minutes at 4˚C. The cells were then resuspended in a minimal volume of cytosol extraction

buffer (100mM KOAc, 2mM MgOAc, 2mM Glutathione, 20mM HEPES pH7.4, 0.5mM AEBSF). The cell suspension was then frozen as beads by using a pipette to produce droplets which were dropped into liquid N_2. The cell beads were then ground with liquid N_2 for approximately 40 minutes using a mortar and pestle to produce a fine powder. The powdered lysate was thawed on ice then centrifuged at 9384x*g* for 6 minutes at 4˚C, the supernatant was transferred to clean tubes and subjected to ultra-centrifugation at 100,000x*g* for 30 minutes at 4˚C. The clear yellow tinted fraction of the supernatant was loaded onto a PD10 desalting column (GE healthcare) which had previously been equilibrated with 30ml of cytosol extraction buffer containing 7.5% glycerol at 4˚C. The cytosol fractions were eluted from the column with 5 ml of cytosol extraction buffer containing 7.5% glycerol at 4˚C, 500µl fractions were collected and the protein content assessed by measuring OD_{260nm}. Fractions with an OD_{260nm}/ml $\geq$ 30 were snap frozen in liquid N_2 and stored at -80˚C for use in translocation assays.

2.4.7 Rapid isolation of genomic DNA from *S. cerevisiae*

Yeast cells were grown overnight to stationary phase in 2ml of YPD medium at 30ºC with shaking. Cells were harvested by centrifugation at 16,000x*g* for 5 minutes. The cells were then resuspended in 200µl of lysis buffer (2% Triton X-100, 1% SDS, 100mM NaCl, 10mM Tris-HCl pH 8.0, 1mM EDTA pH 8.0). This suspension was snap frozen in liquid N_2, then thawed rapidly at 95ºC for 1 minute, this freeze /thaw step was repeated and then the sample was vigorously mixed for 30 seconds. One volume of chloroform was added to the cell lysate and vigorously mixed for 2 minutes, then centrifuged at 16,000x*g* for 3 minutes. The upper aqueous phase was transferred to a clean tube containing 800µl of ice cold 100% ethanol, mixed by inversion, and left on ice for 30 minutes. Precipitated genomic DNA was harvested by centrifugation at 16,000x*g* for 5 minutes. The supernatant was removed and the pellet washed with 500µl of 70% ethanol and re-centrifuged and ethanol discarded. The pellet was air dried at room temperature, then resuspended in 50µl of ddH_2O.

2.2.8 Total proteins extracts from *S. cerevisiae* for SDS-PAGE

Yeast cells were grown in appropriate medium and temperature to an OD_{600nm} 0.2. Then 5 OD_{600nm} units of cells were harvested by centrifugation at 3800x*g* for 5 minutes. The cell pellets were washed with sterile ddH_2O and harvested again. The pellets were heated at 95ºC for 5 minutes then resuspended in 200µl of 1 x Laemmli buffer (10%(w/v) Glycerol, 3%(w/v) SDS, 62.5mM Tris.HCl pH6.8, 0.01%(w/v) Bromophenol blue, 5% β-mercaptoethanol). Acid washed beads were added to the meniscus and the samples vigorously shaken for 30 seconds at 6 m/sec in a Ribolyser (Hybaid). The samples were immediately heated at 95ºC (65ºC for membrane proteins) for 5 minutes. The total cell extract samples were recovered by piercing the end of the tube and placing it into a clean microfuge tube inside of a larger 15ml tube then this is centrifuged at 1000x*g*. Samples were subsequently stored at –20ºC.

2.4.9 Pulse labelling of *S. cerevisiae* proteins *in vivo*

Yeast cells transformed with wild-type CPY/OPY or signal sequence P2 mutants were grown in YNB medium, with 2% Glucose and appropriate supplements, to an OD_{600nm} = 0.2, and labelled with 10µCi of [^{35} S] methionine/cysteine mix (Perkin Elmer NEG772) per OD_{600nm} units of cells for 5 minutes at 30°C (20 minutes at 17°C for MWY63). Radio-labelling was terminated by addition of ice cold sodium azide to a final concentration of 20 mM. For each sample 5 or 10 OD_{600nm} units of cells were harvested for immunoprecipitation.

2.5 *In vitro* DNA manipulations

2.5.1 Agarose Gel Electrophoresis of DNA

DNA samples were supplemented with 10x DNA loading buffer (100mM Tris pH8.2, 50%(w/v) glycerol, 5mM EDTA, 0.25%(w/v) bromophenol blue) to a final

concentration of 1x loading buffer prior to separation on 0.6-2% agarose gels, depending on their size. Gels were run at 70V submerged in 1 x TAE buffer (40mM Tris, 20mM sodium acetate, 1mM EDTA, pH8.2). After electrophoresis agarose gels were soaked in 1 x TAE buffer containing 10μg/ml ethidium bromide for 15 minutes. The DNA fragments were visualised upon a 30 second exposure to blue light (460nm) in a LAS3000 FujiFilm Intelligent Dark Box.

2.5.2 Extraction of DNA fragments from Agarose gels

The fragments were excised from agarose gels using a sterile scalpel under long wave ultraviolet light. The DNA was purified from the gel using a QIAquick™ gel extraction kit following manufacturer's guidelines (Qiagen).

2.5.3 Restriction endonuclease digestion of plasmid DNA

Plasmid DNA was digested using commercially available enzymes with the buffers supplied. Digests were typically performed with 1-2 μg of plasmid DNA and 10units of enzyme. Digests were incubated at the recommended temperature for 1-2 hours. Products were then analysed by agarose gel electrophoresis.

2.5.4 Alkaline phosphatase treatment of DNA fragments

After restriction digest, DNA fragments can be dephosphorylated to prevent self ligation. To treat 1-5ug of DNA, 5 units of Antarctic alkaline phosphatase (NEB) was added to the reaction which was supplemented with 10x Antarctic Phosphatase reaction buffer to a final buffer concentration of 1x. The reaction mixture was incubated at 37°C for 15 minutes. Following this treatment, the sample was incubated at 65°C for 5 minutes in order to inactivate the alkaline phosphatase. Fragments were then separated by Agarose gel electrophoresis followed by gel extraction using a QIAquick™ kit (Qiagen)

2.5.5 Filling in DNA cohesive ends

DNA fragments with 5' overhangs following restriction digestion were filled using the large fragment (Klenow) of DNA ligase I (Roche). To a 1µg DNA sample, 1mM of the appropriate dNTPs and 1 unit of Klenow were added. The sample was incubated at 37°C for 15 minutes after which the Klenow enzyme was deactivated by incubation at 65 ^{o}C for 10 minutes.

2.5.6 Polymerase Chain Reaction (PCR)

Polymerase chain reactions (PCR) were performed using the DNA polymerase Phusion (Finnzymes) according to the manufacture's guidelines. A 50µl reaction mix typically contained 1x High fidelity buffer (supplied with enzyme), 200µM dNTPs, 0.5µM of each primer (forward & reverse), 0.5-1µl of template DNA and 0.5µl Phusion polymerase. Reactions were placed in a thermo-cycler with initial denaturation at 98°C for 30 seconds followed by 30 cycles of 15 seconds at 98°C, 20 seconds at 55°C and 30 seconds /Kb PCR product at 72°C, all followed by a final extension at 72°C for 5 minutes. Products were analysed by agarose gel electrophoresis.

2.5.7 Ligations of DNA fragments

DNA ligations were performed with a vector backbone to insert ratio ranging from 1:3 to 1:6. Reactions were performed in 10µl volumes and each reaction contained DNA fragments to be ligated, 2 units of T4 DNA ligase and 1µl of 10x ligation buffer supplied with the enzyme (NEB). The reactions were incubated at 24°C for 2 hours or at 16°C overnight. After incubation 5µl of the ligation reaction was used to transform $CaCl_2$ competent *E. coli*.

2.5.8 Site directed mutagenesis (SDM)

Site directed mutagenesis was carried out using the PCR based Quick Change™ site directed mutagenesis protocol (Stratagene). Primers were designed to make the appropriate mutations. Reaction mixtures of 50µl contained 20ng of template DNA, 125ng of both forward and reverse mutagenic primers, 1x Pfu PCR buffer, 240µM dNTPs and 2.5units *PfuTurbo* DNA polymerase. These reaction were placed in a thermo-cycler and subjected to an initial denaturation at 95 for 30 seconds, followed by 16 PCR cycles of; 95°C for 30 seconds, 45°C for 1 minute and 68°C for 2 minutes /kb of plasmid. Completed reactions were treated with 10 units *Dpn I* restriction enzyme at 37°C for 2 hours to digest the methylated, parental DNA of which 2µl was used to transform DH5α strain of *E. coli* by electroporation.

2.5.9 DNA sequencing

DNA sequencing was performed using the ABI PRISM™ BigDye™ (v1.1) terminator cycle reaction mix (Applied Biosystems). Reactions were carried out in 10 µl volumes and consisted of 2 µl Big Dye™ terminator mix, 5µl template DNA and 1.6pmoles of sequencing specific primer. The reactions were placed in a thermo-cycler and subjected to 30 cycles of; 98°C for 30 seconds, 45°C for 15 seconds and 60°C for 4 minutes. The reactions were diluted by addition of ddH_2O to a volume of 100µl, then the DNA precipitated by addition of 250µl of 95% ethanol and 10µl NaOAc pH 5.2 and left to stand at room temperature for 15 minutes. The samples were then centrifuged at 16,000x*g* for 20 minutes, the supernatant was removed and the pellet washed twice with 70% ethanol. After removal of all the liquid the pellets were dried by heating open tubes briefly at 65°C. Samples were sent for automated sequencing analysis.

2.6 Protein analysis techniques

2.6.1 SDS-PAGE analysis of proteins

Proteins samples were resolved by migration through SDS polyacrylamide gel electrophoresis (PAGE). SDS-PAGE resolving gels contained between 9-12.5% polyacrylamide, depending on the size of the protein to be resolved. A 30% (w/v) acrylamide stock solution (29:1 acrylamide:bisacrylamide, Bio-Rad) in resolving buffer (375mM Tris-HCl pH 8.8, 0.1%(w/v) SDS), a mixture of 20ml was polymerised by addition of 100µl of 10% (w/v) ammonium persulphate (APS) and 10-20µl tetramethylethylenediamine (TEMED). A stacking gel was layered over the resolving gel and consisted of 4.5% (w/v) acrylamide in stacking buffer (125mM Tris-HCl pH 6.8, 0.1% (w/v) SDS) a 10ml mixture was polymerised with 30µl 10% APS and 20 µl TEMED. Samples to be analysed were prepared in Laemmli buffer and loaded into wells in the stacking gel. Gels were run at 25mA (or at 5mA overnight) in SDS-PAGE running buffer (25mM Tris-HCl, 192mM glycine, 0.1% (w/v) SDS). Once complete gels were either fixed in 10% acetic acid with 10% methanol for 30 minutes before drying at 80°C under vacuum, or transferred to PVDF membrane for western blotting.

2.6.2 Western blotting analysis

Following electrophoresis, proteins were transferred to PVDF (Millipore) membrane by Western transfer. A sandwich consisting of three sheets of Whatman 3MM filter paper soaked in transfer buffer (20mM Tris-base, 150mM glycine, 5% (v/v) methanol), the gel, PVDF membrane (soaked in 100% Methanol prior to transfer buffer), and three more sheets of soaked filter paper. The sandwich was placed in a transfer tank containing transfer buffer, the gel adjacent to the cathode, and transferred at 100mA overnight or 800mA for 1.5 hours. The membrane was washed in blocking solution (2% (w/v) dried skimmed milk powder in TBS-NP40 (130mM NaCl, 2.6mM KCl, 2mM Tris-base pH7.6, 0.1% (v/v) NP40) for a minimum of 30 minutes, changing the blocking solution at least once. The membrane was then

incubated in 30ml blocking solution containing primary antibody (CPY diluted 1:10,000) for 1 hour after which the antibody solution was removed and the membrane washed in blocking solution three times for 5 minutes, The membrane was then incubated with a secondary antibody conjugated to horse radish peroxidase (HRP), immunoreactive against the species that the primary antibody was raised in, (for CPY, anti-sheep IgG diluted 1:10,000) in blocking solution for 1 hour. Excess secondary antibody was washed off as previously. The membrane was washed twice in TBS-NP40 for 5 minutes. Protein was identified by incubating the membrane in enhanced chemiluminescence reagent (NEN) for 1 minute and visualised in a LAS 3000 imager FujiFilm Intelligent Dark Box.

2.6.3 Immunoprecipitation of proteins

This method was adapted from that previously described (Stirling et al., 1992). Yeast spheroplasts were prepared by incubation of cells in IP spheroplast buffer (1.4M sorbitol, 50mM Tris.HCl pH7.4, 2mM $MgCl_2$, 10mM NaN_3) including 6units/OD_{600nm} of Zymolyase 100 T (AMS Biotechnology) at 30°C for 30 minutes then spheroplasts were harvested by centrifugation at 1000x*g* (spheroplast procedure was omitted for *in vitro* translated samples). Spheroplasts or *in vitro* translation samples were resuspended in IP lysis buffer (1% SDS, 50mM Tris-HCl, pH7.4 and 5mM EDTA) and incubated at 95°C for 5 minutes, followed by two minutes on ice. Samples were then diluted with 1mL of IP buffer (1.25% Triton-X-100, 190mM NaCl, 62.5mM Tris-HCl pH7.4, 6.25mM EDTA) and 50ul insoluble protein A cells (Cowan strain) added to preclear the lysates which were rotated at 4°C for 1 hour. The insoluble cells were removed by micro-centrifugation at 12000x*g*. Antiserum was added to the supernatant and preincubated for 1 hour prior to addition of 80µl of Protein A sepharose (10% (w/v) in IP Buffer) which was then incubated for a further hour. Protein A sepharose was sedimented by centrifugation washed three times with IP buffer and sedimented a final time discarding all the supernatant.

Samples for SDS-PAGE analysis had 50µl of 2x SDS-PAGE sample buffer added to the sepharose pellet and were incubated for 5 minutes at 95°C. Radio-labelled proteins were then subjected to SDS-PAGE electrophoresis and visualised either by

phosphorimager FLA3000 or autoradiograph. Quantification was performed using AIDA image analyzer software.

Samples for scintillation counting were dissociated from the sepharose beads with 50µl of 3% SDS at 95°C for 5 minutes, then samples were dried onto whatman glass GF/A filter discs which were added to a scintillation vial containing 4ml of scintillation fluid and then counted ([^{35}S] 1 minute at 0-2000 KeV, [^{14}C] 5 minutes at 4-156 KeV) on a Tricarb 2100TR liquid scintillation counter (Packard).

2.7 *In vitro* translation and translocation techniques

2.7.1 *In vitro* transcription of mRNA

Transcription templates of various ppαF mRNAs were from PCR products of either pEH3 or pGF22 plasmids using primers for MR, MS or MSR ppαF and the respective reverse primer (Table 2), and were transcribed with SP6 polymerase. Templates for mRNA transcriptions of OαF were from PCRs of pGF24 or pGF25 for MR and MS Oα-F respectively. Transcriptions were carried out from with T7 polymerase. All transcriptions were performed according to the manufacture's protocol (Promega).

2.7.2 *In vitro* translation/translocation assay in rabbit reticulocyte lysate

Translations were performed in rabbit reticulocyte lysate system according to the manufacture's protocol (Promega) with the inclusion of either 2.04µCi [^{35}S] Methionine, 0.04µCi [^{14}C] Acetyl Coenzyme A or 0.05µCi [^{14}C] Amino acid mix (Perkin Elmer, NEG709A, NEC313 or NEC445 respectively) per 10µl of reaction and 0.4µl mRNA and incubated at 30°C for 30 minutes.

Co-translational translocation reactions were performed with Oα-F mRNAs. These reaction were carried out as the translations above but were supplemented with 0.1µM purified SRP and 1.5µl ERKM dog microsomes prior to incubation at 30°C for 30 minutes.

Post-translational translocations were performed with α-F mRNAs. Translation was terminated by addition of Cycloheximide to a final concentration of 2mM. Then 10 µl of translation reaction was incubated with 2µl of yeast microsomes for 20 minutes at 30°C.

2.7.3 Isolation and analysis of ribosome nascent chain complex (RNC)

RNCs were isolated by ultra-centrifugation of translations though a low salt sucrose cushion (0.5M sucrose, 40mM HEPES pH7.5, 150mM KOAc, 2mM MgOAc, 2mM DTT, 0.002% Nikkol) at 300,000x*g* for 40 minutes. The pellet isolated will contain the RNCs. This pellet was resuspended in 2x Laemmli buffer for analysis by SDS-PAGE followed by phosphorimager analysis. Alternatively the RNCs were analysed by scintillation counting. For this analysis background radioactivity was minimised by TCA precipitation of the RNCs as follows; the pellet was resuspended in 50µl of ddH_2O and 50µl of 2M NaOH then incubated at 37°C for 10 minutes, these samples were then cooled on ice and 60µl of 0.1% Triton-X-100 mixed in followed by 200µl of TCA and incubated on ice for 30 minutes. The samples were then filtered under vacuum on GF/A glass filters, the filters were washed three times with 3ml of ice-cold 5% TCA and then once with 3ml of 100% acetone. The filters were then air dried and place in to a scintillation vial containing 4ml of scintillation fluid and then counted ([^{35}S] 1 minute at 0-2000 KeV, [^{14}C] 5 minutes at 4-156 KeV) on a Tricarb 2100TR liquid scintillation counter (Packard).

2.7.4 Translations/translocation assay in yeast cytosol

The following mix was added to 25µl of cytosol:
1µl 100mM DTT, 2.5µl 20x Salt (0.4M HEPES-KOH, pH7.4, 3.2M KOAc, 0.04M MgOAC), 2µl 1mM amino acid mix minus methionine, 2.5µl 0.5M Creatine phosphate, 1.25µl 8mg/ml Creatine phosphokinase, 1.25µl ATP/GTP (20mM/4mM), 1µl 12.5mM $CaCl_2$, 1.25µl micrococcal nuclease (0.5-1mg/ml).

This reaction mixture was incubated at 20°C for 30 minutes to allow the micrococcal nuclease to remove endogenous nucleic acids. The nuclease treatment was terminated by the addition of 3µl of 12.5mM EGTA. To this nuclease treated cytosol 15 µCi of [^{35}S] Methionine was added and then 10µl aliquots were used for each sample. If the sample was a co-translocation reaction 1.5µl of yeast microsomes was added to the cytosol at this point, then to each reaction mix 0.5µl of mRNA was added and then incubated at 20°C for 45 minutes. If the sample was a post–translocation reaction microsomes were omitted from the previous incubation step and instead, after the 45 minute incubation translation was terminated by the addition of cycloheximide to a final concentration of 0.4mM, the microsomes were then added and the reaction incubated for a further 45 minutes at 20°C. After incubation 20µl of 2x Laemmli loading buffer was added to each sample prior to SDS-PAGE analysis and visualisation by phosphorimaging.

2.7.5 N-terminal biotinylation

In vitro translations (20 µl scale), programmed with lysine-free OαF mRNAs, were performed as above in the presence of [^{35}S] methionine. Proteins were recovered by precipitation with ammonium sulphate and then re-precipitated with ethanol. The samples were then denatured in PBS+1% SDS for 10 minutes at 65°C. Free N-termini were then modified by treatment with 1mM sulpho-NHS-SS-Biotin (Pierce) for 20 minutes at 37ºC. After removal of free biotinylation reagent by acetone precipitation, samples were resuspended in PBS+0.1% SDS and then biotinylated proteins recovered on immobilized-streptavidin beads (Pierce). Beads were washed five times within PBS+0.1% SDS and bound protein eluted in SDS-PAGE sample buffer containing 2% (v/v) β-mercaptoethanol. Radio-labelled proteins were visualized with SDS-PAGE and phosphorimaging.

2.8 Bioinformatics data processing

The set of *S. cerevisiae* signal sequence-containing proteins was obtained from the signal peptide database (SPdb) v 5.1 (Choo et al., 2005). This set of 291 sequences was manually filtered for duplicates, dubious ORFs (as defined by the *Saccharomyces* genome database (SGD) and proteins known to be localized to mitochondria, to yield a final filtered set of 275 ORFs. For a complete list of ORFs see Supp. Table 1. The P2 amino acid frequency distribution did not differ significantly between the filtered and unfiltered sets (χ^2=5.17, 19 d .f.). Graphical and statistical analysis was performed using Prism 4.0 (GraphPad). MetAP cleavage was assumed for P2 residues A, C, G, P, S, V and T (Boissel et al., 1988, Huang et al., 1987). Yeast cytosolic and mitochondrial protein datasets were generated by random selection from SGD from proteins within known localization. N-α-acetylation prediction was performed according to (Martinez et al., 2008), where appropriate the P3 residue was also taken into consideration. MN, which is only predicted to lead to N-acetylation in 55% of cases, was scored as acetylated. Total amino acid frequency was determined from codon tables for the entire yeast genome at SGD. Human and *C. elegans* signal sequence data sets were also obtained from the signal peptide database (SPdb) v 5.1 (Choo et al., 2005). Peak hydrophobicity was determined by Kyte Doolittle using a window size of 11 (Kyte and Doolittle, 1982, Ng et al., 1996).

Results

3. Results

3.1 Background

It has previously been discussed in chapter 1 that N-terminal modification events are wide spread occurring on a vast majority of proteins. The bioinformatics data in section 1.5 demonstrate that proteins with signal sequences have a strong bias at position 2 (P2) against particular amino acids which may promote N-terminal modification. This bias is so striking that this study aims to determine if it has some biological significance. An obvious difference between these proteins compared to those of the cytosol is that they are targeted to the ER via their N-terminal signal sequences, this part of the sequence being essential for their correct localisation. One reason this bias may exist is that if N-terminal modification were to occur on signal sequences it may affect targeting of this group of secretory proteins. Interestingly N-terminal modification of secretory proteins has not been extensively studied. This may be because signal sequences are removed in the process of translocation, revealing a new N-terminus and therefore the status of the initiator methionine being transient may have been deemed unimportant.

3.2 Mutation of position 2 of CPY signal sequence

To gain a more in depth understanding of the observation which indicates that secretory proteins with signal sequences have a strong bias to retain their initiator methionine, this phenomenon was investigated experimentally. For the initial investigations the well characterised secretory protein pre-pro-Carboxypeptidase Y (ppCPY) was used as a substrate, this protein has been previously introduced in section 1.2.5. ppCPY has a signal sequence of 20 amino acids of which the second residue is a lysine and according to the criteria outlined in section 1.3 ppCPY is therefore not predicted to be a substrate for N-terminal modification. As the second amino acid residue of a sequence is the major determinant of N-terminal modification, it was decided that residues should be introduced at position 2 in the ppCPY signal sequence to promote modification and observe any affect this may

have on translocation into the ER. Changes to the ppCPY signal sequence were made by a PCR method of site directed mutagenesis (SDM), this method uses mutagenic primers during cycles of PCR to insert a codon for various amino acids between the initiator methionine codon and lysine at position 2 of the natural sequence (Figure 3.1). Insertion mutations were introduced rather than replacements to maintain the elements of the original signal sequence. Following SDM, codon insertion was confirmed by DNA sequencing of the resulting constructs.

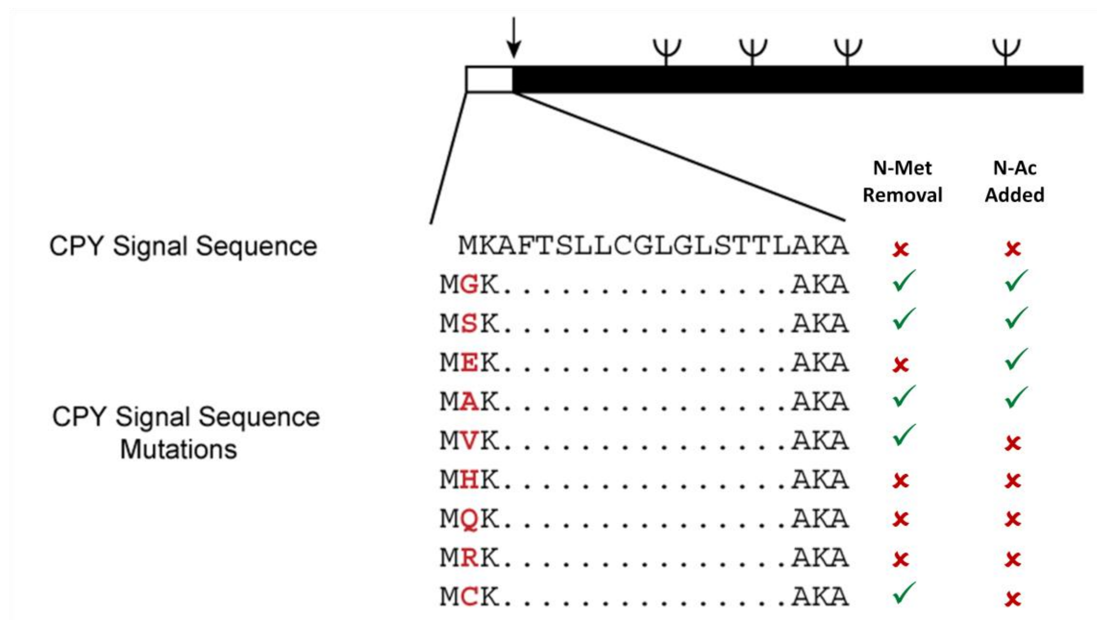


Figure 3.1 ppCPY signal sequence and position 2mutations. Cartoon of CPY showing the signal sequence and mutations of amino acid insertion at position 2, made by SDM. Position of N-glycosylation (ψ) and signal peptidase cleavage (↓) sites are indicated. Also shown, is a prediction of possible N-terminal modification events (N-Met removal & N-Ac added) based on the residue at position 2.

3.3 Translocation of ppCPY position 2 signal sequence mutants *in vivo*

Translocated ppCPY is immediately N-glycosylated, leading to a decrease in mobility during SDS-PAGE (Figure 3.2). This provides a simple assay by which to follow its progress to the ER.

The ppCPY signal sequence mutants were therefore assessed for their ability to translocate to the ER by transforming plasmid constructs expressing the wild-type and various P2 insertion mutants of ppCPY into GFY3 (*Δprc1, Δpep4*) a strain in

which the genomic ORF encoding ppCPYp (*PRC1*) has been deleted so only the recombinant ppCPY will be expressed. In addition GFY3 also has a deletion of the *PEP4(Δpep4)* gene encoding proteinase A, which is required for cleavage of the CPY pro sequence for progression to the mature form. Mature CPY migrates with similar mobility during SDS-PAGE to untranslocated ppCPY, assessing translocation of CPY can therefore be made easier by eliminating the ability of ppCPY to reach the mature form (Figure 3.2). The various ppCPY construct transformants of GFY3 were pulse-labelled for 5 minutes with [^{35}S] methionine/cysteine mix and total CPY was immunoprecipitated, separated by SDS-PAGE and the radio-labelled forms visualised by phosphorimaging (Figure 3.3).

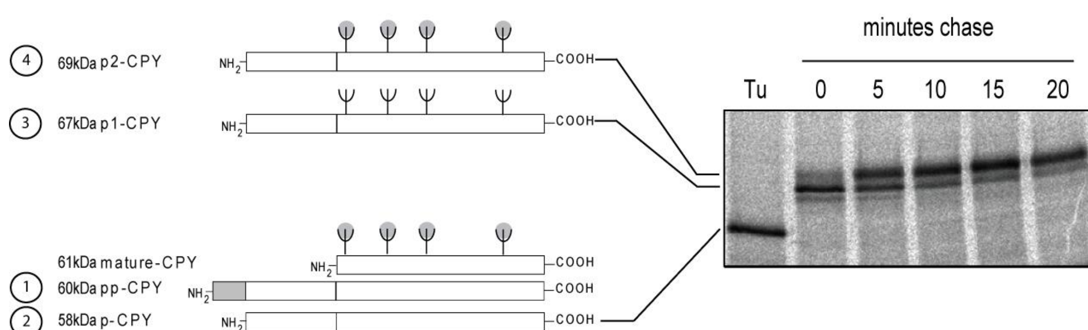


Figure 3.2 Carboxypeptidase Y processing and pulse-chase in a *Δpep4* yeast strain. Cartoon of the various forms of CPY showing the order of processing (1-4) in a pep4 strain and their relative migration. A *Δpep4* yeast strain was subjected to a 5 minute pulse-labelling with [^{35}S] methionine/cysteine followed by cycloheximide chase, samples were taken at time points between 0-20 minutes, immunoprecipitated with CPY antiserum and analysed by SDS-PAGE and phosphorimaging. ppCPY (1) is not seen due to its transient nature while mature-CPY does not arise in the *pep4* mutant. Also shown is cell treated with Tunicamycin (Tu).

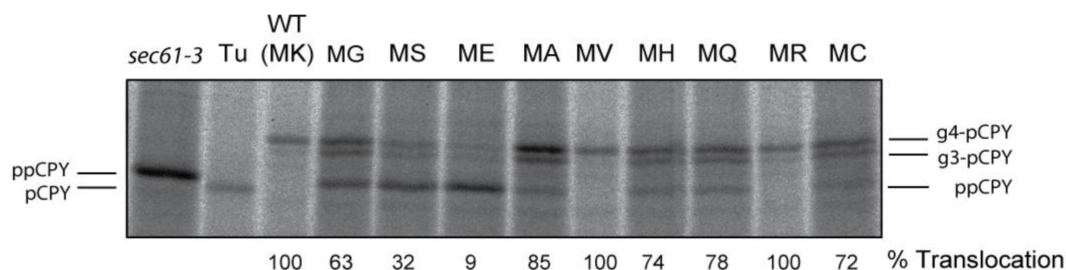


Figure 3.3 N-terminal modification leads to defective ER translocation of CPY. Translocation of ppCPY was analysed in yeast strain GFY3 *(Δpep4 Δprc1)* expressing plasmid copy of either wild-type (WT) or various signal sequence mutants (lanes 4-12) of ppCPY. Cells were pulse-labelled with [^{35}S] methionine/cysteine, immunoprecipitated with CPY antiserum and then analysed by SDS-PAGE and visualised by phosphorimager. Position of fully-(four-times, g4-pCPY), triply- (g3-pCPY) glycosylated CPY is indicated. Also shown are wild-type CPY treated with Tunicamycin (Tu) (pCPY) and also wild-type CPY expressed in *sec61-3* strain (ppCPY) at 17°C.

As expected the wild-type protein was translocated indicated by the presence of a single species corresponding to the glycosylated form (g4-CPY) of the protein (Figure 3.3). In contrast, many of the mutants show an accumulation of ppCPY indicative of a translocation defect, as observed when wild-type protein is expressed in a translocation deficient *sec61-3* cold sensitive strain at the non-permissive temperature. In all the signal sequence mutants tested we observed some under-glycosylation (g3-pCPY). This can be seen to some extent for wild-type pCPY but is exacerbated for the P2 mutants.

Interestingly, the majority of the mutants which have a translocation defect are predicted to undergo N-terminal modification (Figure 3.1). In addition the most severe effects are seen for glycine (MG), serine (MS) and glutamate (ME) mutants, which correlates to those having the strongest bias for appearing in P2 of cytosolic proteins versus signal sequence containing secretory proteins (Figure 1.12).

3.4 The affect of MetAP activity on signal sequences

The ppCPY mutants MA, MC, ME and MG are all predicted to be N-terminally modified in some way, in addition they all display some degree of translocation defect. This phenotype could be caused by N-terminal modification or alternatively these P2 residues themselves could affect the function of the signal sequence. To explore whether the translocation defects seen were a result of N-terminal modification the initial investigations focused upon N-terminal methionine removal by MetAPs. *S. cerevisiae* has two isoforms of MetAP, encoded by the genes *MAP1* and *MAP2*, the null mutant strains lacking either of these genes are viable as these proteins show some degree of functional redundancy with reference to N-terminal methionine removal (Chen et al., 2002, Li and Chang, 1995). A *Δmap1/Δmap2* double null mutant is inviable showing MetAP activity is an essential process. However the drug Fumagillin specifically and irreversibly inhibits Map2p activity (Chen et al., 2004, Sin et al., 1997). Treating a *Δmap1* strain with Fumagillin, abolishes Map2p activity leading to a complete defect in N-terminal methionine removal. To explore whether defective translocation for any of the ppCPY mutant could be attributed to MetAP activity, the approach was to investigate the effect upon translocation of these mutants in yeast strains lacking MetAP activity (Figure 3.4). The *MAP1 MAP2* strain (GFY3) and a *Δmap1 MAP2* strain (GFY7) were transformed with the ppCPY constructs. The resulting transformants were used for analysis of CPY in pulse-labelling experiments as performed in section 3.3 but this time yeast cell cultures either remained untreated or were treated with Fumagillin prior to labelling with [^{35}S] methionine/cysteine.

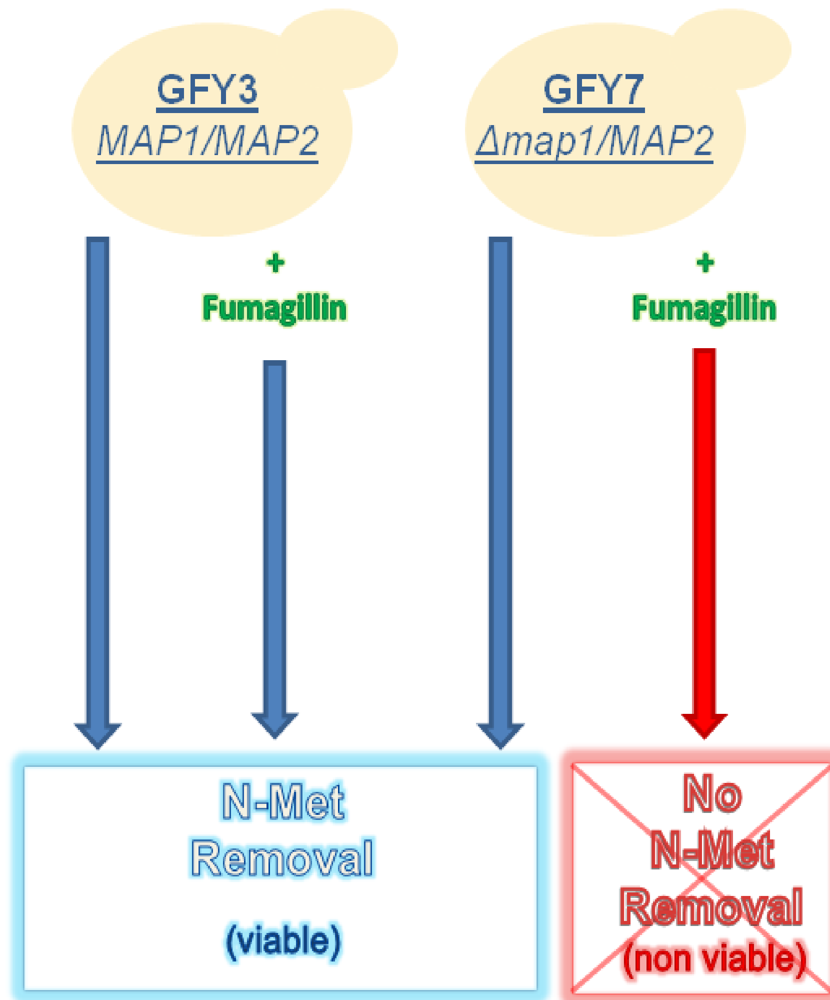


Figure 3.4 Schematic diagram of MetAP activity in the absence & presence of Fumagillin. Both GFY3 (*MAP1,MAP2)* and GFY7 (*Δmap1,MAP2)* strains have sufficient MetAP activity to carry out N-terminal methionine removal. Fumagillin. is a drug that can specifically and irreversibly inhibit Map2p activity. In GFY3 treated with Fumagillin N-terminal methionine removal still occurs via the activity of Map1p. However, treating GFY7 strain with Fumagillin, thus inhibiting activity of is only MetAP (Map2p), effectively makes a strain lacking all MetAP activity and consequently completely defective in N-terminal methionine removal.

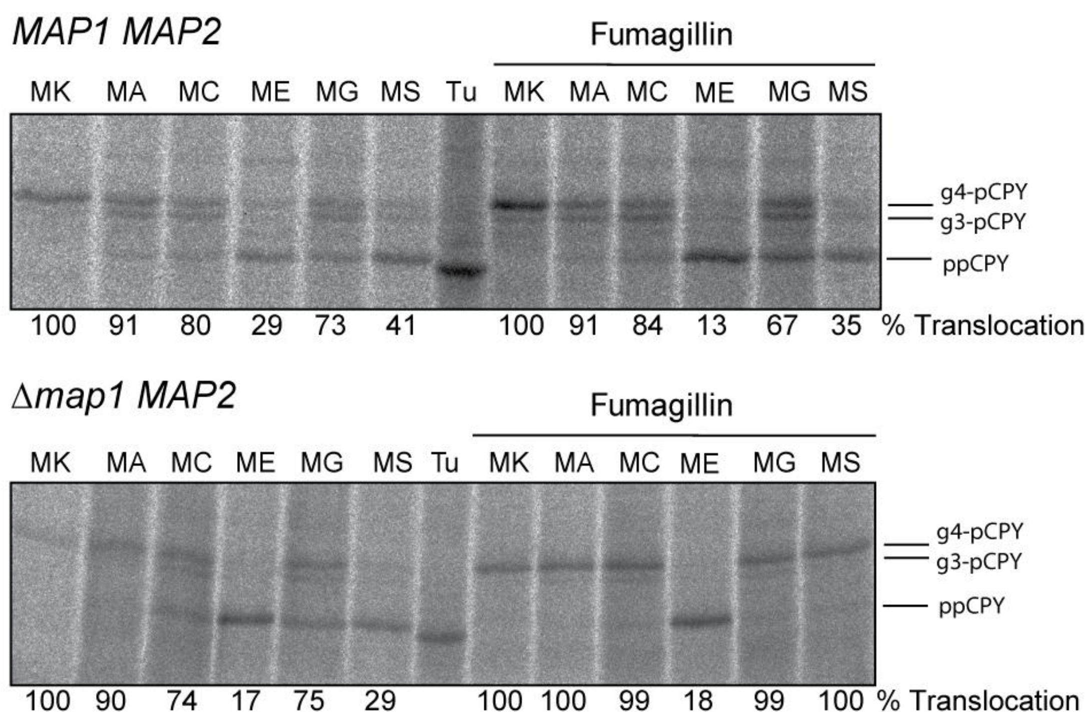


Figure 3.5 N-terminal Methionine removal affects CPY translocation. Translocation of CPY and P2 mutants were analysed in *MAP1 MAP2* strain (GFY3) and an isogenic *Δmap1 MAP2* strain (GFY7) in the absence or presence of treatment with 3μM Fumagillin for 30 minutes prior to radio-labelling. Cells were pulse-labelled with [^{35}S] methionine/cysteine and then CPY was immunoprecipitation then analysed by SDS-PAGE and visualised by phosphorimager.

As seen previously (Figure 3.3) the expression of ppCPY position 2 mutants in the *MAP1/MAP2* (GFY3) led to accumulation of ppCPY (Figure 3.5), this was not altered by the treatment with Fumagillin to inhibit Map2p activity in this strain. Similar results were seen when these ppCPY constructs were expressed in the *Δmap1* strain, in absence of one of the two MetAPs (Map1p). There was little effect on translocation, only the MA mutant showed some recovery of its translocation defect. In striking contrast, when the *Δmap1* strain was treated with Fumagillin, thus creating a strain deficient in MetAP activity, a complete restoration of translocation was observed for the MA, MC, MG and MS mutants. These data indicate that it is not the insertion of the amino acids A, C, G or S that affects the signal sequence rather that the inability of these substrates to translocate efficiently is dependent on MetAP activity. Unlike the other mutants, the translocation defect of the ME mutant persists in the presence of Fumagillin and so is independent of MetAP activity. This is unsurprising based on the known specificity of MetAPs, which would predict

initiator methionine cleavage of the MA, MC, MG and MS mutants but not the ME mutant (Huang et al., 1987). It is possible that ME is subject to some other form of N-terminal modification. From these data we can conclude that removal of the initiator Methionine of a signal sequence by MetAPs, can lead to a significant effect on translocation of ppCPY.

3.5 The effect of N-α-Acetylation on signal sequences

Initiator Methionine removal by MetAPs is unlikely to be the sole cause of translocation defects described above, as the observations show that the translocation defect associated with the ME mutant is independent of the activity of MetAPs. If another common modification event is considered, N-α-Acetylation, it is predicted that this modification would occur on the ME mutant of the ppCPY signal sequence (Figure 3.1) based on previous findings (Martinez et al., 2008, Polevoda and Sherman, 2003b). When considering N-terminal modification by MetAPs it is sufficient to predict modifications based on the P2 residue as this is generally the major determinant, with a few exceptions where the residue at position 3 influences N-terminal methionine removal for example if this position is occupied by proline (Moerschell et al., 1990). Prediction of N-α-Acetylation is more complicated, the primary determinant for this modification is still the residue at position 2, but is also influenced by a more extensive motif at the N-terminus of a protein (Moerschell et al., 1990, Polevoda and Sherman, 2000). Our predictions may be made more accurate by considering some of the other sequence motifs for N-α-Acetylation. The simplest way is to analyse the sequence of ppCPY and the various mutant constructs is by using an N-terminus prediction tool Termi*N*ator available at ExPASy Proteomics tools (http://www.expasy.ch/tools/), this predictive tool has been compiled based on extensive studies of N-α-Acetylation (Frottin et al., 2006, Martinez et al., 2008). The Termi*N*ator tool predicts that both the ME and the MS mutants of ppCPY could be N-α-Acetylated (Table 3.1). This is interesting given that these mutants also display the most striking translocation defects in the GFY3 strain (Figure 3.3).

Table 3.1 N-terminal modification prediction of CPY and P2 mutants by Termi*N*ator

CPY construct	N-Met removal	N-Acetylation	Predicted N-terminus	Likelihood %
Wild-type (MK)	×	×	M(1)	83
MA	✓	×	A(2)	71
MC	✓	×	C(2)	87
ME	×	✓	Ac-M(1)	93
MG	✓	×	G(2)	80
MS	✓	✓	Ac-S(2)	75

N-terminal modification as predicted by Termi*N*ator (version3). This tool predicts N-terminal modification of a particular sequence by comparing it to sequences known to be modified. The original position of predicted N-terminus is shown in brackets.

To investigate whether N-α-acetylation could play a role in the translocation defects observed for the MS and ME mutants, a strategy was devised to prevent N-terminal acetylation to see if this would reverse the translocation defect. N-α-acetylation can be carried out by several different N-α-acetyl transferase complexes (NATs), as detailed in section 1.3.2 each NAT complex has a distinct set of defined substrates. The MS mutant would be N-α-acetylated by the NatA complex after removal of the initiator methionine and the ME mutant would have the initiator methionine N-α-acetylated by the NatB complex (Figure 1.10). Therefore mutant strains were constructed which have been shown to prevent acetylation of either NatA or NatB substrate proteins in other studies (Polevoda et al., 1999). These mutations correspond to deletions of the open reading frames for the catalytic subunits for each complex, namely Δ*ard1* (NatA) and Δ*nat3* (NatB). The translocation efficiency of wild-type (MK), MS and the ME mutants of was analysed in these strains (Figure 3.6) by pulse-labelling experiments as in section 3.3.

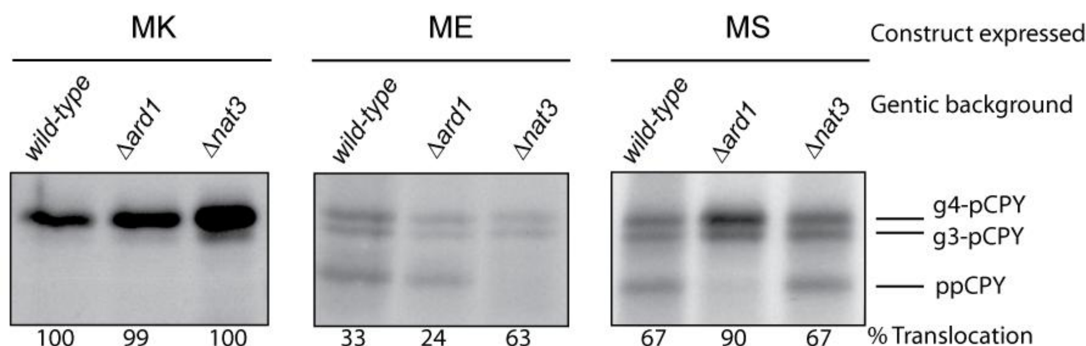


Figure 3.6 N-α-Acetylation affects translocation of CPY mutants to the ER. Translocation of wild-type (MK) CPY and mutants ME and MS were analysed in wild-type (GFY3) and in NatA (*Δard1)* or NatB (*Δnat3)* catalytic subunit mutant strains GFY11 and GFY12, respectively. Yeast cells were pulse labelled with [^{35}S] methionine/cysteine, extracts were immunoprecipitated with CPY antiserum and then analysed by SDS-PAGE and visualised by autoradiography.

The translocation of wild-type ppCPY is unaffected by activity of NATs and translocates in each of the catalytic subunit mutants. However, the data for the ME mutant shows translocation defects in both the 'wild-type' and Δ*ard1* strains, only in the Δ*nat3* strain is translocation largely restored. Whilst for the MS mutant a translocation defect is seen in the 'wild-type' and Δ*nat3* strains and only restored in the Δ*ard1* strain. Thus translocation of the CPY mutants ME and MS can be restored by disrupting catalytic subunits of their specific NAT complex required for their N-terminal acetylation. Therefore these experiments strongly suggest that it is the N-α-acetylation of these substrates that somehow restricts their translocation to the ER.

3.6 An *in vitro* analysis of N-terminal modification effects on signal sequences

The data in sections 3.4 & 3.5 suggest that modification of a signal sequence by either MetAPs and/or NAT complexes has an effect on sorting of the secretory protein ppCPY to the ER *in vivo*. To determine if the effect of N-terminal modification seen for ppCPY could be shown for another substrate and further to this could also be reconstituted *in vitro* it was decided to perform an investigation of N-

terminal modification *in vitro*. For these reasons a well characterised secretory protein prepro-alpha factor (ppαF) was used. This secretory protein is a mating pheromone encoded by the gene *MF(ALPHA)1*. Like the previous substrate ppCPY, the wild-type form of ppαF is not predicted to be a substrate for either MetAPs or NATs as it has arginine as its P2 residue.

3.6.1 Translocation of ppαF and position 2 mutants

Mutation of ppαF was performed by PCR amplification of the wild type open reading frame using a 5' primer containing the desired mutations. The PCR products of wild-type and mutated ppαF were subsequently used as a template for the production of mRNA transcripts. The first mutation tested involved insertion of a serine at position 2 which is predicted to be both processed by MetAPs and subsequently N-α-acetylated (Figure 3.7).

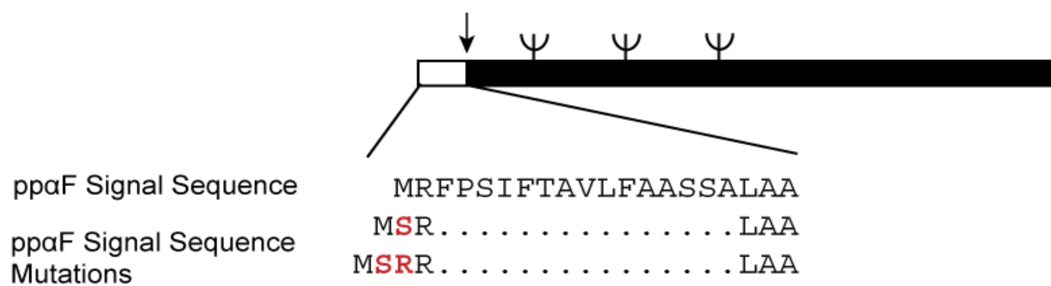


Figure 3.7 ppαF signal sequence mutation at position 2. Cartoon of ppαF showing the signal sequence and mutations of amino acid insertion at position 2, introduced by PCR amplification using mutating primers. The position of N-glycosylation (ψ) and signal peptidase cleavage (↓) sites are indicated.

Transcribed mRNA encoding wild-type (MR) and the position 2 serine mutant (MS) of ppαF were translated and radio-labelled, *in vitro* in a rabbit reticulocyte lysate system. Translated products were then incubated in the absence or presence of yeast microsomes to assess translocation of these proteins (Figure 3.8A).

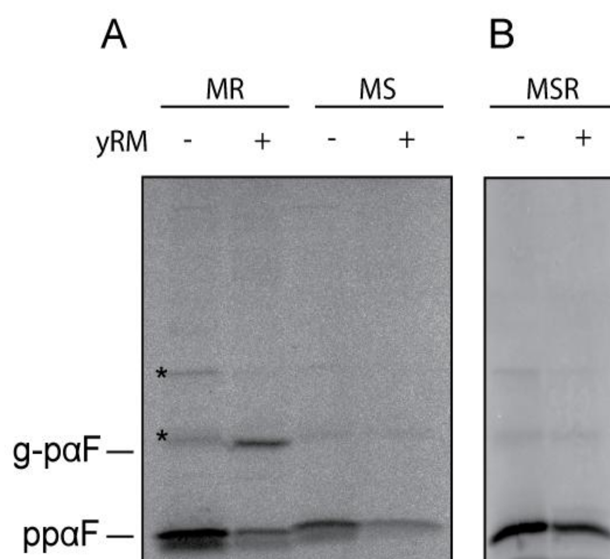


Figure 3.8 Translocation assay for in vitro expressed ppαF. Wild-type (MR), MS and MSR ppαF were translated *in vitro* in rabbit reticulocyte lysate, translation was inhibited with 2 mM cycloheximide followed by incubation in the absence or presence of yeast microsomes (yRM). Position of non-translocated (ppαF) and signal-sequence cleaved, glycosylated (g-pαF) are indicated. Note that species generated in the absence of microsomes (*) which most likely corresponds to ubiquitinylated ppαF the faster migrating of these migrates slightly slower but close to the position of g-pαF.

As expected, wild-type ppαF is glycosylated indicating that it is translocated in the presence of microsomes. However, the MS mutant shows no evident glycosylation or detectable signal sequence cleavage. This indicates that mutation of ppαF signal sequence to an N-terminally processable MS prevents translocation. These data reveal that effects observed following mutation of ppCPY can be replicated for another secretory substrate and reconstituted *in vitro*. Moreover, the use of mammalian cytosol suggests that the mechanism may be conserved. Based on the findings with the ppCPY mutants it was predicted that this block in translocation of the ppαF-MS mutant could be a result of its N-α-acetylation after removal of the initiator methionine by MetAP. N-α-acetylation of the MS mutant would also have the effect of altering the charge of the signal sequence by +1 (the acetyl group being -1). Positive charge at the N-terminus is a common feature of a signal sequences (von Heijne, 1985). Therefore it was necessary to determine if it was this change in charge, rather than the addition of the acetyl group itself, which causes the translocation defect of ppαF-MS. To investigate if this was the case a mutant of

ppαF-MS was made, where an additional Arginine (+1) was inserted at position 3 (ppαF-MSR) (Figure 3.7), thus maintaining the positive charge in the signal sequence even if it was N-α-acetylated. The ppαF-MSR was also unable to translocate (Figure 3.8B). This observation informs us that the translocation defect seen for the MS mutant is not due simply to the change in positive charge of the signal sequence following its acetylation.

3.6.2 Assessing N-α-Acetylation of the ppαF position 2 serine mutant

Whilst it has been concluded up to this point that the translocation defects are a consequence of N-α-acetylation the addition of an acetyl group to our substrates has not been directly shown. To verify if ppαF-MS is acetylated in an *in vitro* system, the approach was to directly translate the substrates in the presence of [^{14}C] Acetyl-CoA, which is the acetyl group donor for N-α-acetylation and then assess incorporation the [^{14}C] acetyl group by scintillation counting the immunoprecipitated ppαF.

In addition to N-terminal acetylation proteins can be post-translationally acetylated on their internal lysine residues. ppαF has nine lysine residues, none of which are in the signal sequence, and of these lysines five are predicted to be sites for internal acetylation according to PAIL a tool which predicts such events http://bdmpail.biocuckoo.org/prediction.php (Li et al., 2006). Therefore versions of ppαF and ppαF-MS were made in which all the lysines had been substituted for arginine residues. The lysine-less versions ppαF behave similarly to the native ppαF in terms of their translocation (Figure 3.9 A&B); the wild-type (MR) can translocate but the MS mutant cannot. It is also worth noting that as there are no lysines in these proteins the ubiquitinated forms of ppαF are absent in these translations.

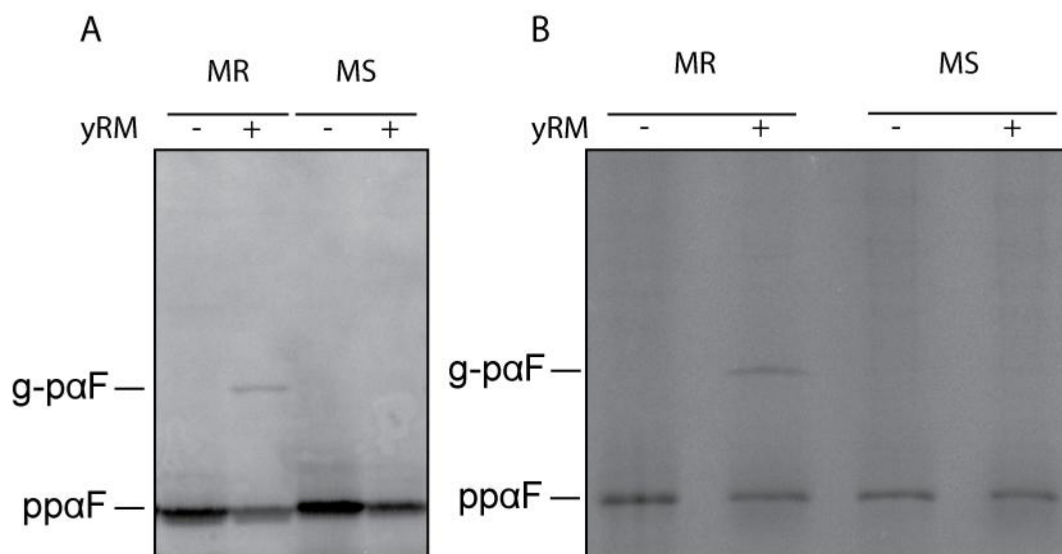


Figure 3.9 Translocation assay for in vitro expressed lysine-less ppαF. Wild-type (MR) and MS ppαF were translated *in vitro* in A) rabbit reticulocyte lysate or B) yeast cytosol, translation was inhibited with cycloheximide followed by incubation in the absence or presence of yeast microsomes (yRM). The position of non-translocated (ppαF) and signal-sequence cleaved, glycosylated (g-pαF) are indicated.

Proceeding with the approach to assess N-terminal acetylation *in vitro*, translations of lysine-less ppαF and ppαF-MS were carried out in rabbit reticulocyte lysate in the presence of [^{14}C] Acetyl-CoA or [^{35}S] methionine. The negative controls for these experiments were translations without the addition of mRNA.

Firstly translations were performed in the presence of [^{35}S] methionine to confirm that both lysine-less ppαF and ppαF-MS were translated at similar levels, shown by both scintillation counting and SDS-PAGE analysis of samples (Figure 3.10A). In contrast, when these proteins were translated in the presence of [^{14}C] Acetyl-CoA we found significant incorporation of [^{14}C] only in the ppαF-MS substrate indicating it to be N-α-acetylated. Scintillation counts for wild-type (MR) ppαF were similar to the no mRNA control (Figure 3.10B). This data shows that the ppαF-MS protein is acetylated. Overall these data show that the bias against certain residues at position 2 of signal sequences is biologically significant, since if the residue occupying this position is one that promotes N-terminal modification by MetAP and/or NATs it may result in a defect in protein translocation.

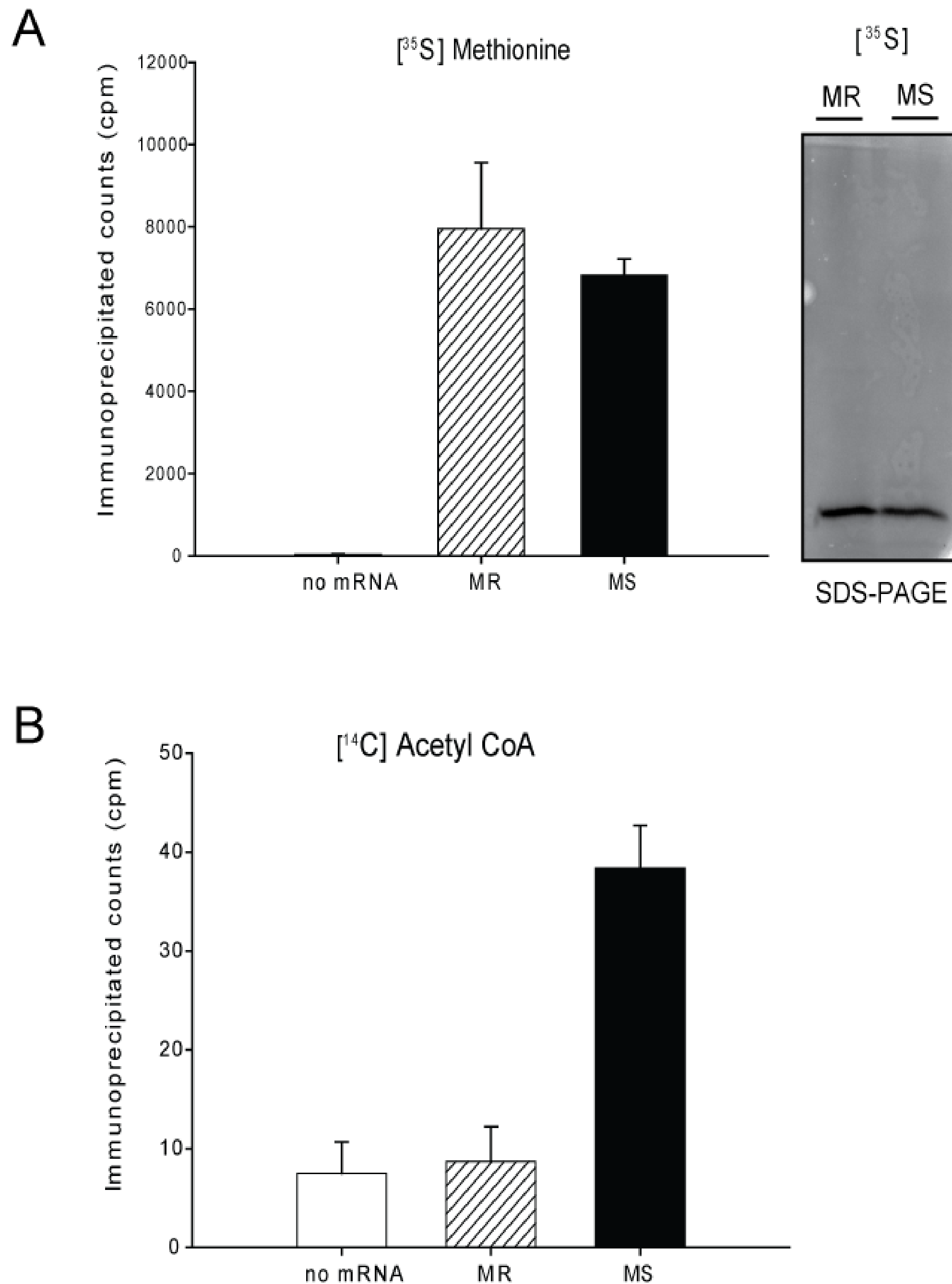


Figure 3.10 Assessing N-α-Acetylation of ppαF. Lysine-less wild-type and MS ppαF were translated *in vitro* in rabbit reticulocyte lysate in the presence of either [^{35}S] Methionine or [^{14}C] Acetyl-CoA and ppαF was immunoprecipitated with anti-ppαF antiserum. Radioactive incorporation was then determined by scintillation counting. To determine background counts, a non-programmed translation reaction (no mRNA) was processed in an identical manner. **A** Scintillation counts of immunoprecipitated [^{35}S] Methionine labelled ppOαF In parallel, the ^{35}S-labelled products were also analysed by SDS-PAGE and phosphorimaging. **B** Scintillation counts of immunoprecipitated [^{14}C] Acetyl-CoA labelled ppαF. Two tailed heteroscedastic student's T-test performed between MR & MS P= < 0.001

3.7 Effect of N-terminal modification on a SRP dependent signal sequence

As discussed in section 1.2 there are two pathways for secretory protein translocation, the post-translational pathway and the co-translational pathway. The latter is utilised by substrates which have more hydrophobic signal sequences. So far this study has focused on the signal sequences of CPY and ppαF both of which are targeted via the post-translational pathway. Therefore, it was decided this investigation should be extended to examine a SRP-dependent substrate to see if N-terminal modification could lead to similar effects to those observed for SRP-independent proteins.

3.7.1 Translocation of position 2 mutants of ppOPY *in vivo*

The SRP-dependent substrate ppOPY chosen for investigation, this is a modified version of ppCPY in which the native signal sequence has been replaced by that of Ost1p (discussed in section 1.2.5). The Ost1p signal sequence has arginine at position 2 and therefore not predicted to be N-terminally processed. Insertion mutations at P2 of the ppOPY signal sequence were made by SDM (Figure 3.11), as was performed in ppCPY.

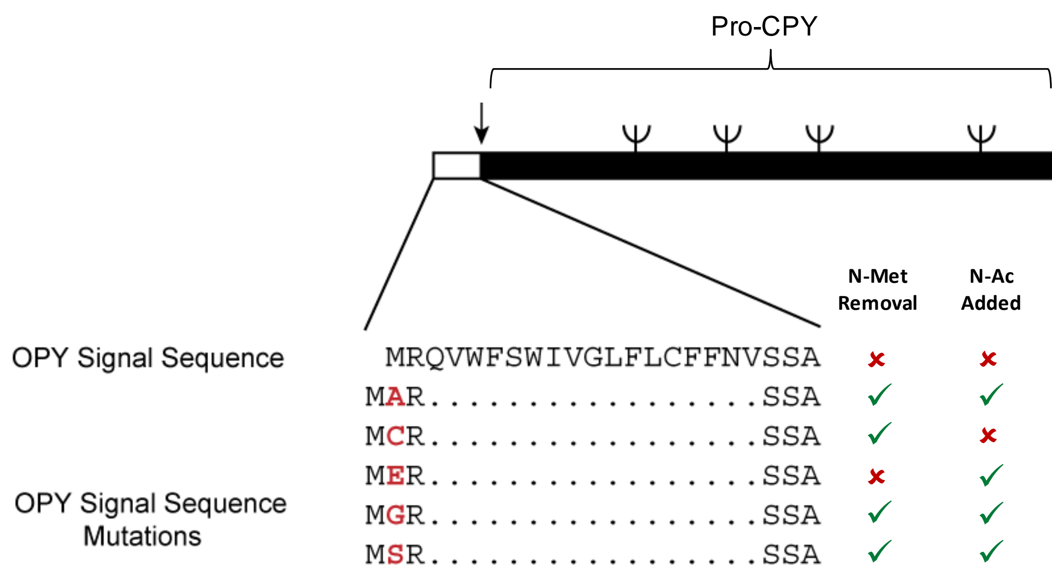


Figure 3.11 ppOPY signal sequence mutation at position 2. Cartoon of ppOPY construct showing the Ost1p signal sequence and mutations of amino acid insertion at position 2, made by SDM. Also shown, a prediction of possible N-terminal modification events based on the residue at P2. The position of signal peptidase cleavage site (↓) and N-glycosylation (ψ) sites in pro-CPY region of the fusion protein are indicated.

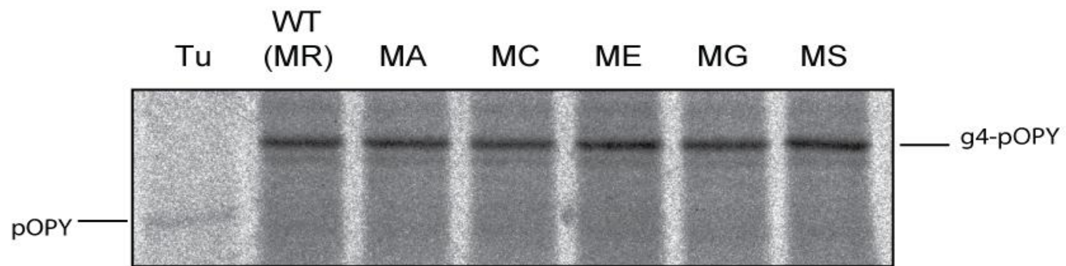


Figure 3.12 Analysis of ppOPY translocation *in vivo*. ppOPY position 2 mutants were analysed in a 'wild-type' (GFY3) yeast strain. Cell cultures were pulse-labelled with [^{35}S] methionine/cysteine, then extracts were immunoprecipited with CPY antiserum and analysed by SDS-PAGE then visualised by phosphorimaging. Position of fully-(four-times, g4-pOPY) glycosylated OPY and wild-type OPY treated with Tunicamycin (Tu) (pOPY) are also indicated.

Pulse-labelling of the various ppOPY position 2 mutants were performed in the yeast strain GFY3 *in vivo.* This was done to assess whether these mutants cause any translocation defect as seen for these mutations in ppCPY (Figure 3.3). Contrary to our findings for ppCPY, introducing these mutations in ppOPY had no affect on translocation indicating that N-terminally processable position 2 residues can be tolerated in a SRP-dependent signal sequence (Figure 3.12). This finding can explained if either, SRP-dependent signal sequences can be N-terminally processed and are still targeted via SRP and subsequently translocate across the ER membrane, or alternatively these substrates are not processed as predicted.

3.7.2 Assessing N-terminal modification of the Ost1p signal sequence *in vitro*

To assess if the Ost1p signal sequence is N-terminally processed a lysine-less substrate ppOα-factor (ppOαF) was made by substituting the signal sequence of the lysine-less ppαF with that of Ost1p or Ost1p mutant signal sequence with a serine position 2 (Figure 3.13A). Making lysine-less substrates avoids the complication of internal acetylation as described in section 3.6.2. To assess the translocation of theses constructs, mRNA encoding wild-type ppOαF-MR and the position 2 serine mutant ppOαF-MS were translated and radio-labelled *in vitro* in the absence or presence of dog microsomes (Figure 3.13B). Both ppOαF-MR and ppOαF-MS are able to translocate into microsomes, as was see for the ppOPY substrates *in vivo* (Figure 3.12).

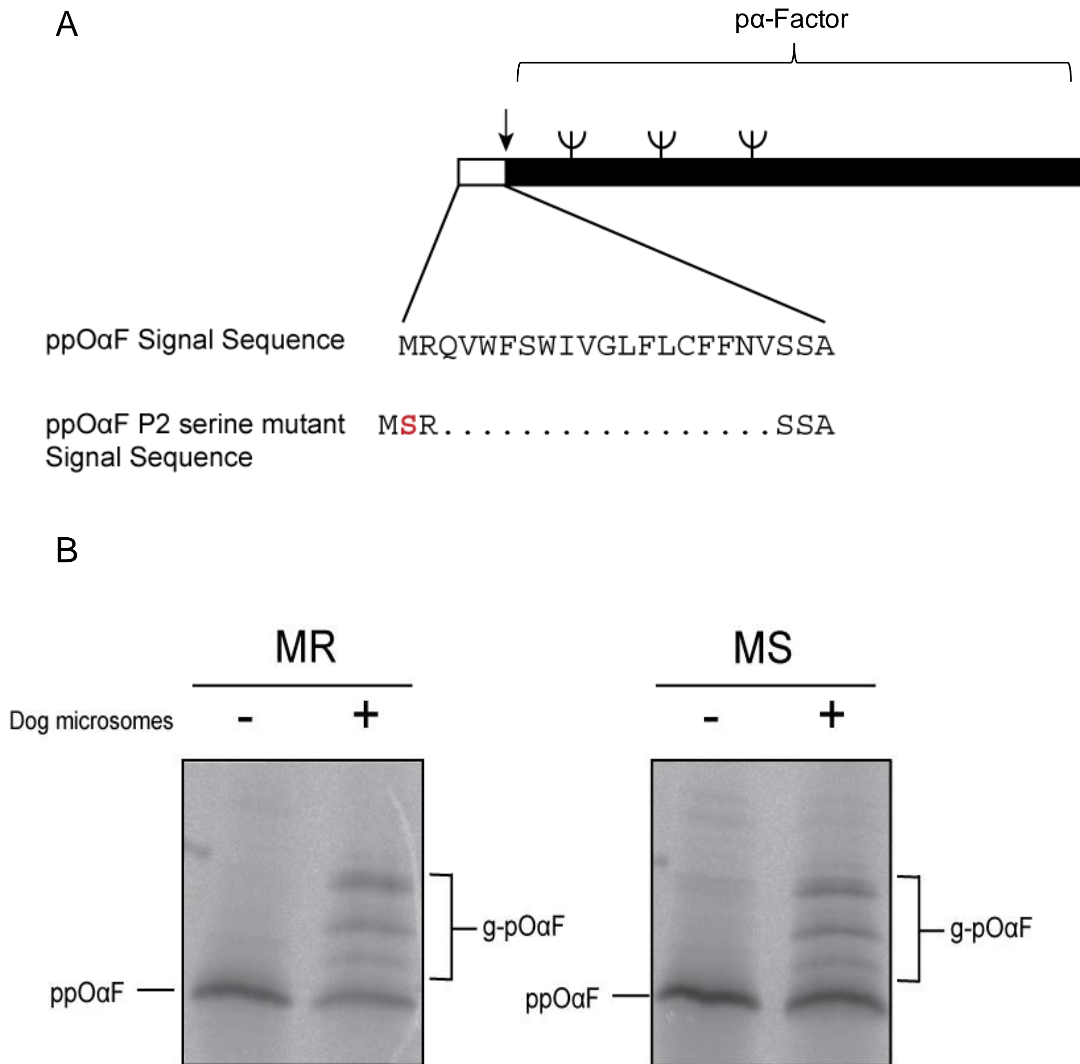


Figure 3.13 Translocation assay for *in vitro* expressed lysine-less ppOαF. A)Wild-type (MR) and P2 serine mutant ppOαF constructs. The position of signal peptidase cleavage site (↓) and N-glycosylation (ψ) sites in pαF region of the fusion protein are indicated. **B)** mRNA transcribed from these constructs was translated *in vitro* in rabbit reticulocyte lysate supplemented with SRP, in the absence or presence of dog microsomes. The position of non-translocated (ppOαF) and signal-sequence cleaved, glycosylated (g-pOαF) are indicated.

To assess N-terminal modification of ppOαF substrates, translations were carried out in the presence of either [^{35}S] methionine or [^{14}C] Acetyl-CoA for scintillation counting as previously performed for ppαF in section 3.6.2. It was found that both MR and MS forms of ppOα-factor were translated at comparable levels, shown by the translation in the presence of [^{35}S] methionine (Figure 3.14A), yet only in the MS form of ppOαF was there any significant incorporation when translated in the presence of [^{14}C] Acetyl-CoA (Figure 3.14B). This demonstrates that MS ppOαF is acetylated. However, if this finding is compared to that observed for the same experiment performed with ppαF (Figure 3.10) it can be seen that the amount of acetylation for ppOαF MS is markedly less than that for ppαF MS, the ppOαF MS having around half the counts per minute. The [^{35}S] methionine scintillation counts show that ppOαF constructs were translated at a similar level to those of ppαF.

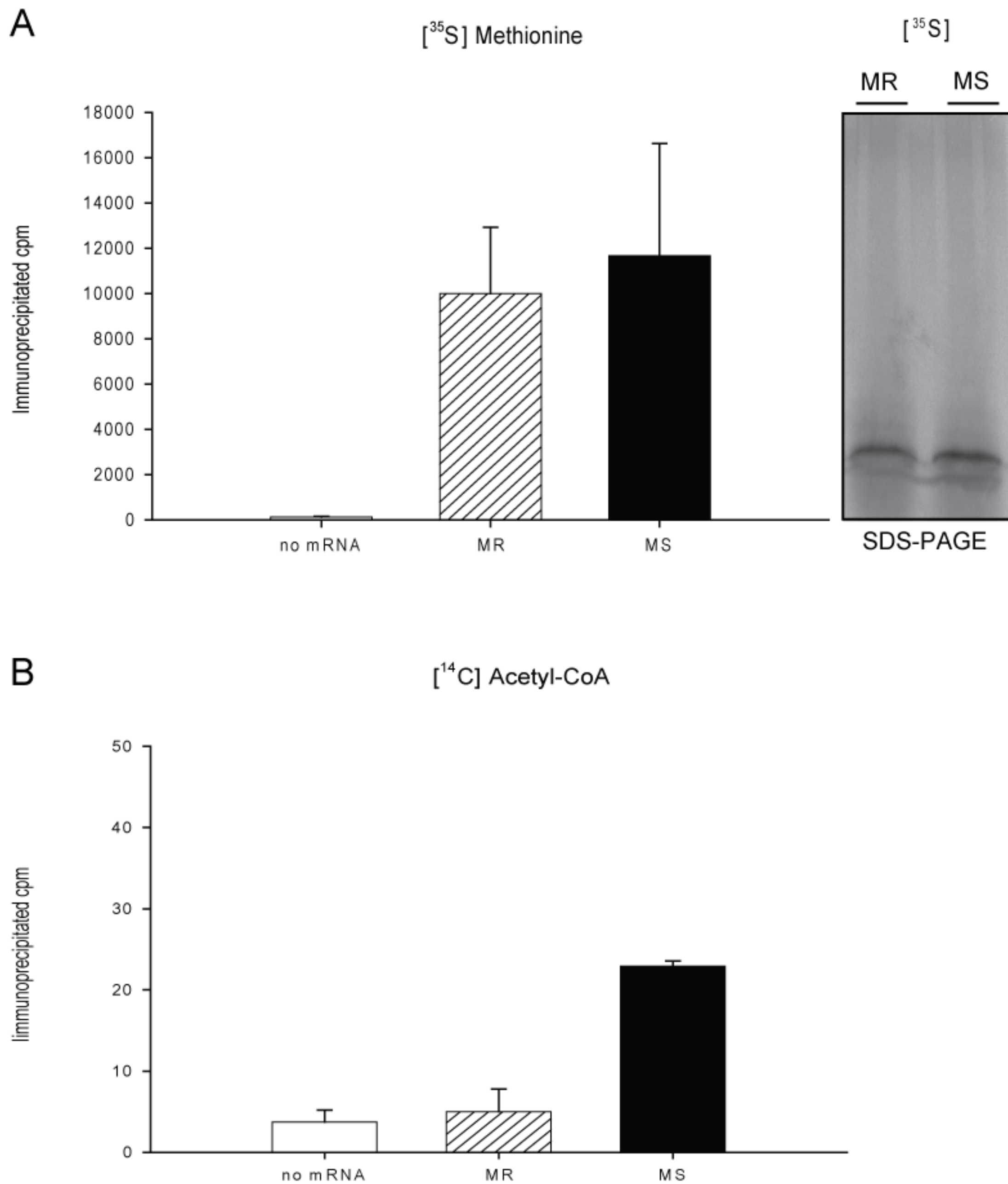


Figure 3.14 Assessing N-α-Acetylation of ppOαF. Lysine-less wild-type and MS ppOαF were translated *in vitro* in rabbit reticulocyte lysate in the presence of either [^{35}S] Methionine or [^{14}C] Acetyl-CoA and ppOαF was immunoprecipitated with anti-ppαF antiserum. Radioactive incorporation was then determined by scintillation counting. To determine background counts a non-programmed translation reaction (no mRNA) was processed in an identical manner. **A)** Scintillation counts of immunoprecipitated [^{35}S] Methionine labelled ppOαF in parallel, the ^{35}S-labelled translation products were also analysed by SDS-PAGE and phosphorimaging. **B)** Scintillation counts of immunoprecipitated [^{14}C] Acetyl-CoA labelled ppOαF. Two tailed heteroscedastic student's T-test performed between MR & MS P= < 0.01

Although this data above indicates ppOαF-MS is N-α-acetylated in the *in vitro* translation experiments, it was reasoned that this experiment was not a fair representation for an SRP targeted substrate, as it would occur *in vivo*. In these experimental conditions the substrate has been fully translated into the cytosol (rabbit reticulocyte lysate) and may have lost some of its interactions with SRP. Normally *in vivo* co-translational substrates would be targeted to the ER membrane as a short polypeptide emerging from a ribosome with SRP known as a ribosome nascent chain complex (RNC). Therefore truncated ppOαF MR and MS mRNA transcripts were made without a stop codon. When these mRNA transcripts are translated the nascent polypeptides are not released from the ribosomes due to the absence of a stop codon, therefore simulating RNCs targeted by SRP. These RNC constructs were 92 amino acids in length and include two methionine residues, one of these being the initiator methionine, and the other at the C-terminus followed by a HA tag epitope for immunoprecipitation. Unfortunately these RNCs were unable to be efficiently immunoprecipitated with either anti-HA or anti-ppαF antiserums (Figure 3.15). Translations of the RNC constructs were performed in the presence of either [^{35}S] Methionine or [^{14}C] Acetyl-CoA for scintillation counting. To remove the unincorporated radio-chemicals, RNCs were isolated by ultra centrifugation through a sucrose cushion. The isolated RNC pellet was then resuspended in water and sequentially treated with sodium hydroxide and Triton-X-100, the proteins were then precipitated with TCA, and these precipitated proteins were then used in scintillation counting experiments.

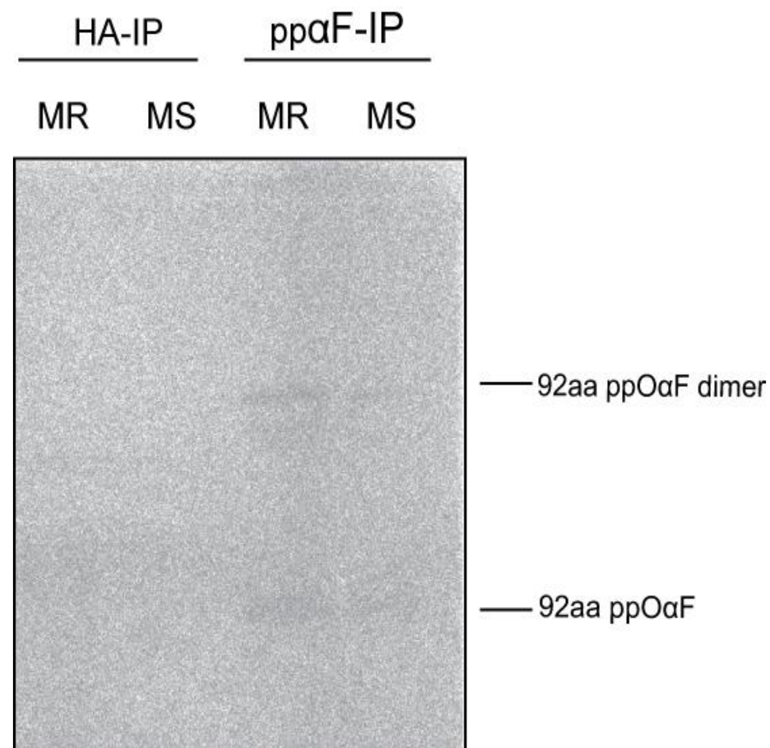


Figure 3.15 Immunoprecipitation of truncated ppOαF. Translations of 92 amino acid truncated ppOαF (wild-type (MR) and MS mutant) in rabbit reticulocyte lysate in the presence of [^{35}S] Methionine. Translated products were immunoprecipitated with either anti-HA or anti-ppαF antiserum and analysed by SDS-PAGE and phosphorimager. Note the appearance of dimerised product is dependent on sample preparation and is also seen with the full length ppOαF

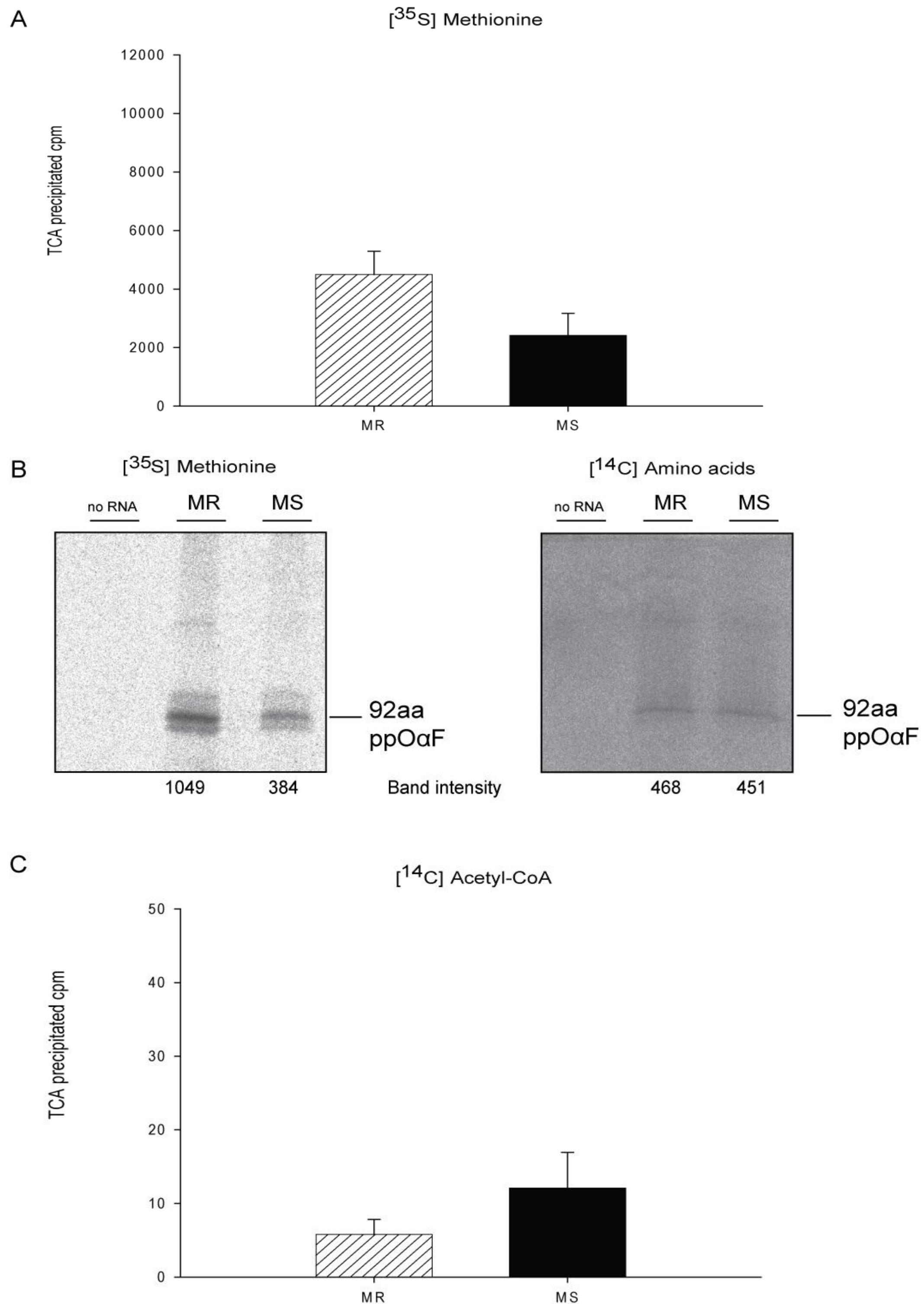


Figure 3.16 Assessing N-α-Acetylation of truncated ppOαF. Lysine-less wild-type and MS truncated ppOαF (92 amino acids) was translated *in vitro* in rabbit reticulocyte lysate in the presence of either [^{35}S] Methionine, [^{14}C] Acetyl-CoA or [^{14}C] Amino acids and RNCs were isolated by centrifugation. Radioactive incorporation was then determined by scintillation counting and SDS-PAGE. To determine background counts a non-programmed translation reaction (no mRNA) was processed in an identical manner. **A** Scintillation counts of TCA precipitated RNCs [^{35}S] Methionine labelled ppOαF Two tailed heteroscedastic student's T-test performed between MR & MS P= <0.05. **B** RNCs translated in the presence of [^{35}S] Methionine and [^{14}C] Amino acids were also analysed by SDS-PAGE and phosphorimaging. **C** Scintillation counts of TCA precipitated [^{14}C] Acetyl-CoA labelled ppOαF.

The RNCs produced by translation with truncated transcripts in the presence of [^{35}S] methionine showed a marked difference in scintillation counts between the MR and MS ppOαF (Figure 3.16A), with counts for the MS mutant being roughly half that of MR. one reason for this difference could be that the initiator methionine of MS ppOαF is removed by MetAPs, another reason may be that the MR ppOαF translated more efficiently. To verify which, ppOαF RNC constructs were translated in the presence of [^{14}C] amino acid mixture without methionine to confirm their translation efficiency. It was observed that both substrates incorporated similar amounts of [^{14}C] amino acid mixture when we quantified the bands using Aida image analysis software, showing that they are translated with equal efficiency (Figure 3.16B). The RNCs constructs were translated in the presence of [^{14}C] Acetyl-CoA to ascertain if the MS substrate would be N-α-acetylated after removal of the initiator methionine. This data indicates that there is no significant difference in [^{14}C]-acetyl incorporation between MR and MS ppOαF, and therefore MS is unlikely to be N-α-acetylated (Figure 3.16C).

Another way to assess whether the Ost1p MS mutant signal sequence is N-terminally processed was by testing for the presence of free α-NH_2 groups using a biotinylation assay. This assay uses sulfo-NHS-SS-biotin cross-linking reagent to modify free α-NH_2 groups attaching biotin which allows extraction of the modified protein via immobilized-streptavidin. This assay was performed using lysine-less wild-type (MR) and MS ppOαF translated *in vitro* in the presence of [35S] methionine. The data (Figure 3.17 (manuscript submitted Forte *et al*)) show both wild type (MR) and MS mutant ppOαF can be isolated from the extracts using immobilized-streptavidin in equivalent proportions. This indicates that both these proteins have free N-termini. These data suggest that in the context of an SRP-dependent signal sequence, the MS N-terminus was not acetylated.

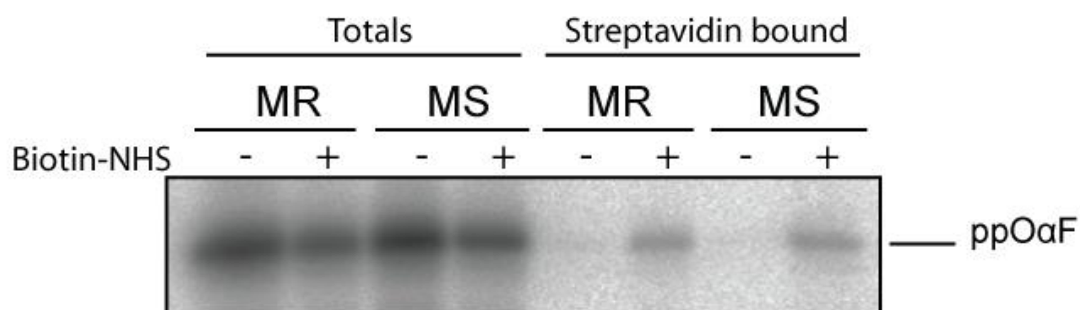


Figure 3.17 Biotinylation of free N-termini of ppOαF. (Figure provided by Dr M Pool) Lysine-less wild-type (MR) and MS ppOαF were translated *in vitro* in rabbit reticulocyte lysate in the presence of [35S] methionine, denatured and modified with sulfo-NHS-SS-biotin. Biotinylated proteins were re-isolated on immobilized-streptavidin and visualised by SDS-PAGE and phosphorimaging.

In summary these data show that changing the signal sequence position 2 residues to one predicted to promote N-terminal modification, has no effect on the translocation of an SRP dependent substrate *in vivo*. Further to this when N-terminal modification events are examined *in vitro*, it was found that N-terminal methionine removal may occur but that it is unlikely that these SRP-dependent substrates are N-α-acetylated.

Discussion

4. Discussion

In the introduction to this study it was noted that secretory proteins with signal sequences have a striking bias against residues at position 2 which promote N-terminal modification. This phenomenon was investigated both *in vivo* and *in vitro* to assess whether it had any biological significance.

In the data presented it is shown that changing the second residue of the ppCPY signal sequence to one which can be modified by MetAPs and/or NATs, can cause defective translocation of proteins to varying extents *in vivo*. Additional analysis of the translocation defects observed for the MA, MC, MG and MS mutants of ppCPY demonstrates that this defect seemed to be dependent on the activity of MetAPs. This conclusion can be justified as when the wild-type strain (*MAP1/MAP2*) was treated with Fumagillin, which inactivates Map2p only, Map1p compensates for the inhibited Map2p and the defect persisted. Likewise, the defect is also seen in a *Δmap1/Map2* strain, where Map2p which can compensate to some extent for the loss of Map1p (Chen et al., 2002, Li and Chang, 1995). Only when the *Δmap1/MAP2* strain is treated with Fumagillin, making the cells totally defective in MetAP activity, is translocation restored. This restorative effect correlates with the predicted specificity of MetAPs. While initiator methionine removal is a determinant of defective translocation, it is clearly not the sole factor. This is shown by the ME mutant of ppCPY which exhibits defective translocation, but it is not predicted to be a substrate for MetAPs. Furthermore, the translocation defect of the ME mutant is not restored by blocking all MetAP activity.

Instead, the translocation defect observed for the ME mutant of ppCPY was attributed to N-α-acetylation of the initiator methionine by NatB, as translocation was only restored by deletion of the *nat3* gene which encodes the catalytic subunit of this complex. Although the translocation defect of the MS mutant was shown to be influenced by MetAP activity, it was predicted that this substrate could also be N-α-acetylated by NatA on the P2 serine residue after removal of the initiator methionine by MetAP. It was subsequently shown that the translocation of the MS mutant could also be restored if the gene which encodes the catalytic subunit of the NatA complex was deleted (*Δard1*). This suggests that the observed phenotype is due in the first instance to N-terminal methionine removal which subsequently leads to N-α-

acetylation of the exposed P2 serine residue. These findings strongly suggest that N-α-acetylation is the major cause of the translocation defect seen, and it may be that initiator methionine removal is a prelude to this. It is unclear whether initiator methionine removal alone can cause defective translocation. To test this, it would be necessary to use a substrate which has a P2 residue which promotes N-terminal methionine removal but cannot be N-α-acetylated. Such a mutant may not exist as various studies show that any P2 residue which promotes initiator methionine removal can potentially lead to subsequent N-α-acetylation of this P2 residue (Polevoda et al., 1999, Tsunasawa et al., 1985) the exceptions being valine and cysteine for which there is conflicting evidence as to how likely they are to be N-α-acetylated. In any case, this may be minor point as it is clear from the data for the ME mutant of ppCPY that N-α-acetylation alone could be enough to cause defective translocation.

This theory was supported by *in vitro* analysis of a different substrate, ppαF. Changing the signal sequence P2 to serine (ppαF-MS) also caused a defect in translocation. Further investigation proceeded to demonstrate the occurrence of N-α-acetylation on ppαF-MS, by direct transfer of an acetyl group to this substrate from Acetyl Coenzyme A.

In the initial data set in Figure 3.3, there are P2 mutants of ppCPY which were not investigated further. The ppCPY signal sequence mutant with arginine (MR) as the second residue is additional control as having this amino acid at P2 has not been reported to lead to any N-terminal modification events. This mutation in the signal sequence did not have an effect on the translocation efficiency of ppCPY compared to the native protein (MK).

The ppCPY MV mutant is an exception, as it is predicted to be N-terminally modified but its translocation is comparable to the wild-type protein. An explanation for this is that a substrate with valine at P2 is not always modified as predicted and its initiator methionine may remain in place. MetAPs have specificity for small residues at P2, valine is one of the larger of those accepted at this position and does not always lead to efficient excision of the initiator methionine (Frottin et al., 2006). As discussed previously it is also unclear if subsequent N-α-acetylation follows N-

terminal methionine excision on a valine residue. This could be tested by Edman degradation and/ or mass spectrometry analysis.
In contrast, the MH and MQ mutants of ppCPY, exhibit translocation defects although these residues at P2 are not commonly believed to result in N-terminal modification. A recent study however, has suggested that MH can be N-α-acetylated by NatC (Martinez et al., 2008). The frequency with which histidine appears at P2 of signal sequences is relatively low. However, glutamine is common at P2 of yeast signal sequences (Figure 1.12) so it is unclear why this residue would affect ppCPY translocation. Analysing the properties of histidine and glutamine residues (M.J. Betts, 2003), and their predicted effect on the integrity of the ppCPY signal sequence using Expasy tool SignalP, there are no obvious implications for these mutations. While glutamine is common P2 of signal sequences, its tolerance may dependent upon the context of the signal sequence and/or its targeting pathway.

Many of the P2 signal sequence mutants of ppCPY exhibit an exacerbated under glycosylation (g3-pCPY) phenotype, in addition to a translocation defect. Some of this g3-pCPY form is seen for the wild-type protein. ppCPY is known to have four glycosylation acceptor sites located throughout the mature region of the protein, all of which have the motif Asn-Xaa-Thr. The most likely cause of the g3-pCPY form seen for the wild-type protein may be the due to probability of glycosylation of the C-terminal acceptor site. The Expasy tool NetNGlyc (http://expasy.org/tools/) predicts that this C-terminal glycosylation site has a reduced glycosylation potential (potential = 0.39) compared to the other three sites (potentials = 0.69-0.74), and may be less efficiently glycosylated as it is below the glycosylation threshold (potential ≥ 0.5). This may be partially due to the fact that the further away a Asn-Xaa-Thr glycosylation motif is from the N-terminus, reduces the frequency of its glycosylation especially if this motif is less than 60 residues from C-terminus of the protein (Nilsson and von Heijne, 2000), as is the case for the C-terminal glycosylation site of ppCPY. However, many of the P2 mutants of ppCPY show an exacerbated g3-pCPY phenotype. It is unclear how the addition of an extra residue in the signal sequence could affect the glycosylation of the mature region of the translocated protein. Therefore it is only possible to speculate that the extension of the signal sequence by one residue affects the efficiency of ppCPY glycosylation.

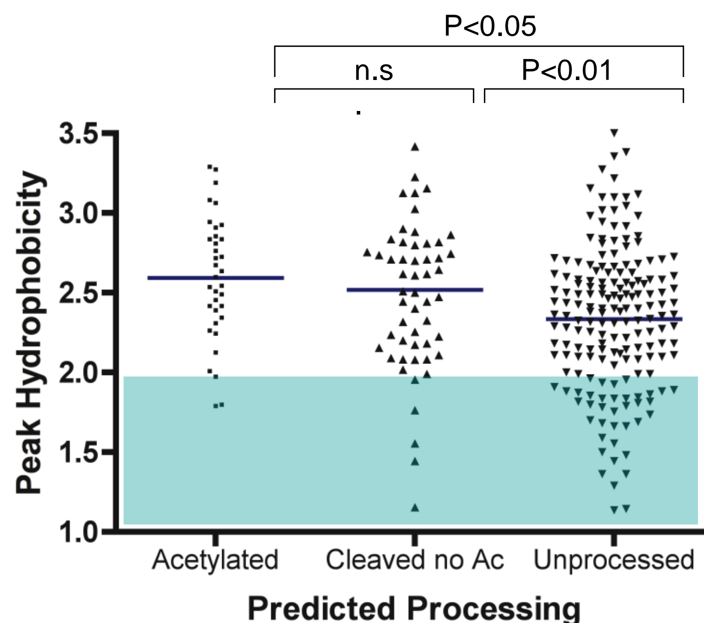
Since there are two distinct translocation pathways in *S. cerevisiae*, and both ppCPY and ppαF are translocated via the SRP-independent pathway (Ng et al., 1996), it was necessary to consider an SRP-dependent substrate. The signal sequence of Ost1p was used for this purpose in the form of the substrates ppOPY and ppOαF. Insertion mutations were made in ppOPY comparable those which caused translocation defects in ppCPY. However, none of the mutations made had any effect on the translocation of ppOPY *in vivo*, or ppOαF *in vitro*. On further investigation of N-terminal modification of ppOαF *in vitro*, it was found that if a serine was introduced at P2 initiator methionine removal would occur but that the serine unlikely to be subsequently N-α-acetylated as shown previously for ppαF-MS.

Despite both SRP-independent and SRP-dependent substrates having signal sequences, the data presented suggest that only proteins targeted via the SRP-independent pathway are susceptible to N-terminal acetylation and the subsequent defect in translocation that this causes. The reason for this is likely to be due to the differences in their mechanism of translocation and also how these differences relate to the coincidence of N-terminal modification events. MetAPs, NatA and SRP are all associated with or close to the exit site of translating ribosomes (reviewed in Kramer et al., 2009). Map1p interacts with nascent polypeptides of 45-50 amino acids in length (Jackson and Hunter, 1970) whilst SRP interacts with those of around 50-70 residues (Kurzchalia et al., 1986). NatA is believed to interact with longer nascent polypeptides (Gautschi et al., 2003). Unlike SRP, NatA contacts a variety of emerging polypeptides, as shown by cross-linking experiment it contacts polypeptide regardless of whether they are substrates (Gautschi et al., 2003). Increased amounts of SRP are recruited to ribosomes which are translating polypeptides with a SRP-dependent signal sequence (Raue et al., 2007). There is some evidence that SRP may cause some element of steric hindrance for NatA interaction with polypeptides that are SRP substrates (Raue et al., 2007), as cross-linking experiments revealed NatA can contact other emerging polypeptides but not the SRP dependent substrate Dap2p (Raue et al., 2007).

One other important difference between the SRP-dependent and independent translocation of polypeptides is that, when SRP attaches itself to its very short substrate, translation is stalled until the translating ribosome docks on a translocon at

the ER membrane (Walter and Blobel, 1981). In contrast, SRP-independent substrates are almost entirely translated before they are translocated, and therefore have a prolonged exposure to the cytosol, where there is increased opportunity for factors such as MetAPs and NATs to act. Also it is worth noting that although it is known that chaperones bind to SRP-independent substrates, little is known about how quickly or to which parts of the nascent protein they interact compared to SRP and its substrates, and therefore chaperones that maintain SRP-independent substrates in a translocation competent conformation may not mask the signal sequence from N-terminal modifying factors as efficiently as SRP does for its substrates.

Whilst the bioinformatics data in section 1.5 show a strong predicted bias against signal sequences being N-terminally modified, there is also a significant portion, approximately 30%, which are predicted to be modified. This could be explained if these signal sequences were of SRP dependent substrates, as the experimental data indicates these are unlikely to be modified as predicted. To determine whether as a substrate is likely to be SRP dependent, the signal sequence can be analysed in terms of peak hydrophobicity, as SRP dependent signal sequences tend to have a peak hydrophobicity > +2 , compared to those which belong to substrates that translocate independently of SRP which generally have a peak hydrophobicity < +2. The signal sequences of *S. cerevisiae* were analysed both in terms of peak hydrophobicity and whether or not they are predicted to be N-α-acetylated or have their initiator methionine cleaved (Figure 4.1, (manuscript submitted Forte *et al*)). These data clearly show that those signal sequences which are predicted to be N-terminally modified have a low proportion with a peak hydrophobicity <+2 while the mean peak hydrophobicity is significantly higher than for signal sequences which are predicted to be unprocessed. The majority of signal sequences which are predicted to be N-terminally modified are therefore likely to be SRP dependent, and contrary to predictions are unlikely to be N-α-acetylated *in vivo*.



	n	Mean ± SEM	ANOVA	h<2 n	h<2 %
Ac	35	2.593 ±0.0657	P>0.05	3	8
Cleaved no Ac	55	2.518 ±0.0673	P<0.05	6	11
Unprocessed	175	2.333 ±0.0352	P<0.01	43	25

Figure 4.1 Combined analysis of signal sequence peak hydrophobicity and predicted N-terminal processing (provided courtesy of Dr M Pool). For each signal sequence of *S. cerevisiae* N-terminal modification was predicted and these were then grouped accordingly unprocessed, N-terminal methionine cleaved only or N-α-acetylated. For each group peak hydrophobicity (Kyte and Doolittle, 1982) of each signal sequence and the mean for each group was determined and plotted, also ANOVA analysis was performed to compare groups. SRP independent signal sequences generally have peak a hydrophobicity less than +2 and this area of the graph is shade blue. For complete data sets see appendix

The hypothesis that *S. cerevisiae* signal sequences may be subject to a bias against N-terminal methionine cleavage (Figure 1.12) can be extended beyond yeast to higher eukaryotes (Figure 4.2 (manuscript submitted Forte *et al*)). This analysis reveals that the signal sequences of higher eukaryotes also have a bias to retain an unprocessed initiator methionine. These data show that *C. elegans* has a similar trend to *S. cerevisiae*, with almost 70% of signal sequences predicted to retain their initiating methionine compared to ~20% proteome. In humans the bias is less striking, with ~50% of signal sequences compared to ~15% of the proteome predicted to have an unprocessed methionine at their N-termini. The reduced bias against N-terminal modification in human signal sequences could be due to a majority of the substrates identified to date having been assigned as being SRP-

dependent, whilst there has been speculation of a SRP-independent pathway in mammals no clear pathway has been identified.

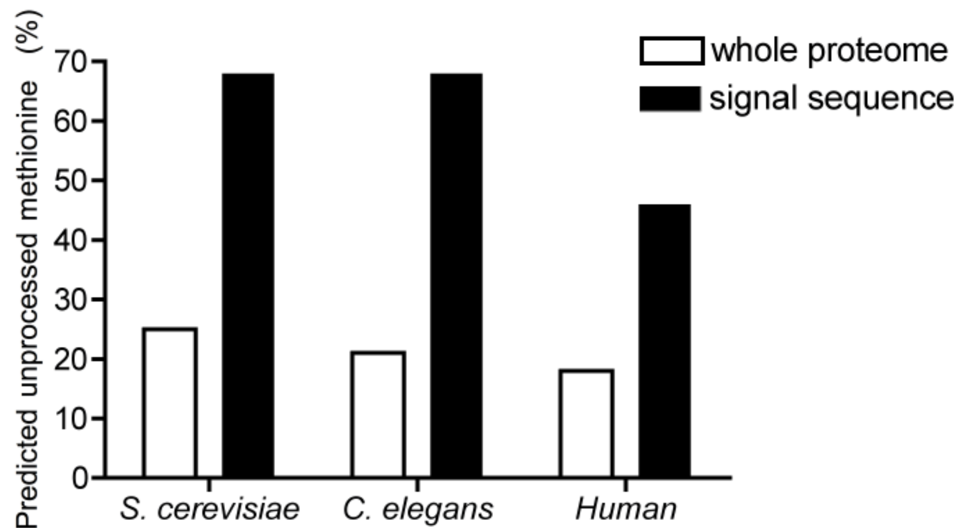


Figure 4.2 Bias for retention of an unprocessed methionine appears to be conserved in the signal sequences of higher eukaryotes. (Provided courtesy of Dr M Pool). Predicted frequency of an unprocessed initiating methionine in signal sequences from *S. cerevisiae* (n=277), *C. elegans* (n=378) and human (n=595), compared to the respective proteomes as a whole (Martinez et al., 2008). For complete data sets see appendix.

The secretory proteins are not the only group targeted from the cytosol to a subcellular localisation; a large proportion of mitochondrial proteins that are encoded by the nuclear genome and are also targeted from the cytosol to the mitochondria via their N-terminal presequences (Schatz and Dobberstein, 1996). When the N-termini of this group of mitochondrial proteins of *S. cerevisiae* were analysed bioinformatically (Figure 4.3 (manuscript submitted Forte *et al*)), a strong bias for these proteins to retain their initiator methionine was observed. This indicates that retention of the N-terminus of both secretory and mitochondrial proteins may be an important feature of proteins which are targeted out of the cytosol via sequences at their N-termini.

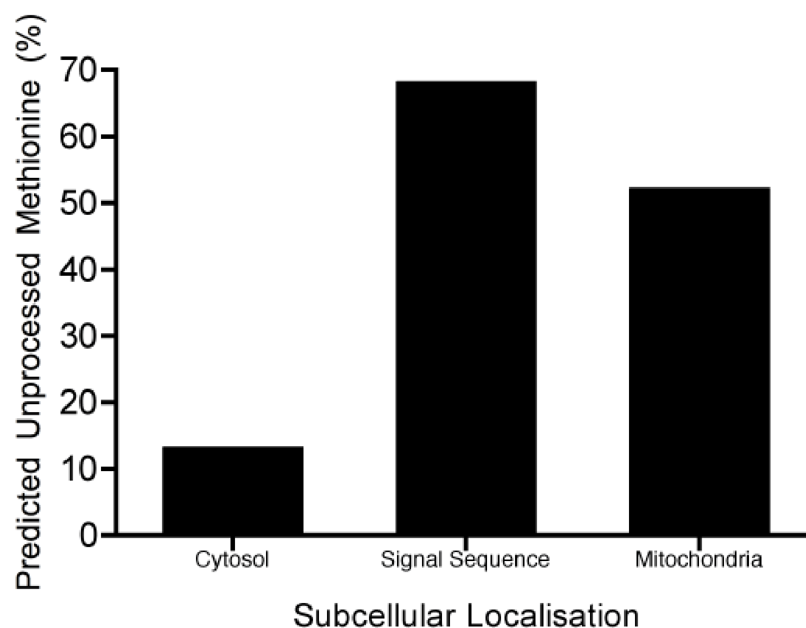


Figure 4.3 Prediction retention of N-terminal methionine in the subcellular localisations of *S. cerevisiae*. (Provided courtesy of Dr M Pool) Predicted frequency of unprocessed N-terminal methionine retention of cytosolic (n=277), signal sequence-containing (n=252), and mitochondrial (n=281) proteins from *S. cerevisiae*. For complete data sets see appendix.

This study has clearly established that N-α-acetylation can alter the fate of secretory proteins which are usually translocated. A major question which remains is how exactly N-α-acetylation causes this phenomenon? Many functional properties have been assigned to the N-α-acetylation of proteins, making it difficult to decipher which of these is the cause of the translocation defect seen for the mutated secretory substrates. To identify how N-α-acetylation could cause defective translocation would require extensive further experimentation. However, some of the possibilities are discussed below.

Recent works by Hwang *et al*, have shown that N-α-acetylation of a protein can act as a signal for its degradation (Hwang et al., 2010), this is an unlikely possibility in this study as this would mean the proteins would be rapidly turned over and hence unseen in the experiments performed. Another possibility is that the addition of an N-terminal group could mask the signal sequences from factors that recognise it and aid the translocation of the protein. Also the physical properties of the acetyl group could hinder the passage of a polypeptide through the translocon, it is not known if

an N-α-acetylated secretory protein can translocate. A study of Peroxiredoxin II (Seo et al., 2009) offers an interesting explanation, this work shows that N-α-acetylation of a protein can cause changes in its conformation, this is interesting as it could cause the aforementioned masking of the signal sequences.

4.1 Further experimentation and studies

The data presented in this study have revealed important processes which may influence the processes required for protein translocation to the ER. However, there still remain many significant questions which could be resolved by further investigation.

One important matter to resolve is whether the defective translocation seen for any of the other ppCPY mutants (MA, MG and MC) can be attributed to N-α-acetylation. Although the Termi*N*ator prediction tool predicted these substrates are unlikely to be N-α-acetylated (Table 3.1), other studies have shown these could be substrates for NatA after initiator methionine excision (Polevoda and Sherman, 2003b). To resolve this would involve pulse-labelling of these constructs in a *Δard1* strain to see if translocation is restored, as was performed for the MS mutant. In addition the outcome to this experiment could also determine whether N-terminal methionine excision alone can cause a defect in translocation.

One of the major findings of this study is that N-α-acetylation may affect the translocation of secretory proteins, but the underlying mechanism has not been ascertained. To determine if the mutated proteins that show translocation defects are still able to interact with targeting factors and/or translocon components, cross-linking and co-immunoprecipitation experiments could be performed in comparison to the wild-type protein. If the proteins could be efficiently purified and retain their structure, comparative studies of the conformation of the native and N-α-acetylated mutant form of the protein could be carried out to determine the extent of any conformational change due to the acetyl group.

The bioinformatics data could be used to find candidate native secretory proteins that are strongly predicted to be N-α-acetylated, in particular those predicted or shown to be SRP-independently translocated, Mass Spectrometry peptide analysis of such proteins could be carried out in a wild-type and appropriate *nat* mutant strains to

determine their N-α-acetylation status. This would determine if there are N-α-acetylated native proteins that can be translocated into the ER. Likewise, it would be important to determine N-α-acetylation status of the mutant proteins which show a translocation defect *in vivo*. Again, the most conclusive method to determine this would be via Mass Spectrometry peptide analysis of these proteins.

The analysis of the effects on N-terminal modification could be extended to mitochondrial proteins with presequences. Performing experiments similar to those for secretory proteins in this study, would give valuable insight into the extent of the N-terminal modification bias and its wider significance.

4.2 Concluding remarks

Proteins can be subject to a number of modifications which can influence their function/localisation. N-terminal modification represents the earliest possible opportunity in a protein's biogenesis for modification, occurring as the N-terminus of a nascent polypeptide emerges from the ribosome. This study has shown that there is a clear difference between the predicted occurrences of N-terminal modification in cytosolic and secretory proteins, the latter having a strong observed bias against these modification events. Experimental data in this study shows that when the N-terminus of normally unmodified secretory proteins are altered to promote N-α-acetylation their translocation to the ER is inhibited, providing an explanation for the predicted bias observed in these proteins. As discussed above there are a number of ways in which N-α-acetylation may inhibit the translocation process. One can speculate based on knowledge of the nature of signal sequences, along with known effects N-α-acetylation has on other proteins, that it is possible that acetylation of a signal sequence may change its conformation or affect its interaction with cytosolic targeting factors. As secretory proteins need to be maintained in an unfolded state to translocate, such influences could have a detrimental effect on the ability of these proteins to translocate. The data presented in this study may therefore show a novel sorting step for proteins targeted to subcellular locations via their N-terminal sequence. N-terminal modification may therefore represent a commitment for a protein to undergo immediate folding or conformational change and remain in the cytosol.

References

ADAMS, J. M. & CAPECCHI, M. R. (1966) N-formylmethionyl-sRNA as the initiator of protein synthesis. *Proc Natl Acad Sci U S A,* 55**,** 147-55.

ADDLAGATTA, A. & MATTHEWS, B. W. (2006) Structure of the angiogenesis inhibitor ovalicin bound to its noncognate target, human Type 1 methionine aminopeptidase. *Protein Sci,* 15**,** 1842-8.

ARFIN, S. M. & BRADSHAW, R. A. (1988) Cotranslational processing and protein turnover in eukaryotic cells. *Biochemistry,* 27**,** 7979-84.

ARFIN, S. M., KENDALL, R. L., HALL, L., WEAVER, L. H., STEWART, A. E., MATTHEWS, B. W. & BRADSHAW, R. A. (1995) Eukaryotic methionyl aminopeptidases: two classes of cobalt-dependent enzymes. *Proc Natl Acad Sci U S A,* 92**,** 7714-8.

ARNESEN, T., VAN DAMME, P., POLEVODA, B., HELSENS, K., EVJENTH, R., COLAERT, N., VARHAUG, J. E., VANDEKERCKHOVE, J., LILLEHAUG, J. R., SHERMAN, F. & GEVAERT, K. (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proc Natl Acad Sci U S A,* 106**,** 8157-62.

ARNOLD, R. J., POLEVODA, B., REILLY, J. P. & SHERMAN, F. (1999) The action of N-terminal acetyltransferases on yeast ribosomal proteins. *J Biol Chem,* 274**,** 37035-40.

BACHER, G., LUTCKE, H., JUNGNICKEL, B., RAPOPORT, T. A. & DOBBERSTEIN, B. (1996) Regulation by the ribosome of the GTPase of the signal-recognition particle during protein targeting. *Nature,* 381**,** 248-51.

BACHMAIR, A., FINLEY, D. & VARSHAVSKY, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. *Science,* 234**,** 179-86.

BALLOU, L., HERNANDEZ, L. M., ALVARADO, E. & BALLOU, C. E. (1990) Revision of the oligosaccharide structures of yeast carboxypeptidase Y. *Proc Natl Acad Sci U S A,* 87**,** 3368-72.

BECKMANN, R., BUBECK, D., GRASSUCCI, R., PENCZEK, P., VERSCHOOR, A., BLOBEL, G. & FRANK, J. (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science,* 278**,** 2123-6.

BEN-BASSAT, A., BAUER, K., CHANG, S. Y., MYAMBO, K., BOOSMAN, A. & CHANG, S. (1987) Processing of the initiation methionine from proteins: properties of the Escherichia coli methionine aminopeptidase and its gene structure. *J Bacteriol,* 169**,** 751-7.

BERNDT, U., OELLERER, S., ZHANG, Y., JOHNSON, A. E. & ROSPERT, S. (2009) A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc Natl Acad Sci U S A,* 106**,** 1398-403.

BLOBEL, G. & DOBBERSTEIN, B. (1975) Transfer to proteins across membranes. II. Reconstitution of functional rough microsomes from heterologous components. *J Cell Biol,* 67**,** 852-62.

BOISSEL, J. P., KASPER, T. J. & BUNN, H. F. (1988) Cotranslational amino-terminal processing of cytosolic proteins. Cell-free expression of site-directed mutants of human hemoglobin. *J Biol Chem,* 263**,** 8443-9.

BOUTIN, J. A. (1997) Myristoylation. *Cell Signal,* 9**,** 15-35.

CHANG, S. Y., MCGARY, E. C. & CHANG, S. (1989) Methionine aminopeptidase gene of Escherichia coli is essential for cell growth. *J Bacteriol,* 171**,** 4071-2.

CHANG, Y. H., TEICHERT, U. & SMITH, J. A. (1990) Purification and characterization of a methionine aminopeptidase from Saccharomyces cerevisiae. *J Biol Chem,* 265**,** 19892-7.

CHEN, D. C., YANG, B. C. & KUO, T. T. (1992) One-step transformation of yeast in stationary phase. *Curr Genet,* 21**,** 83-4.

CHEN, L. L., LI, J., LI, J. Y., LUO, Q. L., MAO, W. F., SHEN, Q., NAN, F. J. & YE, Q. Z. (2004) Type I methionine aminopeptidase from Saccharomyces cerevisiae is a potential target for antifungal drug screening. *Acta Pharmacol Sin,* 25**,** 907-14.

CHEN, S., VETRO, J. A. & CHANG, Y. H. (2002) The specificity in vivo of two distinct methionine aminopeptidases in Saccharomyces cerevisiae. *Arch Biochem Biophys,* 398**,** 87-93.

CHIRICO, W. J., WATERS, M. G. & BLOBEL, G. (1988) 70K heat shock related proteins stimulate protein translocation into microsomes. *Nature,* 332**,** 805-10.

CHOO, K. H., TAN, T. W. & RANGANATHAN, S. (2005) SPdb--a signal peptide database. *BMC Bioinformatics,* 6**,** 249.

COLOMBO, S., LONGHI, R., ALCARO, S., ORTUSO, F., SPROCATI, T., FLORA, A. & BORGESE, N. (2005) N-myristoylation determines dual targeting of mammalian NADH-cytochrome b5 reductase to ER and mitochondrial outer membranes by a mechanism of kinetic partitioning. *J Cell Biol,* 168**,** 735-45.

CONNOLLY, T. & GILMORE, R. (1993) GTP hydrolysis by complexes of the signal recognition particle and the signal recognition particle receptor. *J Cell Biol,* 123**,** 799-807.

CYR, D. M., LU, X. & DOUGLAS, M. G. (1992) Regulation of Hsp70 function by a eukaryotic DnaJ homolog. *J Biol Chem,* 267**,** 20927-31.

DESHAIES, R. J., KOCH, B. D., WERNER-WASHBURNE, M., CRAIG, E. A. & SCHEKMAN, R. (1988) A subfamily of stress proteins facilitates translocation of secretory and mitochondrial precursor polypeptides. *Nature,* 332**,** 800-5.

DESHAIES, R. J., SANDERS, S. L., FELDHEIM, D. A. & SCHEKMAN, R. (1991) Assembly of yeast Sec proteins involved in translocation into the endoplasmic reticulum into a membrane-bound multisubunit complex. *Nature,* 349**,** 806-8.

DESHAIES, R. J. & SCHEKMAN, R. (1990) Structural and functional dissection of Sec62p, a membrane-bound component of the yeast endoplasmic reticulum protein import machinery. *Mol Cell Biol,* 10**,** 6024-35.

DRIESSEN, H. P., DE JONG, W. W., TESSER, G. I. & BLOEMENDAL, H. (1985) The mechanism of N-terminal acetylation of proteins. *CRC Crit Rev Biochem,* 18**,** 281-325.

DUMMITT, B., MICKA, W. S. & CHANG, Y. H. (2003) N-terminal methionine removal and methionine metabolism in Saccharomyces cerevisiae. *J Cell Biochem,* 89**,** 964-74.

FELDHEIM, D., ROTHBLATT, J. & SCHEKMAN, R. (1992) Topology and functional domains of Sec63p, an endoplasmic reticulum membrane protein required for secretory protein translocation. *Mol Cell Biol,* 12**,** 3288-96.

FINKE, K., PLATH, K., PANZNER, S., PREHN, S., RAPOPORT, T. A., HARTMANN, E. & SOMMER, T. (1996) A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of S. cerevisiae. *EMBO J,* 15**,** 1482-94.

FLINTA, C., PERSSON, B., JORNVALL, H. & VON HEIJNE, G. (1986) Sequence determinants of cytosolic N-terminal protein processing. *Eur J Biochem,* 154**,** 193-6.

FROTTIN, F., MARTINEZ, A., PEYNOT, P., MITRA, S., HOLZ, R. C., GIGLIONE, C. & MEINNEL, T. (2006) The proteomics of N-terminal methionine cleavage. *Mol Cell Proteomics,* 5**,** 2336-49.

GAUTSCHI, M., JUST, S., MUN, A., ROSS, S., RUCKNAGEL, P., DUBAQUIE, Y., EHRENHOFER-MURRAY, A. & ROSPERT, S. (2003) The yeast N(alpha)-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. *Mol Cell Biol,* 23**,** 7403-14.

GAUTSCHI, M., LILIE, H., FUNFSCHILLING, U., MUN, A., ROSS, S., LITHGOW, T., RUCKNAGEL, P. & ROSPERT, S. (2001) RAC, a stable ribosome-associated complex in yeast formed by the DnaK-DnaJ homologs Ssz1p and zuotin. *Proc Natl Acad Sci U S A,* 98**,** 3762-7.

GIETZ, R. D. & SCHIESTL, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc,* 2**,** 31-4.

GIGLIONE, C., SERERO, A., PIERRE, M., BOISSON, B. & MEINNEL, T. (2000) Identification of eukaryotic peptide deformylases reveals universality of N-terminal protein processing mechanisms. *Embo J,* 19**,** 5916-29.

GOLDSTEIN, A. L. & MCCUSKER, J. H. (1999) Three new dominant drug resistance cassettes for gene disruption in Saccharomyces cerevisiae. *Yeast,* 15**,** 1541-53.

GONDA, D. K., BACHMAIR, A., WUNNING, I., TOBIAS, J. W., LANE, W. S. & VARSHAVSKY, A. (1989) Universality and structure of the N-end rule. *J Biol Chem,* 264**,** 16700-12.

GORLICH, D. & RAPOPORT, T. A. (1993) Protein translocation into proteoliposomes reconstituted from purified components of the endoplasmic reticulum membrane. *Cell,* 75**,** 615-30.

HAIGH, N. G. & JOHNSON, A. E. (2002) A new role for BiP: closing the aqueous translocon pore during protein integration into the ER membrane. *J Cell Biol,* 156**,** 261-70.

HALDAR, M. K., SCOTT, M. D., SULE, N., SRIVASTAVA, D. K. & MALLIK, S. (2008) Synthesis of barbiturate-based methionine aminopeptidase-1 inhibitors. *Bioorg Med Chem Lett*.

HALIC, M., GARTMANN, M., SCHLENKER, O., MIELKE, T., POOL, M. R., SINNING, I. & BECKMANN, R. (2006) Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science,* 312**,** 745-7.

HAMMAN, B. D., CHEN, J. C., JOHNSON, E. E. & JOHNSON, A. E. (1997) The aqueous pore through the translocon has a diameter of 40-60 A during cotranslational protein translocation at the ER membrane. *Cell,* 89**,** 535-44.

HAMMAN, B. D., HENDERSHOT, L. M. & JOHNSON, A. E. (1998) BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. *Cell,* 92**,** 747-58.

HANAHAN, D. (1983) Studies on transformation of Escherichia coli with plasmids. *J Mol Biol,* 166**,** 557-80.

HANEIN, D., MATLACK, K. E., JUNGNICKEL, B., PLATH, K., KALIES, K. U., MILLER, K. R., RAPOPORT, T. A. & AKEY, C. W. (1996) Oligomeric rings of the Sec61p complex induced by ligands required for protein translocation. *Cell,* 87**,** 721-32.

HANN, B. C. & WALTER, P. (1991) The signal recognition particle in S. cerevisiae. *Cell,* 67**,** 131-44.

HARTMANN, E., SOMMER, T., PREHN, S., GORLICH, D., JENTSCH, S. & RAPOPORT, T. A. (1994) Evolutionary conservation of components of the protein translocation complex. *Nature,* 367**,** 654-7.

HIREL, P. H., SCHMITTER, M. J., DESSEN, P., FAYAT, G. & BLANQUET, S. (1989) Extent of N-terminal methionine excision from Escherichia coli proteins is governed by the side-chain length of the penultimate amino acid. *Proc Natl Acad Sci U S A,* 86**,** 8247-51.

HOLST, B., BRUUN, A. W., KIELLAND-BRANDT, M. C. & WINTHER, J. R. (1996) Competition between folding and glycosylation in the endoplasmic reticulum. *Embo J,* 15**,** 3538-46.

HUANG, S., ELLIOTT, R. C., LIU, P. S., KODURI, R. K., WEICKMANN, J. L., LEE, J. H., BLAIR, L. C., GHOSH-DASTIDAR, P., BRADSHAW, R. A., BRYAN, K. M. & ET AL. (1987) Specificity of cotranslational amino-terminal processing of proteins in yeast. *Biochemistry,* 26**,** 8242-6.

HWANG, C. S., SHEMORRY, A. & VARSHAVSKY, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science,* 327**,** 973-7.

JACKSON, R. & HUNTER, T. (1970) Role of methionine in the initiation of haemoglobin synthesis. *Nature,* 227**,** 672-6.

JERMY, A. J., WILLER, M., DAVIS, E., WILKINSON, B. M. & STIRLING, C. J. (2006) The Brl domain in Sec63p is required for assembly of functional endoplasmic reticulum translocons. *J Biol Chem,* 281**,** 7899-906.

JIANG, Y., CHENG, Z., MANDON, E. C. & GILMORE, R. (2008) An interaction between the SRP receptor and the translocon is critical during cotranslational protein translocation. *J Cell Biol,* 180**,** 1149-61.

JUNG, G., UENO, H. & HAYASHI, R. (1999) Carboxypeptidase Y: structural basis for protein sorting and catalytic triad. *J Biochem,* 126**,** 1-6.

KABANI, M., BECKERICH, J. M. & GAILLARDIN, C. (2000) Sls1p stimulates Sec63p-mediated activation of Kar2p in a conformation-dependent manner in the yeast endoplasmic reticulum. *Mol Cell Biol,* 20**,** 6923-34.

KEENAN, R. J., FREYMANN, D. M., STROUD, R. M. & WALTER, P. (2001) The signal recognition particle. *Annu Rev Biochem,* 70**,** 755-75.

KENDALL, R. L. & BRADSHAW, R. A. (1992) Isolation and characterization of the methionine aminopeptidase from porcine liver responsible for the co-translational processing of proteins. *J Biol Chem,* 267**,** 20667-73.

KRAMER, G., BOEHRINGER, D., BAN, N. & BUKAU, B. (2009) The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nat Struct Mol Biol,* 16**,** 589-97.

KURZCHALIA, T. V., WIEDMANN, M., GIRSHOVICH, A. S., BOCHKAREVA, E. S., BIELKA, H. & RAPOPORT, T. A. (1986) The signal sequence of nascent preprolactin interacts with the 54K polypeptide of the signal recognition particle. *Nature,* 320**,** 634-6.

KWON, Y. T., BALOGH, S. A., DAVYDOV, I. V., KASHINA, A. S., YOON, J. K., XIE, Y., GAUR, A., HYDE, L., DENENBERG, V. H. &

VARSHAVSKY, A. (2000) Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. *Mol Cell Biol,* 20**,** 4135-48.

KWON, Y. T., KASHINA, A. S., DAVYDOV, I. V., HU, R. G., AN, J. Y., SEO, J. W., DU, F. & VARSHAVSKY, A. (2002) An essential role of N-terminal arginylation in cardiovascular development. *Science,* 297**,** 96-9.

KYTE, J. & DOOLITTLE, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol,* 157**,** 105-32.

LEE, F. J., LIN, L. W. & SMITH, J. A. (1989) N alpha-acetyltransferase deficiency alters protein synthesis in Saccharomyces cerevisiae. *FEBS Lett,* 256**,** 139-42.

LEOPOLDINI, M., RUSSO, N. & TOSCANO, M. (2007) Which one among Zn(II), Co(II), Mn(II), and Fe(II) is the most efficient ion for the methionine aminopeptidase catalyzed reaction? *J Am Chem Soc,* 129**,** 7776-84.

LI, A., XUE, Y., JIN, C., WANG, M. & YAO, X. (2006) Prediction of Nepsilon-acetylation on internal lysines implemented in Bayesian Discriminant Method. *Biochem Biophys Res Commun,* 350**,** 818-24.

LI, J. Y., CHEN, L. L., CUI, Y. M., LUO, Q. L., LI, J., NAN, F. J. & YE, Q. Z. (2003) Specificity for inhibitors of metal-substituted methionine aminopeptidase. *Biochem Biophys Res Commun,* 307**,** 172-9.

LI, X. & CHANG, Y. H. (1995) Amino-terminal protein processing in Saccharomyces cerevisiae is an essential function that requires two distinct methionine aminopeptidases. *Proc Natl Acad Sci U S A,* 92**,** 12357-61.

LIU, S., WIDOM, J., KEMP, C. W., CREWS, C. M. & CLARDY, J. (1998) Structure of human methionine aminopeptidase-2 complexed with fumagillin. *Science,* 282**,** 1324-7.

LONGTINE, M. S., MCKENZIE, A., 3RD, DEMARINI, D. J., SHAH, N. G., WACH, A., BRACHAT, A., PHILIPPSEN, P. & PRINGLE, J. R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast,* 14**,** 953-61.

LOWTHER, W. T. & MATTHEWS, B. W. (2000) Structure and function of the methionine aminopeptidases. *Biochim Biophys Acta,* 1477**,** 157-67.

LYKO, F., MARTOGLIO, B., JUNGNICKEL, B., RAPOPORT, T. A. & DOBBERSTEIN, B. (1995) Signal sequence processing in rough microsomes. *J Biol Chem,* 270**,** 19873-8.

LYMAN, S. K. & SCHEKMAN, R. (1997) Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BiP. *Cell,* 88**,** 85-96.

M.J. BETTS, R. B. R. (2003) Amino acid properties and consequences of subsitutions. IN M.R. BARNES, I. C. G. (Ed.) *Bioinformatics for Geneticists.* Wiley.

MARTINEZ, A., TRAVERSO, J. A., VALOT, B., FERRO, M., ESPAGNE, C., EPHRITIKHINE, G., ZIVY, M., GIGLIONE, C. & MEINNEL, T. (2008) Extent of N-terminal modifications in cytosolic proteins from eukaryotes. *Proteomics,* 8**,** 2809-31.

MATLACK, K. E., MISSELWITZ, B., PLATH, K. & RAPOPORT, T. A. (1999) BiP acts as a molecular ratchet during posttranslational transport of prepro-alpha factor across the ER membrane. *Cell,* 97**,** 553-64.

MATOBA, S. & OGRYDZIAK, D. M. (1998) Another factor besides hydrophobicity can affect signal peptide interaction with signal recognition particle. *J Biol Chem,* 273**,** 18841-7.

MAURER-STROH, S., EISENHABER, B. & EISENHABER, F. (2002a) N-terminal N-myristoylation of proteins: prediction of substrate proteins from amino acid sequence. *J Mol Biol,* 317**,** 541-57.

MAURER-STROH, S., EISENHABER, B. & EISENHABER, F. (2002b) N-terminal N-myristoylation of proteins: refinement of the sequence motif and its taxon-specific differences. *J Mol Biol,* 317**,** 523-40.

MCCLELLAN, A. J., ENDRES, J. B., VOGEL, J. P., PALAZZI, D., ROSE, M. D. & BRODSKY, J. L. (1998) Specific molecular chaperone interactions and an ATP-dependent conformational change are required during posttranslational protein translocation into the yeast ER. *Mol Biol Cell,* 9**,** 3533-45.

MEINNEL, T., MECHULAM, Y. & BLANQUET, S. (1993) Methionine as translation start signal: a review of the enzymes of the pathway in Escherichia coli. *Biochimie,* 75**,** 1061-75.

MEINNEL, T., PEYNOT, P. & GIGLIONE, C. (2005) Processed N-termini of mature proteins in higher eukaryotes and their major contribution to dynamic proteomics. *Biochimie,* 87**,** 701-12.

MEINNEL, T., SERERO, A. & GIGLIONE, C. (2006) Impact of the N-terminal amino acid on targeted protein degradation. *Biol Chem,* 387**,** 839-51.

MICHEL, H., HUNT, D. F., SHABANOWITZ, J. & BENNETT, J. (1988) Tandem mass spectrometry reveals that three photosystem II proteins of spinach chloroplasts contain N-acetyl-O-phosphothreonine at their NH2 termini. *J Biol Chem,* 263**,** 1123-30.

MILLER, J. D., WILHELM, H., GIERASCH, L., GILMORE, R. & WALTER, P. (1993) GTP binding and hydrolysis by the signal recognition particle during initiation of protein translocation. *Nature,* 366**,** 351-4.

MISSELWITZ, B., STAECK, O. & RAPOPORT, T. A. (1998) J proteins catalytically activate Hsp70 molecules to trap a wide range of peptide sequences. *Mol Cell,* 2**,** 593-603.

MITRA, K., FRANK, J. & DRIESSEN, A. (2006) Co- and post-translational translocation through the protein-conducting channel: analogous mechanisms at work? *Nat Struct Mol Biol,* 13**,** 957-64.

MOERSCHELL, R. P., HOSOKAWA, Y., TSUNASAWA, S. & SHERMAN, F. (1990) The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes c created by oligonucleotide transformation. *J Biol Chem,* 265**,** 19638-43.

MOGK, A., SCHMIDT, R. & BUKAU, B. (2007) The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol,* 17**,** 165-72.

NELSON, R. J., ZIEGELHOFFER, T., NICOLET, C., WERNER-WASHBURNE, M. & CRAIG, E. A. (1992) The translation machinery and 70 kd heat shock protein cooperate in protein synthesis. *Cell,* 71**,** 97-105.

NG, D. T., BROWN, J. D. & WALTER, P. (1996) Signal sequences specify the targeting route to the endoplasmic reticulum membrane. *J Cell Biol,* 134**,** 269-78.

NILSSON, I. & VON HEIJNE, G. (2000) Glycosylation efficiency of Asn-Xaa-Thr sequons depends both on the distance from the C terminus and on the presence of a downstream transmembrane segment. *J Biol Chem,* 275**,** 17338-43.

PALADE, G. (1975) Intracellular aspects of the process of protein synthesis. *Science,* 189**,** 347-58.

PANZNER, S., DREIER, L., HARTMANN, E., KOSTKA, S. & RAPOPORT, T. A. (1995) Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. *Cell,* 81**,** 561-70.

PERSSON, B., FLINTA, C., VON HEIJNE, G. & JORNVALL, H. (1985) Structures of N-terminally acetylated proteins. *Eur J Biochem,* 152**,** 523-7.

PLATH, K., MOTHES, W., WILKINSON, B. M., STIRLING, C. J. & RAPOPORT, T. A. (1998) Signal sequence recognition in posttranslational protein transport across the yeast ER membrane. *Cell,* 94**,** 795-807.

PLATH, K. & RAPOPORT, T. A. (2000) Spontaneous release of cytosolic proteins from posttranslational substrates before their transport into the endoplasmic reticulum. *J Cell Biol,* 151**,** 167-78.

POLEVODA, B., ARNESEN, T. & SHERMAN, F. (2009a) A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. *BMC Proc,* 3 Suppl 6**,** S2.

POLEVODA, B., BROWN, S., CARDILLO, T. S., RIGBY, S. & SHERMAN, F. (2008) Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. *J Cell Biochem,* 103**,** 492-508.

POLEVODA, B., CARDILLO, T. S., DOYLE, T. C., BEDI, G. S. & SHERMAN, F. (2003) Nat3p and Mdm20p are required for function of yeast NatB Nalpha-terminal acetyltransferase and of actin and tropomyosin. *J Biol Chem,* 278**,** 30686-97.

POLEVODA, B., HOSKINS, J. & SHERMAN, F. (2009b) Properties of Nat4, an N(alpha)-acetyltransferase of Saccharomyces cerevisiae that modifies N termini of histones H2A and H4. *Mol Cell Biol,* 29**,** 2913-24.

POLEVODA, B., NORBECK, J., TAKAKURA, H., BLOMBERG, A. & SHERMAN, F. (1999) Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae. *Embo J,* 18**,** 6155-68.

POLEVODA, B. & SHERMAN, F. (2000) Nalpha -terminal acetylation of eukaryotic proteins. *J Biol Chem,* 275**,** 36479-82.

POLEVODA, B. & SHERMAN, F. (2001) NatC Nalpha-terminal acetyltransferase of yeast contains three subunits, Mak3p, Mak10p, and Mak31p. *J Biol Chem,* 276**,** 20154-9.

POLEVODA, B. & SHERMAN, F. (2003a) Composition and function of the eukaryotic N-terminal acetyltransferase subunits. *Biochem Biophys Res Commun,* 308**,** 1-11.

POLEVODA, B. & SHERMAN, F. (2003b) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. *J Mol Biol,* 325**,** 595-622.

POOL, M. R., STUMM, J., FULGA, T. A., SINNING, I. & DOBBERSTEIN, B. (2002) Distinct modes of signal recognition particle interaction with the ribosome. *Science,* 297**,** 1345-8.

RAJALA, R. V., DATLA, R. S., MOYANA, T. N., KAKKAR, R., CARLSEN, S. A. & SHARMA, R. K. (2000) N-myristoyltransferase. *Mol Cell Biochem,* 204**,** 135-55.

RAMOS, C., WINTHER, J. R. & KIELLAND-BRANDT, M. C. (1994) Requirement of the propeptide for in vivo formation of active yeast carboxypeptidase Y. *J Biol Chem,* 269**,** 7006-12.

RAPIEJKO, P. J. & GILMORE, R. (1997) Empty site forms of the SRP54 and SR alpha GTPases mediate targeting of ribosome-nascent chain complexes to the endoplasmic reticulum. *Cell,* 89**,** 703-13.

RAPOPORT, T. A. (2007) Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature,* 450**,** 663-9.

RAPOPORT, T. A., MATLACK, K. E., PLATH, K., MISSELWITZ, B. & STAECK, O. (1999) Posttranslational protein translocation across the membrane of the endoplasmic reticulum. *Biol Chem,* 380**,** 1143-50.

RAUE, U., OELLERER, S. & ROSPERT, S. (2007) Association of protein biogenesis factors at the yeast ribosomal tunnel exit is affected by the translational status and nascent polypeptide sequence. *J Biol Chem,* 282**,** 7809-16.

RAY, M. K., DATTA, B., CHAKRABORTY, A., CHATTOPADHYAY, A., MEZA-KEUTHEN, S. & GUPTA, N. K. (1992) The eukaryotic initiation factor 2-associated 67-kDa polypeptide (p67) plays a critical role in regulation of protein synthesis initiation in animal cells. *Proc Natl Acad Sci U S A,* 89**,** 539-43.

RODERICK, S. L. & MATTHEWS, B. W. (1993) Structure of the cobalt-dependent methionine aminopeptidase from Escherichia coli: a new type of proteolytic enzyme. *Biochemistry,* 32**,** 3907-12.

ROSPERT, S., DUBAQUIE, Y. & GAUTSCHI, M. (2002) Nascent-polypeptide-associated complex. *Cell Mol Life Sci,* 59**,** 1632-9.

RUPP, S. & WOLF, D. H. (1995) Biogenesis of the yeast vacuole (lysosome). The use of active-site mutants of proteinase yscA to determine the necessity of the enzyme for vacuolar proteinase maturation and proteinase yscB stability. *Eur J Biochem,* 231**,** 115-25.

SCHATZ, G. & DOBBERSTEIN, B. (1996) Common principles of protein translocation across membranes. *Science,* 271**,** 1519-26.

SEO, J. H., LIM, J. C., LEE, D. Y., KIM, K. S., PISZCZEK, G., NAM, H. W., KIM, Y. S., AHN, T., YUN, C. H., KIM, K., CHOCK, P. B. & CHAE, H. Z. (2009) Novel protective mechanism against irreversible hyperoxidation of peroxiredoxin: Nalpha-terminal acetylation of human peroxiredoxin II. *J Biol Chem,* 284**,** 13455-65.

SERERO, A., GIGLIONE, C., SARDINI, A., MARTINEZ-SANZ, J. & MEINNEL, T. (2003) An unusual peptide deformylase features in the human mitochondrial N-terminal methionine excision pathway. *J Biol Chem,* 278**,** 52953-63.

SIN, N., MENG, L., WANG, M. Q., WEN, J. J., BORNMANN, W. G. & CREWS, C. M. (1997) The anti-angiogenic agent fumagillin covalently binds and inhibits the methionine aminopeptidase, MetAP-2. *Proc Natl Acad Sci U S A,* 94**,** 6099-103.

SOLBIATI, J., CHAPMAN-SMITH, A., MILLER, J. L., MILLER, C. G. & CRONAN, J. E., JR. (1999) Processing of the N termini of nascent polypeptide chains requires deformylation prior to methionine removal. *J Mol Biol,* 290**,** 607-14.

SONG, O. K., WANG, X., WATERBORG, J. H. & STERNGLANZ, R. (2003) An Nalpha-acetyltransferase responsible for acetylation of the N-terminal residues of histones H4 and H2A. *J Biol Chem,* 278**,** 38109-12.

SORENSEN, S. O., VAN DEN HAZEL, H. B., KIELLAND-BRANDT, M. C. & WINTHER, J. R. (1994) pH-dependent processing of yeast

procarboxypeptidase Y by proteinase A in vivo and in vitro. *Eur J Biochem,* 220**,** 19-27.

STEEL, G. J., BROWNSWORD, J. & STIRLING, C. J. (2002) Tail-anchored protein insertion into yeast ER requires a novel posttranslational mechanism which is independent of the SEC machinery. *Biochemistry,* 41**,** 11914-20.

STEEL, G. J., FULLERTON, D. M., TYSON, J. R. & STIRLING, C. J. (2004) Coordinated activation of Hsp70 chaperones. *Science,* 303**,** 98-101.

STEVENS, T., ESMON, B. & SCHEKMAN, R. (1982) Early stages in the yeast secretory pathway are required for transport of carboxypeptidase Y to the vacuole. *Cell,* 30**,** 439-48.

STIRLING, C. J. & HEWITT, E. W. (1992) The S. cerevisiae SEC65 gene encodes a component of yeast signal recognition particle with homology to human SRP19. *Nature,* 356**,** 534-7.

STIRLING, C. J., ROTHBLATT, J., HOSOBUCHI, M., DESHAIES, R. & SCHEKMAN, R. (1992) Protein translocation mutants defective in the insertion of integral membrane proteins into the endoplasmic reticulum. *Mol Biol Cell,* 3**,** 129-42.

TSUNASAWA, S., IZU, Y., MIYAGI, M. & KATO, I. (1997) Methionine aminopeptidase from the hyperthermophilic Archaeon Pyrococcus furiosus: molecular cloning and overexpression in Escherichia coli of the gene, and characteristics of the enzyme. *J Biochem,* 122**,** 843-50.

TSUNASAWA, S., STEWART, J. W. & SHERMAN, F. (1985) Amino-terminal processing of mutant forms of yeast iso-1-cytochrome c. The specificities of methionine aminopeptidase and acetyltransferase. *J Biol Chem,* 260**,** 5382-91.

TUCKER, L. A., ZHANG, Q., SHEPPARD, G. S., LOU, P., JIANG, F., MCKEEGAN, E., LESNIEWSKI, R., DAVIDSEN, S. K., BELL, R. L. & WANG, J. (2008) Ectopic expression of methionine aminopeptidase-2 causes cell transformation and stimulates proliferation. *Oncogene*.

TYSON, J. R. & STIRLING, C. J. (2000) LHS1 and SIL1 provide a lumenal function that is essential for protein translocation into the endoplasmic reticulum. *EMBO J,* 19**,** 6440-52.

VAN DEN BERG, B., CLEMONS, W. M., JR., COLLINSON, I., MODIS, Y., HARTMANN, E., HARRISON, S. C. & RAPOPORT, T. A. (2004) X-ray structure of a protein-conducting channel. *Nature,* 427**,** 36-44.

VAN DEN HAZEL, H. B., KIELLAND-BRANDT, M. C. & WINTHER, J. R. (1996) Review: biosynthesis and function of yeast vacuolar proteases. *Yeast,* 12**,** 1-16.

VETRO, J. A. & CHANG, Y. H. (2002) Yeast methionine aminopeptidase type 1 is ribosome-associated and requires its N-terminal zinc finger domain for normal function in vivo. *J Cell Biochem,* 85**,** 678-88.

VON HEIJNE, G. (1984) How signal sequences maintain cleavage specificity. *J Mol Biol,* 173**,** 243-51.

VON HEIJNE, G. (1985) Signal sequences. The limits of variation. *J Mol Biol,* 184**,** 99-105.

WALKER, K. W. & BRADSHAW, R. A. (1998) Yeast methionine aminopeptidase I can utilize either Zn2+ or Co2+ as a cofactor: a case of mistaken identity? *Protein Sci,* 7**,** 2684-7.

WALLER, J. P. (1963) The Nh2-Terminal Residues of the Proteins from Cell-Free Extracts of E. Coli. *J Mol Biol,* 7**,** 483-96.

WALTER, P. & BLOBEL, G. (1980) Purification of a membrane-associated protein complex required for protein translocation across the endoplasmic reticulum. *Proc Natl Acad Sci U S A,* 77**,** 7112-6.

WALTER, P. & BLOBEL, G. (1981) Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. *J Cell Biol,* 91**,** 557-61.

WALTER, P., IBRAHIMI, I. & BLOBEL, G. (1981) Translocation of proteins across the endoplasmic reticulum. I. Signal recognition protein (SRP) binds to in-vitro-assembled polysomes synthesizing secretory protein. *J Cell Biol,* 91**,** 545-50.

WILKINSON, B. M., CRITCHLEY, A. J. & STIRLING, C. J. (1996) Determination of the transmembrane topology of yeast Sec61p, an essential component of the endoplasmic reticulum translocation complex. *J Biol Chem,* 271**,** 25590-7.

WILLER, M., FORTE, G. M. & STIRLING, C. J. (2008) Sec61p is required for ERAD-L: genetic dissection of the translocation and ERAD-L functions of Sec61P using novel derivatives of CPY. *J Biol Chem,* 283**,** 33883-8.

WINZELER, E. A., SHOEMAKER, D. D., ASTROMOFF, A., LIANG, H., ANDERSON, K., ANDRE, B., BANGHAM, R., BENITO, R., BOEKE, J. D., BUSSEY, H., CHU, A. M., CONNELLY, C., DAVIS, K., DIETRICH, F., DOW, S. W., EL BAKKOURY, M., FOURY, F., FRIEND, S. H., GENTALEN, E., GIAEVER, G., HEGEMANN, J. H., JONES, T., LAUB, M., LIAO, H., LIEBUNDGUTH, N., LOCKHART, D. J., LUCAU-DANILA, A., LUSSIER, M., M'RABET, N., MENARD, P., MITTMANN, M., PAI, C., REBISCHUNG, C., REVUELTA, J. L., RILES, L., ROBERTS, C. J., ROSS-MACDONALD, P., SCHERENS, B., SNYDER, M., SOOKHAI-MAHADEO, S., STORMS, R. K., VERONNEAU, S., VOET, M., VOLCKAERT, G., WARD, T. R., WYSOCKI, R., YEN, G. S., YU, K., ZIMMERMANN, K., PHILIPPSEN, P., JOHNSTON, M. & DAVIS, R. W. (1999) Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science,* 285**,** 901-6.

WU, S., GUPTA, S., CHATTERJEE, N., HILEMAN, R. E., KINZY, T. G., DENSLOW, N. D., MERRICK, W. C., CHAKRABARTI, D., OSTERMAN, J. C. & GUPTA, N. K. (1993) Cloning and characterization of complementary DNA encoding the eukaryotic initiation factor 2-associated 67-kDa protein (p67). *J Biol Chem,* 268**,** 10796-801.

XIAO, Q., ZHANG, F., NACEV, B. A., LIU, J. O. & PEI, D. (2010) Protein N-Terminal Processing: Substrate Specificity of Escherichia coli and Human Methionine Aminopeptidases. *Biochemistry,* 49**,** 5588-99.

YAN, W., SCHILKE, B., PFUND, C., WALTER, W., KIM, S. & CRAIG, E. A. (1998) Zuotin, a ribosome-associated DnaJ molecular chaperone. *Embo J,* 17**,** 4809-17.

YOUNG, B. P., CRAVEN, R. A., REID, P. J., WILLER, M. & STIRLING, C. J. (2001) Sec63p and Kar2p are required for the translocation of SRP-dependent precursors into the yeast endoplasmic reticulum in vivo. *Embo J,* 20**,** 262-71.

ZUO, S., GUO, Q., LING, C. & CHANG, Y. H. (1995) Evidence that two zinc fingers in the methionine aminopeptidase from Saccharomyces cerevisiae are important for normal growth. *Mol Gen Genet,* 246**,** 247-53.

Appendix

Appendix

alpha factor K-less 14 C Acetyl CoA scintillation counts **30ul samples**

	no mRNA	no mRNA background adjusted	WT (MR)	WT background adjusted	S (MS)	S background adjusted	background buffer
	15.2	3.8	22.2	10.8667	54.2	42.8667	10.2
	20.8	9.4	16	4.6667	49.4	38.0667	10
	20.8	9.4	22	10.6667	45.6	34.2667	14
mean		7.5333		8.7333		38.4	11.4
SD	3.2332	3.2332	3.5233	3.5233	4.3097	4.3097	2.2539

T-test	noRNA vs WT	0.6864
T-test	noRNA vs S	8.37E-04
T-test	WT vs S	9.17E-04

alpha factor K-less) 35 S methionine scintillation counts **5ul samples**

	no mRNA	no mRNA background adjusted	WT (MR)	WT background adjusted	S (MS)	S background adjusted	background buffer
	43	21	9049	9027	6593	6571	21
	68	46	6133	6111	6628	6606	25
	59	37	8745	8723	7300	7278	20
mean		34.6667		7953.6667		6818.3333	22
SD	12.6623	12.6623	1603.0188	1603.0188	398.4675	398.4675	2.6458

T-test	noRNA vs WT	0.0134
T-test	noRNA vs S	1.14E-03
T-test	WT vs S	0.3444

Ost-alpha factor K-less (pgf24/25) 14 C Acetyl CoA scintillation counts

30ul samples

	no mRNA	no mRNA background adjusted		WT (MR)	WT background adjusted	S (MS)	S background adjusted	background buffer
	15.6	2.2		16.6	3.2	36.6	23.2	13.4
	18.6	5.2		21.6	8.2	35.6	22.2	13.6
	17.2	3.8		17	3.6	36.8	23.4	13.2
mean		3.7333			5		22.9333	13.4
SD	1.5011	1.5011		2.7785	2.7785	0.6429	0.6429	0.2

T-test	noRNA vs WT	0.5361
T-test	noRNA vs S	4.82E-04
T-test	WT vs S	5.77E-03

Ost-alpha factor K-less (pgf24/25) 35 S methionine scintillation counts

5ul samples

	no mRNA	no mRNA background adjusted		WT (MR)	WT background adjusted	S (MS)	S background adjusted	background buffer
	131	117.6		11472	11458.6	17269	17255.6	116
	169	155.6		6626	6612.6	7865	7851.6	106
	140	126.6		11922	11908.6	9920	9906.6	122
mean		133.2667			9993.267		11671.27	114.6667
SD	19.8578	19.8578		2936.376	2936.376	4944.122	4944.122	8.0829

T-test	noRNA vs WT	0.0283
T-test	noRNA vs S	0.0561
T-test	WT vs S	0.6455

ost-alpha factor met HA RNCs 14C Acetyl CoA scintillation counts 30 ul samples

	no mRNA	WT (MR)	WT adjusted-RNA	S (MS)	S adjusted-RNA	Background buffer
	31.2	36.6	3.6	49.6	16.6	9.4
	34.2	39.2	6.2	45.8	12.8	20.8
	33.6	40.6	7.6	40	7	22.8
mean	33	38.8	5.8		12.1333	17.6667
SD	1.5875	2.0298	2.0298	4.8346	4.8346	7.2286

T-test WT vs S 0.1381472

ost-alpha factor met HA RNCs 35 S Methionine scintillation counts 10 ul samples

	no mRNA	WT (MR)	WT adjusted-RNA	S (MS)	S adjusted-RNA	background buffer
	1496	4684	3588	2665	1569	156
	805	6114	5018	3732	2636	191
	987	5982	4886	4133	3037	156
mean	1096		4497.3333		2414	167.6667
SD	358.163	790.267	790.2666	758.7615	758.7615	20.2073

T-test WT vs S 0.0301843

N-terminal sequence and predicted processing of Yeast Signal Sequences

Protein	N-terminal residues	Predicted Met cleavage	Predicted Acetylation*
ADP1	MG	Cleaved	Non-acetylated
AGA1	MT	Cleaved	Non-acetylated
AGA2	MQ	Uncleaved	Non-acetylated
AIM6	ML	Uncleaved	Non-acetylated
ANS1	MK	Uncleaved	Non-acetylated
APE3	MHF	Uncleaved	Non-acetylated
ASP3-1	MR	Uncleaved	Non-acetylated
ASP3-2	MR	Uncleaved	Non-acetylated
ASP3-3	MR	Uncleaved	Non-acetylated
ASP3-4	MR	Uncleaved	Non-acetylated
ATG15	ML	Uncleaved	Non-acetylated
ATG27	MV	Cleaved	Acetylated
ATH1	MK	Uncleaved	Non-acetylated
AWA1	MFN	Uncleaved	Acetylated
AXL2	MT	Cleaved	Acetylated
BAR1	MS	Cleaved	Acetylated
BGL2	MR	Uncleaved	Non-acetylated
BIG1	MQ	Uncleaved	Non-acetylated
CCW12	MQ	Uncleaved	Non-acetylated
CCW14	MR	Uncleaved	Non-acetylated
CDA1	MK	Uncleaved	Non-acetylated
CDA2	MR	Uncleaved	Non-acetylated
CIS3	MQ	Uncleaved	Non-acetylated
CNE1	MK	Uncleaved	Non-acetylated
CPR4	MWL	Uncleaved	Non-acetylated
CPR5	MK	Uncleaved	Non-acetylated
CRH1	MK	Uncleaved	Non-acetylated
CRR1	MR	Uncleaved	Non-acetylated
CSG2	MS	Cleaved	Acetylated
CTS1	MS	Cleaved	Acetylated
CTS2	MV	Cleaved	Acetylated
CWH43	ML	Uncleaved	Non-acetylated
CWP1	MK	Uncleaved	Non-acetylated
CWP2	MQ	Uncleaved	Non-acetylated
CYP2	MK	Uncleaved	Non-acetylated
DAN1	MS	Cleaved	Acetylated
DAN2	MV	Cleaved	Acetylated
DAN3	MV	Cleaved	Acetylated
DAN4	MV	Cleaved	Acetylated
DCW1	ML	Uncleaved	Non-acetylated
DDR2	MK	Uncleaved	Non-acetylated

DFG5	MI	Uncleaved	Non-acetylated
DIA3	MV	Cleaved	Acetylated
DSE2	MK	Uncleaved	Non-acetylated
DSE4	MQ	Uncleaved	Non-acetylated
ECM14	ML	Uncleaved	Non-acetylated
ECM33	MQ	Uncleaved	Non-acetylated
EGT2	MN	Uncleaved	Acetylated *
EMC1	MK	Uncleaved	Non-acetylated
EMP24	MAS	Cleaved	Acetylated
EMP46	MT	Cleaved	Acetylated
EMP47	MMM	Uncleaved	Non-acetylated
EMP70	MI	Uncleaved	Non-acetylated
EPS1	MK	Uncleaved	Non-acetylated
ERJ5	MN	Uncleaved	Acetylated *
ERO1	MR	Uncleaved	Non-acetylated
ERP1	ML	Uncleaved	Non-acetylated
ERP2	MI	Uncleaved	Non-acetylated
ERP3	MS	Cleaved	Acetylated
ERP4	MR	Uncleaved	Non-acetylated
ERP5	MR	Uncleaved	Non-acetylated
ERP6	ML	Uncleaved	Non-acetylated
ERV25	MQ	Uncleaved	Non-acetylated
EUG1	MQ	Uncleaved	Non-acetylated
EXG1	ML	Uncleaved	Non-acetylated
EXG2	MP	Cleaved	Non-acetylated
FET3	MT	Cleaved	Acetylated
FET5	ML	Uncleaved	Non-acetylated
FIG2	MN	Uncleaved	Acetylated *
FIT1	MK	Uncleaved	Non-acetylated
FIT2	MK	Uncleaved	Non-acetylated
FIT3	MK	Uncleaved	Non-acetylated
FLC1	MQ	Uncleaved	Non-acetylated
FLC2	MI	Uncleaved	Non-acetylated
FLC3	MR	Uncleaved	Non-acetylated
FLO1	MT	Cleaved	Acetylated
FLO10	MP	Cleaved	Non-acetylated
FLO11	MQ	Uncleaved	Non-acetylated
FLO5	MT	Cleaved	Acetylated
FLO9	MS	Cleaved	Acetylated
FMN1	MFT	Uncleaved	Non-acetylated
FPR2	MMF	Uncleaved	Non-acetylated
FRE1	MV	Cleaved	Acetylated
FRE2	MHW	Uncleaved	Non-acetylated
FRE3	MY	Uncleaved	Non-acetylated
FRE4	ML	Uncleaved	Non-acetylated
FRE5	ML	Uncleaved	Non-acetylated

FRE6	MHR	Uncleaved	Acetylated
GAB1	MD	Uncleaved	Acetylated
GAS1	ML	Uncleaved	Non-acetylated
GAS2	MN	Uncleaved	Acetylated *
GAS3	MQ	Uncleaved	Non-acetylated
GAS4	MMV	Uncleaved	Non-acetylated
GAS5	ML	Uncleaved	Non-acetylated
GPI16	MI	Uncleaved	Non-acetylated
GPI17	MS	Cleaved	Acetylated
GPI8	MR	Uncleaved	Non-acetylated
GRX6	MI	Uncleaved	Non-acetylated
GTB1	MV	Cleaved	Acetylated
HKR1	MV	Cleaved	Acetylated
HOR7	MK	Uncleaved	Non-acetylated
HPF1	MFN	Uncleaved	Acetylated
HRD3	MI	Uncleaved	Non-acetylated
HSP150	MQ	Uncleaved	Non-acetylated
IRC22	MR	Uncleaved	Non-acetylated
IRE1	MR	Uncleaved	Non-acetylated
JEM1	MI	Uncleaved	Non-acetylated
KAR2	MFF	Uncleaved	Non-acetylated
KAR5	MFE	Uncleaved	Non-acetylated
KEX1	MFY	Uncleaved	Non-acetylated
KEX2	MK	Uncleaved	Non-acetylated
KNH1	ML	Uncleaved	Non-acetylated
KRE1	MMR	Uncleaved	Acetylated
KRE27	MS	Cleaved	Acetylated
KRE5	MR	Uncleaved	Non-acetylated
KRE9	MR	Uncleaved	Non-acetylated
LHS1	MR	Uncleaved	Non-acetylated
LRC1	ML	Uncleaved	Non-acetylated
MEL1	MFA	Uncleaved	Acetylated
MEL2	MFA	Uncleaved	Acetylated
MEL5	MFA	Uncleaved	Acetylated
MEL6	MFA	Uncleaved	Acetylated
MFALPHA	MR	Uncleaved	Non-acetylated
MFALPHA2	MK	Uncleaved	Non-acetylated
MID1	MI	Uncleaved	Non-acetylated
MID2	ML	Uncleaved	Non-acetylated
MKC7	MK	Uncleaved	Non-acetylated
MNL1	MV	Cleaved	Acetylated
MNN5	ML	Uncleaved	Non-acetylated
MPD1	ML	Uncleaved	Non-acetylated
MPD2	MK	Uncleaved	Non-acetylated
MRL1	ML	Uncleaved	Non-acetylated
MSB2	MQ	Uncleaved	Non-acetylated

MTL1	MAS	Cleaved	Acetylated
MUC1	MQ	Uncleaved	Non-acetylated
NCP1	MP	Cleaved	Non-acetylated
NCR1	MN	Uncleaved	Acetylated *
NPC2	MT	Cleaved	Acetylated
NVJ1	MT	Cleaved	Acetylated
OST1	MR	Uncleaved	Non-acetylated
OST3	MN	Uncleaved	Acetylated *
OST6	MK	Uncleaved	Non-acetylated
PAU10	MV	Cleaved	Acetylated
PAU13	MV	Cleaved	Acetylated
PAU15	MV	Cleaved	Acetylated
PDI1	MK	Uncleaved	Non-acetylated
PEP1	MI	Uncleaved	Non-acetylated
PEP4	MFS	Uncleaved	Non-acetylated
PER1	MR	Uncleaved	Non-acetylated
PGA1	MV	Cleaved	Acetylated
PGU1	MI	Uncleaved	Non-acetylated
PHO11	ML	Uncleaved	Non-acetylated
PHO12	ML	Uncleaved	Non-acetylated
PHO3	MFK	Uncleaved	Acetylated
PHO5	MFK	Uncleaved	Acetylated
PIR1	MQ	Uncleaved	Non-acetylated
PIR3	MQ	Uncleaved	Non-acetylated
PLB1	MK	Uncleaved	Non-acetylated
PLB2	MQ	Uncleaved	Non-acetylated
PLB3	MI	Uncleaved	Non-acetylated
PRB1	MK	Uncleaved	Non-acetylated
PRC1	MK	Uncleaved	Non-acetylated
PRY1	MK	Uncleaved	Non-acetylated
PRY2	MK	Uncleaved	Non-acetylated
PRY3	ML	Uncleaved	Non-acetylated
PST1	MQ	Uncleaved	Non-acetylated
PST2	MP	Cleaved	Non-acetylated
PTM1	MR	Uncleaved	Non-acetylated
RAX2	MFV	Uncleaved	Non-acetylated
RNY1	ML	Uncleaved	Non-acetylated
ROT1	MWS	Uncleaved	Non-acetylated
ROT2	MV	Cleaved	Acetylated
RRT12	MK	Uncleaved	Non-acetylated
SAG1	MFT	Uncleaved	Non-acetylated
SCJ1	MI	Uncleaved	Non-acetylated
SCW10	MR	Uncleaved	Non-acetylated
SCW11	MI	Uncleaved	Non-acetylated
SCW4	MR	Uncleaved	Non-acetylated
SED1	MK	Uncleaved	Non-acetylated

SHE10	MG	Cleaved	Non-acetylated
SIA1	MR	Uncleaved	Non-acetylated
SIL1	MV	Cleaved	Acetylated
SIM1	MK	Uncleaved	Non-acetylated
SLG1	MR	Uncleaved	Non-acetylated
SLP1	MAN	Cleaved	Non-acetylated
SOP4	ML	Uncleaved	Non-acetylated
SPI1	ML	Uncleaved	Non-acetylated
SPO19	MK	Uncleaved	Non-acetylated
SPR1	MV	Cleaved	Acetylated
SPS100	MK	Uncleaved	Non-acetylated
SPS22	MN	Uncleaved	Acetylated *
SRL1	ML	Uncleaved	Non-acetylated
SSP120	MR	Uncleaved	Non-acetylated
STA1	MV	Cleaved	Acetylated
STA2	MQ	Uncleaved	Non-acetylated
SUC1	ML	Uncleaved	Non-acetylated
SUC2	ML	Uncleaved	Non-acetylated
SUC4	ML	Uncleaved	Non-acetylated
SUN4	MK	Uncleaved	Non-acetylated
SVS1	MI	Uncleaved	Non-acetylated
SWP1	MQ	Uncleaved	Non-acetylated
TED1	ML	Uncleaved	Non-acetylated
THI22	MV	Cleaved	Acetylated
TIP1	MS	Cleaved	Acetylated
TIR1	MAY	Cleaved	Non-acetylated
TIR2	MAY	Cleaved	Non-acetylated
TIR3	MS	Cleaved	Acetylated
TIR4	MAY	Cleaved	Non-acetylated
TMN2	MK	Uncleaved	Non-acetylated
TMN3	MR	Uncleaved	Non-acetylated
TOS1	ML	Uncleaved	Non-acetylated
UIP5	MS	Cleaved	Acetylated
UTR2	MAI	Cleaved	Non-acetylated
VOA1	MV	Cleaved	Acetylated
VTH1	MAL	Cleaved	Non-acetylated
WBP1	MR	Uncleaved	Non-acetylated
WSC2	MHL	Uncleaved	Non-acetylated
WSC3	ME	Uncleaved	Acetylated
WSC4	MQ	Uncleaved	Non-acetylated
YBL008W-A	MK	Uncleaved	Non-acetylated
YBR013C	MI	Uncleaved	Non-acetylated
YBR200W-A	ML	Uncleaved	Non-acetylated
YCL012C	MK	Uncleaved	Non-acetylated
YCL048W-A	MQ	Uncleaved	Non-acetylated
YCL049C	MFS	Uncleaved	Non-acetylated

YCR012C	MK	Uncleaved	Non-acetylated
YDR053W	MR	Uncleaved	Non-acetylated
YDR134C	MQ	Uncleaved	Non-acetylated
YDR246W-A	MR	Uncleaved	Non-acetylated
YDR262W	MI	Uncleaved	Non-acetylated
YDR366C	MV	Cleaved	Acetylated
YDR415C	MR	Uncleaved	Non-acetylated
YDR524C-B	MQ	Uncleaved	Non-acetylated
YER067W	MT	Cleaved	Acetylated
YFL051C	MS	Cleaved	Acetylated
YFR012W-A	ML	Uncleaved	Non-acetylated
YGP1	MK	Uncleaved	Non-acetylated
YGR079W	MS	Cleaved	Acetylated
YHC3	MS	Cleaved	Acetylated
YHL017W	MD	Uncleaved	Acetylated
YHL042W	MK	Uncleaved	Non-acetylated
YHR138C	MK	Uncleaved	Non-acetylated
YHR214W	MFN	Uncleaved	Acetylated
YIL156W-B	MT	Cleaved	Acetylated
YIL169C	MFN	Uncleaved	Acetylated
YJL052C	MHL	Uncleaved	Non-acetylated
YJL160c	MHY	Uncleaved	Non-acetylated
YJL171C	ML	Uncleaved	Non-acetylated
YJR120W	MR	Uncleaved	Non-acetylated
YKL018C-A	ML	Uncleaved	Non-acetylated
YLR001C	MN	Uncleaved	Acetylated *
YLR040c	MI	Uncleaved	Non-acetylated
YLR042C	MK	Uncleaved	Non-acetylated
YLR104W	MS	Cleaved	Acetylated
YLR194C	MK	Uncleaved	Non-acetylated
YLR406C-A	MI	Uncleaved	Non-acetylated
YLR413W	MN	Uncleaved	Acetylated *
YLR414C	MR	Uncleaved	Non-acetylated
YMR247W-A	MAH	Cleaved	Non-acetylated
YMR272W-B	MR	Uncleaved	Non-acetylated
YMR315W-A	MT	Cleaved	Acetylated
YNL019C	ML	Uncleaved	Non-acetylated
YNL024C-A	MS	Cleaved	Acetylated
YNL033W	ML	Uncleaved	Non-acetylated
YNL217W	ME	Uncleaved	Acetylated
YOR008C-A	MWR	Uncleaved	Acetylated
YOR214C	ML	Uncleaved	Non-acetylated
YOR365C	ML	Uncleaved	Non-acetylated
YOR389W	MR	Uncleaved	Non-acetylated
YOS9	MQ	Uncleaved	Non-acetylated
YPS1	MK	Uncleaved	Non-acetylated

YPS3	MK	Uncleaved	Non-acetylated
YPS5	MQ	Uncleaved	Non-acetylated
YPS6	MQ	Uncleaved	Non-acetylated
YPS7	MT	Cleaved	Acetylated
YSP3	MK	Uncleaved	Non-acetylated
ZPS1	MK	Uncleaved	Non-acetylated

* MN acetylation is predicted in only 55% of cases

N-terminal sequence and predicted processing of Cytosolic Proteins

Protein	N-terminal residues	Predicted Met cleavage	Predicted Acetylation*
AAT2	MS	cleaved	Acetylated
ABZ2	MS	cleaved	Acetylated
ACS2	MT	cleaved	non-acetylated
ACT1	MD	uncleaved	Acetylated
ADE1	MS	cleaved	Acetylated
ADE2	MD	uncleaved	Acetylated
ADE3	MAG	cleaved	non-acetylated
ADK1	MS	cleaved	Acetylated
ALD6	MT	cleaved	non-acetylated
ALD6	MT	cleaved	non-acetylated
ANB1	MS	cleaved	Acetylated
APL3	MD	uncleaved	Acetylated
APL6	MV	cleaved	non-acetylated
APS3	MI	uncleaved	non-acetylated
ARD1	MP	cleaved	non-acetylated
ARF1	MG	cleaved	non-acetylated
ARG4	MS	cleaved	Acetylated
ARK1	MN	uncleaved	Acetylated *
ARP14	MS	cleaved	Acetylated
ARP2	MD	uncleaved	Acetylated
ASC1	MAS	cleaved	Acetylated
ASP1	MK	uncleaved	non-acetylated
ATG1	MG	cleaved	non-acetylated
ATG11	MAD	cleaved	non-acetylated
ATS1	MS	cleaved	Acetylated
ATX1	MAE	cleaved	Acetylated

AVO2	ML	uncleaved	non-acetylated
BAT2	MT	cleaved	non-acetylated
BAT2	MT	cleaved	non-acetylated
BCY1	MV	cleaved	non-acetylated
BEM1	ML	uncleaved	non-acetylated
BIM1	MS	cleaved	Acetylated
BMH1	MS	cleaved	Acetylated
BMH2	MS	cleaved	Acetylated
BNI1	ML	uncleaved	non-acetylated
BTS1	ME	uncleaved	Acetylated
BTT1	MP	cleaved	non-acetylated
CCR4	MN	uncleaved	Acetylated *
CCT3	MQ	uncleaved	non-acetylated
CCT4	MS	cleaved	Acetylated
CCT6	MS	cleaved	Acetylated
CCT7	MN	uncleaved	Acetylated *
CDC14	MR	uncleaved	non-acetylated
CDC19	MS	cleaved	Acetylated
CDC42	MQ	uncleaved	non-acetylated
CDC48	MG	cleaved	non-acetylated
CDC60	MS	cleaved	Acetylated
CDC8	MMG	uncleaved	non-acetylated
CIA1	MAS	cleaved	Acetylated
CKA2	MP	cleaved	non-acetylated
CKB2	MG	cleaved	non-acetylated
CLA4	MS	cleaved	Acetylated
CMD1	MS	cleaved	Acetylated
CMK1	MD	uncleaved	Acetylated
CPR1	MS	cleaved	Acetylated
CSL4	MAC	cleaved	non-acetylated
CYS3	MT	cleaved	non-acetylated
DCC1	MS	cleaved	Acetylated
DCP2	MS	cleaved	Acetylated
DDP1	MG	cleaved	non-acetylated
DED81	MS	cleaved	Acetylated
DIA2	MS	cleaved	Acetylated
DOA1	MG	cleaved	non-acetylated
DOG2	MP	cleaved	non-acetylated
DYS1	MS	cleaved	Acetylated
DYS1	MS	cleaved	Acetylated
EBS1	ME	uncleaved	Acetylated
EDE1	MAS	cleaved	Acetylated
EFB1	MAS	cleaved	Acetylated
EGD1	MP	cleaved	non-acetylated

EGD2	MS	cleaved	Acetylated
ERG10	MS	cleaved	Acetylated
ERG8	MS	cleaved	Acetylated
EXO70	MP	cleaved	non-acetylated
EXO84	MV	cleaved	non-acetylated
FAB1	MS	cleaved	Acetylated
FAS1	MD	uncleaved	Acetylated
FAS2	MK	uncleaved	non-acetylated
FBA1	MG	cleaved	non-acetylated
FES1	ME	uncleaved	Acetylated
FOL2	MHN	uncleaved	Acetylated
FPR1	MS	cleaved	Acetylated
FRS1	MP	cleaved	non-acetylated
FUN11	MS	cleaved	Acetylated
GAL1	MT	cleaved	non-acetylated
GAL10	MT	cleaved	non-acetylated
GCD11	MS	cleaved	Acetylated
GCN2	MS	cleaved	Acetylated
GET3	MD	uncleaved	Acetylated
GGA2	MS	cleaved	Acetylated
GIM3	ME	uncleaved	Acetylated
GIM4	ME	uncleaved	Acetylated
GLK1	MS	cleaved	Acetylated
GTO3	MS	cleaved	Acetylated
GUA1	MAA	cleaved	non-acetylated
GUS1	MP	cleaved	non-acetylated
HAC1	ME	uncleaved	Acetylated
HCH1	MV	cleaved	non-acetylated
HEK2	MS	cleaved	Acetylated
HIS3	MT	cleaved	non-acetylated
HIS4	MV	cleaved	non-acetylated
HOM2	MAG	cleaved	non-acetylated
HOM3	MP	cleaved	non-acetylated
HOM6	MS	cleaved	Acetylated
HTS1	ML	uncleaved	non-acetylated
HUL5	ML	uncleaved	non-acetylated
HXK1	MV	cleaved	non-acetylated
HXK2	MV	cleaved	non-acetylated
HYP2	MS	cleaved	Acetylated
IDP2	MT	cleaved	non-acetylated
ILS1	MS	cleaved	Acetylated
IPP1	MT	cleaved	non-acetylated
IRA2	MS	cleaved	Acetylated
JSN1	MD	uncleaved	Acetylated

KAP104	MAS	cleaved	Acetylated
KAR3	ME	uncleaved	Acetylated
KCS1	MD	uncleaved	Acetylated
KIN2	MP	cleaved	non-acetylated
KIP1	MAR	cleaved	non-acetylated
KSP1	MT	cleaved	non-acetylated
LEU2	MS	cleaved	Acetylated
LIA1	MS	cleaved	Acetylated
LIA1	MS	cleaved	Acetylated
LSM1	MS	cleaved	Acetylated
LSM2	ML	uncleaved	non-acetylated
LTP1	MT	cleaved	non-acetylated
LYS2	MT	cleaved	non-acetylated
LYS5	MV	cleaved	non-acetylated
LYS9	MG	cleaved	non-acetylated
MAP1	MS	cleaved	Acetylated
MAP2	MT	cleaved	non-acetylated
MCM2	MS	cleaved	Acetylated
MDY2	MS	cleaved	Acetylated
MES1	MS	cleaved	Acetylated
MET1	MV	cleaved	non-acetylated
MET14	MAT	cleaved	non-acetylated
MET6	MV	cleaved	non-acetylated
MET8	MV	cleaved	non-acetylated
MIH1	MN	uncleaved	Acetylated *
MON2	MAM	cleaved	non-acetylated
MOT2	MMN	uncleaved	Acetylated
MUQ1	MT	cleaved	non-acetylated
MYO2	MS	cleaved	Acetylated
MYO4	MS	cleaved	Acetylated
NAR1	MS	cleaved	Acetylated
NCS2	ME	uncleaved	Acetylated
NMD3	ME	uncleaved	Acetylated
NPL4	ML	uncleaved	non-acetylated
NRK1	MT	cleaved	non-acetylated
OPI3	MK	uncleaved	non-acetylated
PAB1	MAD	cleaved	non-acetylated
PAC10	MD	uncleaved	Acetylated
PAN1	MY	uncleaved	non-acetylated
PCK1	MS	cleaved	Acetylated
PCT1	MAN	cleaved	non-acetylated
PFD1	MS	cleaved	Acetylated
PFK1	MQ	uncleaved	non-acetylated
PFK27	MG	cleaved	non-acetylated

PFY1	MS	cleaved	Acetylated
PGI1	MS	cleaved	Acetylated
PIG2	MAT	cleaved	non-acetylated
PKC1	MS	cleaved	Acetylated
PPG1	ME	uncleaved	Acetylated
PPQ1	MR	uncleaved	non-acetylated
PPT1	MS	cleaved	Acetylated
PPZ1	MG	cleaved	non-acetylated
PRE2	MQ	uncleaved	non-acetylated
PRE6	MS	cleaved	Acetylated
PRR1	MD	uncleaved	Acetylated
PRS1	MR	uncleaved	non-acetylated
PTC1	MS	cleaved	Acetylated
PTP3	MK	uncleaved	non-acetylated
PUF2	MD	uncleaved	Acetylated
PYC1	MS	cleaved	Acetylated
PYC2	MS	cleaved	Acetylated
RAD6	MS	cleaved	Acetylated
RAS1	MQ	uncleaved	non-acetylated
RBL2	MAP	cleaved	non-acetylated
RFA1	MS	cleaved	Acetylated
RGA1	MAS	cleaved	Acetylated
RIM13	MN	uncleaved	Acetylated *
RNR4	ME	uncleaved	Acetylated
RPL17A	MAR	cleaved	non-acetylated
RPL19B	MAN	cleaved	non-acetylated
RPL23A	MS	cleaved	Acetylated
RPL39	MAA	cleaved	non-acetylated
RPL4A	MS	cleaved	Acetylated
RPN1	MV	cleaved	non-acetylated
RPN2	MS	cleaved	Acetylated
RPS3	MV	cleaved	non-acetylated
RPS6A	MK	uncleaved	non-acetylated
RPS7B	MS	cleaved	Acetylated
RPS8A	MG	cleaved	non-acetylated
RTG3	MMN	uncleaved	Acetylated
RVS161	MS	cleaved	Acetylated
RVS167	MS	cleaved	Acetylated
SAC6	MN	uncleaved	Acetylated *
SAC7	MP	cleaved	non-acetylated
SAR1	MAG	cleaved	non-acetylated
SCP160	MS	cleaved	Acetylated
SEC13	MV	cleaved	non-acetylated
SEC31	MV	cleaved	non-acetylated

SER1	MS	cleaved	Acetylated
SER2	MS	cleaved	Acetylated
SES1	ML	uncleaved	non-acetylated
SHE2	MS	cleaved	Acetylated
SHE3	MS	cleaved	Acetylated
SHM2	MP	cleaved	non-acetylated
SIC1	MT	cleaved	non-acetylated
SIT4	MV	cleaved	non-acetylated
SIW14	MG	cleaved	non-acetylated
SIZ1	MI	uncleaved	non-acetylated
SKO1	MS	cleaved	Acetylated
SLA1	MT	cleaved	non-acetylated
SMY1	MHW	uncleaved	non-acetylated
SRP1	MD	uncleaved	Acetylated
SRP101	MFD	uncleaved	non-acetylated
SRP14	MAN	cleaved	non-acetylated
SRP54	MV	cleaved	non-acetylated
SSA1	MS	cleaved	Acetylated
SSB1	MAE	cleaved	Acetylated
STI1	MS	cleaved	Acetylated
STU1	MS	cleaved	Acetylated
SUI2	MS	cleaved	Acetylated
SUP35	MS	cleaved	Acetylated
SUP45	MD	uncleaved	Acetylated
SYN8	MD	uncleaved	Acetylated
TEF1	MG	cleaved	non-acetylated
TEF2	MG	cleaved	non-acetylated
TIF1	MS	cleaved	Acetylated
TMA46	MP	cleaved	non-acetylated
TMT1	MS	cleaved	Acetylated
TPD3	MS	cleaved	Acetylated
TPM1	MD	uncleaved	Acetylated
TRP1	MS	cleaved	Acetylated
TSA1	MV	cleaved	non-acetylated
TUB1	MR	uncleaved	non-acetylated
TUB2	MR	uncleaved	non-acetylated
UBA1	MS	cleaved	Acetylated
UBP6	MS	cleaved	Acetylated
UFD1	MFS	uncleaved	non-acetylated
URA3	MS	cleaved	Acetylated
URA7	MK	uncleaved	non-acetylated
URA8	MK	uncleaved	non-acetylated
UTR1	MK	uncleaved	non-acetylated
VID30	MS	cleaved	Acetylated

VPS1	MD	uncleaved	Acetylated
YMR099c	MP	cleaved	non-acetylated
YVH1	MAG	cleaved	non-acetylated
ZUO1	MFS	uncleaved	non-acetylated
ZWF1	MS	cleaved	Acetylated

* MN acetylation is predicted in only 55% of cases
N-terminal sequence and predicted processing of Yeast Mitochondrial Proteins

Protein	N-terminal residues	Predicted Met cleavage	Predicted Acetylation*
AAC1	MS	Cleaved	Acetylated
AAC3	MS	Cleaved	Acetylated
AAT1	ML	Uncleaved	Unacetylated
ABF2	MN	Uncleaved	Acetylated *
ABF2	MN	Uncleaved	Acetylated *
ACH1	MT	Cleaved	Unacetylated
ACN9	MN	Uncleaved	Acetylated *
ACO1	ML	Uncleaved	Unacetylated
ACO2	ML	Uncleaved	Unacetylated
ACP1	MFR	Uncleaved	Acetylation
ADH3	ML	Uncleaved	Unacetylated
ADK2	MK	Uncleaved	Unacetylated
AEP2	MWI	Uncleaved	Unacetylated
AGC1	ME	Uncleaved	Acetylation
AIF1	MT	Cleaved	Unacetylated
AIM10	ML	Uncleaved	Unacetylated
AIM25	MHR	Uncleaved	Acetylation
AIM28	MML	Uncleaved	Unacetylated
ALD4	MFS	Uncleaved	Unacetylated
ALD5	ML	Uncleaved	Unacetylated
ARG7	MR	Uncleaved	Unacetylated
ATM1	ML	Uncleaved	Unacetylated
ATP10	MQ	Uncleaved	Unacetylated
ATP12	ML	Uncleaved	Unacetylated
ATP18	ML	Uncleaved	Unacetylated
ATP22	ML	Uncleaved	Unacetylated
ATP25	MN	Uncleaved	Acetylated *
ATP6	MFN	Uncleaved	Acetylation
BAT1	ML	Uncleaved	Unacetylated
CBP1	MFL	Uncleaved	Unacetylated

CBP2	MV	Cleaved	Unacetylated
CBP3	MMS	Uncleaved	Unacetylated
CBP4	MQ	Uncleaved	Unacetylated
CBP6	MS	Cleaved	Acetylated
CBS1	ML	Uncleaved	Unacetylated
CBS2	MS	Cleaved	Acetylated
CCE1	MS	Cleaved	Acetylated
CCP1	MT	Cleaved	Unacetylated
CEM1	MS	Cleaved	Acetylated
CHA1	MS	Cleaved	Acetylated
CIR2	MI	Uncleaved	Unacetylated
CLD1	MFK	Uncleaved	Acetylation
COA1	MML	Uncleaved	Unacetylated
COQ1	MFQ	Uncleaved	Unacetylated
COQ10	MV	Cleaved	Unacetylated
COQ2	MFI	Uncleaved	Unacetylated
COS111	MS	Cleaved	Acetylated
COX11	MI	Uncleaved	Unacetylated
COX14	MS	Cleaved	Acetylated
COX16	MS	Cleaved	Acetylated
COX18	ML	Uncleaved	Unacetylated
COX2	ML	Uncleaved	Unacetylated
COX20	MR	Uncleaved	Unacetylated
COX23	ME	Uncleaved	Acetylation
COX4	ML	Uncleaved	Unacetylated
CPR3	MFK	Uncleaved	Acetylation
CRC1	MS	Cleaved	Acetylated
CTP1	MS	Cleaved	Acetylated
CYC2	ML	Uncleaved	Unacetylated
DEM1	ML	Uncleaved	Unacetylated
DIC1	MS	Cleaved	Acetylated
DIN7	MG	Cleaved	Unacetylated
ERV1	MK	Uncleaved	Unacetylated
FLX1	MV	Cleaved	Unacetylated
FMC1	MD	Uncleaved	Acetylation
FMT1	MV	Cleaved	Unacetylated
GGC1	MP	Cleaved	Unacetylated
GLO4	MK	Uncleaved	Unacetylated
GUF1	ML	Uncleaved	Unacetylated
GUT2	mfs	Uncleaved	Unacetylated
HFA1	MR	Uncleaved	Unacetylated
HMI1	MD	Uncleaved	Acetylation
HOT13	MI	Uncleaved	Unacetylated
HSP10	MS	Cleaved	Acetylated

HSP60	ML	Uncleaved	Unacetylated
HTD2	MK	Uncleaved	Unacetylated
IBA57	MFI	Uncleaved	Unacetylated
ICP55	ML	Uncleaved	Unacetylated
IDH2	ML	Uncleaved	Unacetylated
IDP1	MS	Cleaved	Acetylated
IFM1	ML	Uncleaved	Unacetylated
IMG1	MWS	Uncleaved	Unacetylated
IMG2	MI	Uncleaved	Unacetylated
ISA1	MI	Uncleaved	Unacetylated
ISC1	MY	Uncleaved	Unacetylated
ISD11	MP	Cleaved	Unacetylated
ISM1	MK	Uncleaved	Unacetylated
ISU1	ML	Uncleaved	Unacetylated
LEU5	MT	Cleaved	Unacetylated
LEU9	MV	Cleaved	Unacetylated
LPE10	MR	Uncleaved	Unacetylated
MAE1	ML	Uncleaved	Unacetylated
MAM33	MFL	Uncleaved	Unacetylated
MAS1	MFS	Uncleaved	Unacetylated
MCR1	MFS	Uncleaved	Unacetylated
MCT1	MK	Uncleaved	Unacetylated
MCX1	ML	Uncleaved	Unacetylated
MDH1	ML	Uncleaved	Unacetylated
MDL1	MI	Uncleaved	Unacetylated
MDL2	ML	Uncleaved	Unacetylated
MDM31	MS	Cleaved	Acetylated
MDM32	ML	Uncleaved	Unacetylated
MDM34	MS	Cleaved	Acetylated
MDM35	MG	Cleaved	Unacetylated
MDM38	ML	Uncleaved	Unacetylated
MEF1	MS	Cleaved	Acetylated
MEF2	MWK	Uncleaved	Acetylated
MGE1	MR	Uncleaved	Unacetylated
MGM1	MN	Uncleaved	Acetylated *
MIA40	ML	Uncleaved	Unacetylated
MIC14	MS	Cleaved	Acetylated
MIC17	MAR	Cleaved	Unacetylated
MIR1	MS	Cleaved	Acetylated
MIS1	ML	Uncleaved	Unacetylated
MMF1	MFL	Uncleaved	Unacetylated
MNE1	MK	Uncleaved	Unacetylated
MPM1	MG	Cleaved	Unacetylated
MRF1	MWL	Uncleaved	Unacetylated

MRH4	MS	Cleaved	Acetylated
MRM2	MI	Uncleaved	Unacetylated
MRP49	MS	Cleaved	Acetylated
MRPL1	ML	Uncleaved	Unacetylated
MRPL10	MK	Uncleaved	Unacetylated
MRPL11	ML	Uncleaved	Unacetylated
MRPL13	MS	Cleaved	Acetylated
MRPL15	ME	Uncleaved	Acetylation
MRPL16	MFP	Uncleaved	Unacetylated
MRPL19	MS	Cleaved	Acetylated
MRPL2	MWN	Uncleaved	Acetylated
MRPL20	MI	Uncleaved	Unacetylated
MRPL22	MN	Uncleaved	Acetylated *
MRPL23	MS	Cleaved	Acetylated
MRPL24	MQ	Uncleaved	Unacetylated
MRPL25	MS	Cleaved	Acetylated
MRPL27	MK	Uncleaved	Unacetylated
MRPL28	ML	Uncleaved	Unacetylated
MRPL3	MG	Cleaved	Unacetylated
MRPL31	MFG	Uncleaved	Unacetylated
MRPL32	MN	Uncleaved	Acetylated *
MRPL35	ML	Uncleaved	Unacetylated
MRPL36	ML	Uncleaved	Unacetylated
MRPL37	ML	Uncleaved	Unacetylated
MRPL38	MI	Uncleaved	Unacetylated
MRPL39	MV	Cleaved	Unacetylated
MRPL4	MWK	Uncleaved	Acetylated
MRPL40	MS	Cleaved	Acetylated
MRPL41	MP	Cleaved	Unacetylated
MRPL44	MI	Uncleaved	Unacetylated
MRPL49	ML	Uncleaved	Unacetylated
MRPL50	ML	Uncleaved	Unacetylated
MRPL51	MV	Cleaved	Unacetylated
MRPL6	MS	Cleaved	Acetylated
MRPL7	MQ	Uncleaved	Unacetylated
MRPL8	MT	Cleaved	Unacetylated
MRPL9	MS	Cleaved	Acetylated
MRPS12	ML	Uncleaved	Unacetylated
MRPS16	MS	Cleaved	Acetylated
MRPS17	MAR	Cleaved	Unacetylated
MRPS18	ML	Uncleaved	Unacetylated
MRPS28	MS	Cleaved	Acetylated
MRPS35	MS	Cleaved	Acetylated
MRPS5	MFK	Uncleaved	Acetylation

MRPS9	MFS	Uncleaved	Unacetylated
MRS2	MN	Uncleaved	Acetylated *
MSD1	ML	Uncleaved	Unacetylated
MSE1	MI	Uncleaved	Unacetylated
MSF1	MFL	Uncleaved	Unacetylated
MSK1	MN	Uncleaved	Acetylated *
MSM1	MQ	Uncleaved	Unacetylated
MSP1	MS	Cleaved	Acetylated
MSR1	MFG	Uncleaved	Unacetylated
MSS1	MN	Uncleaved	Acetylated *
MSS2	MQ	Uncleaved	Unacetylated
MST1	MK	Uncleaved	Unacetylated
MSW1	MS	Cleaved	Acetylated
MSY1	ML	Uncleaved	Unacetylated
MTF1	MS	Cleaved	Acetylated
MTF2	MI	Uncleaved	Unacetylated
MTG1	MHI	Uncleaved	Unacetylated
MTM1	MS	Cleaved	Acetylated
MTO1	ML	Uncleaved	Unacetylated
NAM2	ML	Uncleaved	Unacetylated
NAM9	MP	Cleaved	Unacetylated
NDE1	MI	Uncleaved	Unacetylated
NDE2	ML	Uncleaved	Unacetylated
NUC1	MC	Cleaved	Unacetylated
OAC1	MS	Cleaved	Acetylated
OAR1	MHY	Uncleaved	Unacetylated
OCT1	ML	Uncleaved	Unacetylated
ODC1	MT	Cleaved	Unacetylated
ODC2	MS	Cleaved	Acetylated
OGG1	MS	Cleaved	Acetylated
OMS1	MI	Uncleaved	Unacetylated
ORT1	ME	Uncleaved	Acetylation
OXA1	MFK	Uncleaved	Acetylation
PCP1	MS	Cleaved	Acetylated
PDA1	ML	Uncleaved	Unacetylated
PDH1	MFL	Uncleaved	Unacetylated
PET111	ML	Uncleaved	Unacetylated
PET112	ML	Uncleaved	Unacetylated
PET117	MS	Cleaved	Acetylated
PET122	ML	Uncleaved	Unacetylated
PET123	MG	Cleaved	Unacetylated
PET20	ML	Uncleaved	Unacetylated
PET494	MHL	Uncleaved	Unacetylated
PET54	MK	Uncleaved	Unacetylated

PET9	MS	Cleaved	Acetylated
PGK1	MS	Cleaved	Acetylated
PIC2	ME	Uncleaved	Acetylation
PIM1	ML	Uncleaved	Unacetylated
PKP1	MWK	Uncleaved	Acetylated
PKP2	MS	Cleaved	Acetylated
PNT1	MD	Uncleaved	Acetylation
POS5	MFV	Uncleaved	Unacetylated
PPA2	MN	Uncleaved	Acetylated *
PPE1	MG	Cleaved	Unacetylated
PPT2	MS	Cleaved	Acetylated
PRX1	MFS	Uncleaved	Unacetylated
PTC5	MS	Cleaved	Acetylated
PTC6	MR	Uncleaved	Unacetylated
PUS2	ML	Uncleaved	Unacetylated
PUS9	MQ	Uncleaved	Unacetylated
QRI5	ML	Uncleaved	Unacetylated
QRI7	MI	Uncleaved	Unacetylated
RIM1	MFL	Uncleaved	Unacetylated
RIM2	MP	Cleaved	Unacetylated
RMD9	MML	Uncleaved	Unacetylated
RML2	ML	Uncleaved	Unacetylated
RPO41	ML	Uncleaved	Unacetylated
RRF1	MI	Uncleaved	Unacetylated
RSM10	ML	Uncleaved	Unacetylated
RSM28	MR	Uncleaved	Unacetylated
SAL1	ML	Uncleaved	Unacetylated
SEN15	MAN	Cleaved	Unacetylated
SEN2	MS	Cleaved	Acetylated
SEN34	MP	Cleaved	Unacetylated
SEN54	MQ	Uncleaved	Unacetylated
SFC1	MS	Cleaved	Acetylated
SHE9	ML	Uncleaved	Unacetylated
SHM1	MFP	Uncleaved	Unacetylated
SHY1	MS	Cleaved	Acetylated
SLM5	MFH	Uncleaved	Unacetylated
SLS1	MWK	Uncleaved	Acetylated
SSC1	ML	Uncleaved	Unacetylated
SSQ1	ML	Uncleaved	Unacetylated
STP1	MS	Cleaved	Acetylated
SUE1	MI	Uncleaved	Unacetylated
TAM41	ML	Uncleaved	Unacetylated
TAR1	MR	Uncleaved	Unacetylated
THI74	MN	Uncleaved	Acetylated *

TIM10	MS	Cleaved	Acetylated
TIM11	MS	Cleaved	Acetylated
TIM12	MS	Cleaved	Acetylated
TIM8	MS	Cleaved	Acetylated
TIM9	MD	Uncleaved	Acetylation
TPC1	MFK	Uncleaved	Acetylation
TRR2	MI	Uncleaved	Unacetylated
TRX3	ML	Uncleaved	Unacetylated
TUF1	MS	Cleaved	Acetylated
TUM1	MP	Cleaved	Unacetylated
UPS1	MV	Cleaved	Unacetylated
UPS2	MK	Uncleaved	Unacetylated
UPS3	MK	Uncleaved	Unacetylated
VAR1	MK	Uncleaved	Unacetylated
VAS1	MN	Uncleaved	Acetylated *
VPS73	MN	Uncleaved	Acetylated *
YDL119c	MT	Cleaved	Unacetylated
YEA6	MN	Uncleaved	Acetylated *
YFH1	MI	Uncleaved	Unacetylated
YFR045w	MAN	Cleaved	Unacetylated
YHM2	MP	Cleaved	Unacetylated
YIA6	MT	Cleaved	Unacetylated
YLH47	ML	Uncleaved	Unacetylated
YLR164W	MS	Cleaved	Acetylated
YMC1	MS	Cleaved	Acetylated
YMC2	MS	Cleaved	Acetylated
YML6	MT	Cleaved	Unacetylated
YMR166c	MN	Uncleaved	Acetylated *
YPR011c	MAE	Cleaved	Acetylation
YPR021c	ME	Uncleaved	Acetylation
YSP1	MHE	Uncleaved	Unacetylated

* MN acetylation is predicted in ony 55% of cases

Relative Amino Acid Frequency at Position 2 by Compartment in Yeast

P2 Residue	Signal Sequence	Cytosol	Mitochondria	f(ss)/f(cyt)
A	3.25	10.32	1.77	0.317
C	0.00	0.00	0.35	∞
D	0.72	6.75	1.42	0.108
E	0.72	4.37	2.13	0.167
F	6.50	1.19	10.28	4.887
G	0.72	5.56	2.13	0.131
H	2.17	0.79	1.77	2.749
I	6.86	0.79	6.74	8.705
K	16.61	3.17	4.96	5.269
L	13.72	3.17	23.05	4.353
M	1.44	1.19	1.42	1.222
N	3.61	2.38	5.67	1.527
P	1.44	5.56	2.84	0.262
Q	10.11	1.98	2.84	5.132
R	13.00	1.98	2.84	6.598
S	5.78	35.32	20.57	0.165
T	4.33	7.54	3.19	0.579
V	7.58	7.54	2.84	1.013
W	1.08	0.00	2.84	∞
Y	0.36	0.40	0.35	0.916
Total	100.00	100.00	100.00	
n	277	251	281	

Predictied Relative Frequency of N-terminal methionine cleavage

	Signal Sequence	Cytosol	Mitochondria
cleaved	23.10	71.83	33.69
uncleaved	76.90	28.17	66.31

Relative P2 Frequency of Signal Sequences from different Organisms

P2 residue	Human	*C. elegans*	*S. cerevisiae*
A	18.99	1.59	2.94
C	0.34	1.85	0.00
D	1.85	1.06	1.10
E	5.04	1.32	1.10
F	2.02	4.23	6.62
G	9.41	2.12	0.74
H	1.01	2.65	2.21
I	3.03	7.41	6.62
K	9.92	12.96	16.18
L	8.07	7.94	11.40
M	1.51	2.38	1.84
N	1.68	8.47	4.78
P	2.86	2.65	1.10
Q	3.70	4.23	11.76
R	13.28	23.54	13.24
S	5.21	5.82	5.51
T	2.69	5.56	4.04
V	6.05	1.85	7.72
W	2.18	1.32	0.74
Y	1.18	1.06	0.37
n	595	378	277

Predicted Frequency of N-terminal processing of Signal Sequences from different organisms

	Predicted N-terminal Processing				
Organism	Unprocessed	Met-cleaved	Uncleaved & acetylated	Met-cleaved & acetylated	n
S. ceverevisiae	68	5	9	18	277
C. elegans	68	15	11	6	378
Human	46	23	23	8	595

Peak hydrophobicity of yeast signal sequences

	Met-unprocessed	Met-processed
n	178	90
Mean peak hydrophobicity (*)	2.333	2.548
SEM	0.0352	0.0484
Fraction with peak h<2 (%)	22.5	10
t-test (two-tailed)	t=3.552 df=266	
p-value	0.0005	

* Determined according to Kyte & Doolittle (1982) with a window size of 11