Bioinformatics Advance Access published July 15, 2004

Preferred *in vivo* ubiquitination sites

André Catic[1, 4], Cal Collins[2], George M. Church[3] and Hidde L. Ploegh[1]

[1] Department of Pathology, Harvard Medical School,
77 Avenue Louis Pasteur,
Boston, MA 02115, USA

[2] Akaza Research, Inc.,
56 John F. Kennedy Street,
Cambridge, MA 02138, USA

[3] Lipper Center for Computational Genetics and Department of Genetics,
Harvard Medical School,
77 Avenue Louis Pasteur,
Boston, MA 02115, USA

[4] Address correspondence to: A. Catic
Department of Pathology, Ploegh Laboratory,
77 Avenue Louis Pasteur, NRB Room 836,
Boston, MA 02115, USA.
Phone: +001 617 432 4779. Fax: +001 617 432 4775.
Email: Catic@fas.harvard.edu

Keywords: Ubiquitin, Motif, Preference

ABSTRACT

Motivation: The conjugation of ubiquitin to target molecules involves several enzymatic steps. Little is known about the specificity of ubiquitination. How E3 ligases select their substrate and which lysines are targeted for ubiquitin conjugation is largely an enigma. The object of this study is to identify preferred ubiquitination sites. Genetic approaches to study this question have proven difficult, because of the redundancy of ligases and the lack of strictly required motifs. However, a better understanding of acceptor site selection could help predict ubiquitination sites and clarify yet unsolved structure-function relationships of the transfer reaction.

Results: In an effort to define preferences for ubiquitination, we systematically analyzed structure and sequence of 135 known ubiquitination sites in 95 proteins in *Saccharomyces cerevisiae*. The results show clear structural preferences for ubiquitin ligation to target proteins, and compartment-specific amino acid patterns in close proximity to the modified side chain.

Contact: Catic@fas.harvard.edu

Supplementary information: http://www.people.fas.harvard.edu/~catic

INTRODUCTION

Ubiquitination is a process that requires three distinct catalytic activities. First, ubiquitin is activated by an E1 enzyme, then conjugated to an E2 enzyme and finally transferred to a substrate molecule with the assistance of an E3 ligase, forming an isopeptide bond with an internal lysine of the target protein. In yeast, there is a single E1, 11 E2 enzymes and possibly hundred or more E3 proteins, which can be divided into RING or HECT domain-containing enzymes (Weissman, 2001). There may be a specific E2/E3 combination for ubiquitination of a given substrate and several motifs have been described that are required for formation of a functional complex of the conjugation machinery and the target molecule. However, the acceptor lysine residue is usually not involved in these binding interactions (Pickart, 2001). Genetic studies suggest that there is no preference as to which lysine of a target protein is modified by its cognate E2/E3 complex. Nonetheless, these investigations do not address ubiquitination of an endogenously expressed native substrate. While there is the potential for combinatorial diversity in terms of binding of E2/E3 enzymes to their specific substrate, the catalytic transfer itself is limited to fewer options. It is either the HECT domain of respective E3 enzymes or one of 11 E2 polypeptides that are capable of forming a stable isopeptide bond between ubiquitin and the acceptor protein. A BLASTP search for possible HECT domain-containing sequences in *Saccharomyces cerevisiae* reveals 5 potential HECT E3 enzymes (Altschul et al., 1990). Multiple alignment analysis of yeast E2 proteins shows that some of these polypeptides can be grouped into subfamilies (Haas and Siepmann, 1997), based on the conservation of their catalytic core (Figure 1). Combined, these findings suggest that ubiquitination in yeast is achieved by only few catalytic core sequences.

The ubiquitin-like molecule SUMO is conjugated to a substrate protein by a single E2, Ubc9. The catalytic cleft of this enzyme favors interactions with a particular sequence in the target protein, and this motif can be used to predict likely sumoylation sites (Bernier-Villamor et al., 2002). We hypothesized that such a preference might also exist for ubiquitination, to reflect the conserved catalytic mechanism. A correlation between the catalytic core sequence of E2 ligases and the preferred site of modification in a target protein is also implied by the fact that the SUMO ligase Ubc9 and its homologous ubiquitin ligase Ubc2 (synonym: RAD6) compete for altering the same lysine residue in the substrate polypeptide PCNA (Hoege et al., 2002). Moreover, since E2's are known to be compartmentalized (Table 1), we were especially interested in analyzing ubiquitination sites, according to the subcellular localization of the target proteins in question.

METHODS

Dataset: The majority of the dataset analyzed was published in two recent articles (Peng et al., 2003; Hitchcock et al., 2003), in which ubiquitination sites of yeast proteins were identified by tandem mass spectrometry. The authors employed multidimensional protein identification technology (MudPIT), which enables semi-quantitative and unbiased analysis of posttranslational modification sites in proteins (Zhou et al., 2001; Lin et al., 2003). For 91 of the proteins analyzed the modified lysine was identified (four additional proteins were published elsewhere). The ubiquitination sites in this dataset may not reflect all possible ubiquitin-modified lysines in the examined substrates, but likely represent the most abundant, hence the preferred ones. The source materials were cells not exposed to proteasome inhibitors and therefore short-lived proteins may have escaped detection. Given the diversity of source proteins, we consider a bias of the dataset caused by separation methods unlikely.

Structure analysis: We could retrieve structural information on 23 of the 95 proteins in our study, representing a total of 40 ubiquitination sites. Only structures of yeast proteins or close homologues were considered. The structure files were downloaded from the Protein Data Bank and analyzed using the Swiss PDB Viewer (Guex and Peitsch, 1997; Berman et al., 2000).

Compartmentalization: The assignment of proteins to cellular compartments was conducted according to a recent publication, in which 4,156 yeast proteins were tagged with GFP and their subcellular distribution was assessed using fluorescence microscopy (Huh et al., 2003). For the sake of a strict definition, we considered only the plasma membrane and the nucleus. Proteins of ambiguous localization were excluded. The assignment of every protein was verified based on published function or subcellular distribution (www.yeastgenome.org). False positive or negative hits in the GFP screen were excluded. To distinguish between plasma membrane-bound and plasma membrane-associated proteins, we scored the hydrophobicity index with web-based algorithms (Horton and Nakai, 1997; Hirokawa et al., 1998). Potential phosphorylation sites in yeast were evaluated with NetPhos (Blom et al., 1999).

Proteome comparison: 4,963 published *Saccharomyces cerevisiae* proteins (http://us.expasy.org/cgi-bin/lists?yeast.txt) were analyzed and the two flanking amino acids at either side of all the 175,137 lysines were scored for their frequency (terminal lysines with less then 4 flanking partners were excluded). All calculations are based on the χ^2 test with Yates' correction (one degree of freedom).

Supplementary information can be found on our website:
http://www.people.fas.harvard.edu/~catic

RESULTS AND DISCUSSION

Structural preference

With one exception (K370 in GDH1), all ubiquitin-modified lysines are exposed at the surface of the molecule and are readily accessible from the outside. However, ubiquitin modification of the buried lysine in GDH1 must require prior unfolding of the protein. This position might be involved in degradation of misfolded GDH1 molecules. K370 is located in a hydrophobic stretch PPKAA. Note that in our dataset, only one other sequence contains four non-polar amino acids that surround the modified lysine (K506 in ACS2).

There is a clear preference for ubiquitination within loops (26/40), followed by α helices (10/40) (Table 2). Two of the four modification sites in β strands appear in the ubiquitin precursor RPL40A. However, the prevailing sites for poly-ubiquitin chain formation in the ubiquitin sequence are K48 and K63 (Weissman, 2001), both located in loops. We consider the dominance of loops as ubiquitination sites significant, since lysines have a slight propensity to occur within α helices, in general (Williams et al., 1987; Wilmot and Thornton, 1988) as well as in 12 of the 23 proteins analyzed. If ubiquitin attachment in these 23 proteins were to occur without preference for a particular structural element, we would have expected 17.1 modified lysines in α helices, 17.4 in loops and 5.5 in β strands, based on the overall distribution of lysines in this data set. The deviation of the observed occurrence in loops from the expected value is highly significant ($p=0.0097$).

Bias in the primary protein structure

We examined the two positions flanking either side of the 135 modified lysines in our dataset, looking for potential ubiquitination motifs. For comparison, we used the frequency of a given amino acid at the respective flanking position of all the 175,137 non-terminal lysines in the yeast proteome, as published in the Swiss-Prot database. We found that in a di-lysine sequence, the downstream lysine is significantly more frequently ubiquitinated (3.3 fold, $p=0.0079$). Furthermore, cysteine was altogether absent from the analyzed positions ($p=0.017$).

One possible explanation for this absence may be the high reactivity of the cysteine sulfhydryl, interfering with the catalytic core of ubiquitin ligases. An alternative explanation is that the sites of ubiquitination are also possible locations of other posttranslational modifications at the ε-NH_2 group, such as methylation, acetylation, sumoylation, which could be adversely affected by a cysteine side chain. Yet another possible explanation lies in the reversibility of ubiquitin ligation: deubiquitinating enzymes (DUBs) cleave off ubiquitin and may rescue a protein from degradation. The catalytic core of these enzymes, which mostly have a cysteine in their active site, could also be affected by a reactive sulfhydryl side chain in the substrate.

In addition to looking at amino acids immediately adjacent to the modified lysine, we also searched for more ambiguous sequence motifs involving the target lysine residue. PRATT (Jonassen et al., 1995) was used as motif-finding algorithm, and delivered the highest score for one lysine-containing motif, KEEE. Upon further examination, we discovered that 4 out of 5 KEEE patterns in our dataset of 95 proteins are ubiquitinated, which indicates a permissiveness of this motif (22 fold compared to an average lysine in our dataset, $p=2.5 \times 10^{-15}$). Since the four ubiquitinated KEEE motifs occur in genetically and functionally unrelated proteins (CUE5, Ena5, Rps3, YGR268C), we suggest that this sequence is a generally preferred site for attachment of ubiquitin in yeast.

Compartment-specific motifs

We then divided the ubiquitin-conjugated proteins into nuclear and plasma membrane-associated polypeptides. Proteins in the nucleus show a 1.7 fold higher occurrence of phosphorylatable side chains close (within two amino acids) to the ubiquitinated lysine (Table 3) when compared to the entire yeast proteome ($p=0.03$), and a 2.3 fold higher occurrence when compared to the four flanking positions of 90 known ubiquitination sites of non-nuclear proteins in our dataset ($p=0.0003$). Although several side chains do not yield a high score for phosphorylation, it should be noted that the prediction algorithms rely on known consensus sequences for well-studied kinases, and low scores do not exclude phosphorylation *in vivo*. Algorithms to predict O-linked glycosylation at Ser and Thr are currently not available for yeast. We furthermore observed an accumulation of amino acids with polar uncharged residues in the nuclear subset (Ser, Thr, Asn, and Gln; $p=0.04$ compared with proteome and $p=0.019$ compared with non-nuclear ubiquitination sites).

The importance of phosphorylation of Ser, Thr and Tyr, and their interplay with modifications at the ε-NH_2 group of lysine have been recognized previously in well-studied examples, like histones and the degradation of SIC1 (Nash et al., 2001; Petroski and Deshaies, 2003; Fischle et al., 2003). Our data suggests that the interconnection of phosphorylation and ubiquitination is not necessarily restricted to specific proteins, but may include a compartmentalization theme. Of particular interest is the comparison of sequences surrounding known ubiquitination sites of proteins from different compartments. Such an evaluation also takes into account any coincidental correlations that might arise if phosphorylation sites and ubiquitination sites independently shared similar preferences concerning e.g. tertiary and secondary structures. Our data shows that this is not the case, and only nuclear ubiquitination sites display the characteristics noted above.

Combined, our data suggest that native, not necessarily unfolded proteins can be the substrate for ubiquitin conjugation, and as shown for multiple examples (Pickart, 2001), this recognition might require prior phosphorylation (or dephosphorylation).

The ubiquitination sites at the plasma membrane (Table 4) have a 4.4 fold increase of Asp and Glu at position "-2" compared against the entire yeast proteome ($p=1.3 \times 10^{-10}$). Comparison with the respective position of known ubiquitination sites of 57 proteins not associated with membranes yields a 6.1 fold increase of Asp and Glu ($p=8.5 \times 10^{-16}$). Enrichment of Asp and Glu at position "-1" is 4.1 fold ($p=6.3 \times 10^{-9}$) compared with the proteome and 2.4 fold ($p=0.0004$) compared to ubiquitination sites of proteins in our dataset that are not associated with membranes. Of particular significance is the occurrence of two consecutive acidic amino acids at the "-2" and "-1" position (17 fold enhancement compared to proteome, $p=4.6 \times 10^{-25}$, and no occurrence in ubiquitination sites of non-membrane-associated proteins).

The fact that transmembrane proteins and plasma membrane-associated proteins show a similar pattern, suggests that the preference we observe may be compartment-associated. In accordance with our observation, the E2 enzymes Ubc4 and Ubc5, and the HECT E3 RSP5 have been shown to be responsible for modification of several plasma membrane receptors and transporters (Pickart, 2001; Weissman, 2001). Interestingly, the crystal structure of yeast Ubc4 reveals a positively charged protrusion, created by the ε-NH_2 of Lys91, adjacent to the catalytic core Cys86 (Figure 2). Side chains of basic and acid amino acids are charged at neutral pH, and it is conceivable that Lys91 in Ubc4 acts as a hook and uses ionic interactions with the carboxyl groups of Asp or Glu to position the target sequence in proximity to Cys86. A positively charged group at the relative position of Lys91 is highly conserved in Ubc4 and Ubc5 homologues across species.

For all plasma membrane proteins listed, we could retrieve detailed information on additional ubiquitination sites in GAP1 and STE6 (Soetens et al., 2001; Kolling, 2002). Genetic studies show that the sequence DAKTI is a preferred site of modification in STE6, while in our dataset the sequence DEKHN is modified. Our results suggest a more general theme, namely the presence of an acidic amino acid at position "-2" that is important. Another possibly relevant side chain in this particular example is the phosphorylatable amino acid at position "+1" (histidine kinases have been described in yeast; Santos and Shiozaki, 2001). The permease GAP1 has an additional ubiquitination site at K9, which is important for endocytosis. This target lysine is also flanked by an acidic residue at position "-1".

In summary, it is likely that as more data emerge, motifs for ubiquitin modification can be further refined.

Acknowledgements

We thank G. Weber, B.M. Kessler, H. Ovaa, J.W. Harper, and the reviewers for helpful comments.

References

Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215,** 403-410.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourna,P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235-242.

Bernier-Villamor,V., Sampson,D.A., Matunis,M.J. and Lima,C.D. (2002) Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell*, **108**, 345-356.

Blom,N., Gammeltoft,S. and Brunak,S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.*, **294**, 1351-1362.

Cook,W.J., Jeffrey,L.C., Xu,Y. and Chau,V. (1993) Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: crystal structure of yeast Ubc4. *Biochemistry (Washington)*, **32**, 13809-13817.

Fischle,W., Wang,Y. and Allis,C.D. (2003) Binary switches and modification cassettes in histone biology and beyond. *Nature*, **425**, 475-479.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714-2723.

Haas,A.L. and Siepmann,T.J. (1997) Pathways of ubiquitin conjugation. *FASEB J.*, **11**, 1257-1268.

Hirokawa,T., Boon-Chieng,S. and Mitaku,S. (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics*, **14**, 378-379.

Hitchcock,A.L., Auld,K., Gygi,S.P. and Silver,P.A. (2003) A subset of membrane-associated proteins is ubiquitinated in response to mutations in the endoplasmic reticulum degradation machinery. *Proc. Natl. Acad. Sci. USA*, **100**, 12735-12740.

Hoege,C., Pfander,B., Moldovan,G.L., Pyrowolakis,G. and Jentsch,S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature*, **419**, 135-141.

Horton,P. and Nakai,K. (1997) Better prediction of protein cellular localization sites with the k nearest neighbors classifier. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, **5**, 147-152.

Huh,W.K., Falvo,J.V., Gerke,L.C., Carroll,A.S., Howson,R.W., Weisman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686-691.

Jonassen,I., Collins,J.F. and Higgins,D.G. (1995) Finding flexible patterns in unaligned protein sequences. *Protein Sci.*, **4**, 1587-1595.

Kolling,R. (2002) Mutations affecting the phosphorylation, ubiquitination and turnover of the ABC-transporter Ste6. *FEBS Lett.*, **531**, 548-552.

Lin,D., Tabb,D.L. and Yates, J.R., 3rd (2003) Large-scale protein identification using mass spectrometry. *Biochim. Biophys. Acta*, **1646**, 1-10.

Nash,P., Tang,X., Orlicky,S., Chen,Q., Gertler,F.B., Mendenhall,M.D., Sicheri,F., Pawson,T. and Tyers,M. (2001) Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature*, **414**, 514-521.

Peng,J., Schwartz,D., Elias,J.E., Thoreen,C.C., Cheng,D., Marsischky,G., Roelofs,J., Finley,D. and Gygi,S.P. (2003) A proteomics approach to understanding protein ubiquitination. *Nat. Biotechnol.*, **21**, 921-926.

Petroski,M.D. and Deshaies,R.J. (2003) Context of multiubiquitin chain attachment influences the rate of Sic1 degradation. *Mol. Cell*, **11**, 1435-1444.

Pickart,C.M. (2001) Mechanisms underlying ubiquitination. *Annu. Rev. Biochem.*, **70**, 503-533.

Santos,J.L. and Shiozaki,K. (2001) Fungal histidine kinases. *Sci. STKE*, 2001**(98)**, RE1.

Soetens,O., DeCraene,J.O. and Andre,B. (2001) Ubiquitin is required for sorting to the vacuole of the yeast general amino acid permease, Gap1. *J. Biol. Chem.*, **276**, 43949-43957.

Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673-4680.

Weissman,A.M. (2001) Themes and variations on ubiquitylation. *Nat. Rev. Mol. Cell Biol.*, **2**, 169-178.

Williams,R.W., Chang,A., Juretic,D. and Loughran,S. (1987) Secondary structure predictions and medium range interactions. *Biochim. Biophys. Acta*, **916**, 200-204.

Wilmot,C.M. and Thornton,J.M. (1988) Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.*, **203**, 221-232.

Zhou,H., Watts,J.D. and Aebersold,R. (2001) A systematic approach to the analysis of protein phosphorylation. *Nat. Biotechnol.*, **19**, 375-378.

Fig. 1: Alignment of the catalytically active region of E2's in *Saccharomyces cerevisiae*. The putative ubiquitin acceptor cysteine in the catalytic cleft is shaded. The SUMO specific E2 Ubc9 has similarity to the ubiquitin E2 Ubc2. The crystal structure of Ubc9 is instructive for understanding the biochemistry of SUMO transfer, and residues known to participate in interaction with the acceptor site of sumoylated proteins are printed in red (Thompson et al., 1994; Bernier-Villamor, 2002).

Table 1: Ubiquitin specific E2 proteins and their confirmed intracellular localization in *Saccharomyces cerevisiae* (Saccharomyces Genome Database; www.yeastgenome.org).

Table 2: 40 ubiquitination sites from 23 proteins, listed according to their occurrence within secondary structure (β strand, α helix, loop). Indicated are the frequencies of ubiquitination sites identified in the given protein structure.

Table 3: Alignment of nuclear proteins, centered on the ubiquitin-modified lysine. Ser, Thr and Tyr (bold) are significantly enriched in the flanking positions. These amino acids can be phosphorylated by appropriate kinases. Side chains with a predicted high probability of phosphorylation, as assessed by comparison with defined phosphorylation sites, are centered. Green = polar residues, blue = basic side chains, red = acidic side chains.

Table 4: Plasma membrane-bound and -associated proteins. The frequency of acidic amino acids is significantly increased, particularly N terminal of the modified lysine. To distinguish between plasma membrane-bound and plasma membrane-associated (=assoc) proteins, we scored the hydrophobicity index. The comparison with known ubiquitination sites of non-membrane-associated proteins was conducted excluding polypeptides associated with membranes of intracellular organelles.

Fig. 2: Space-filling model of yeast Ubc4 (Cook et al., 1993; Guex and Peitsch, 1997; Berman et al., 2000). The enzyme mainly consists of a conserved E2 core domain. The sulfhydryl side chain of Cys86 is exposed, while the backbone of the molecule forms a slight depression. The ε-NH_2 group of Lys91 creates a dominant protrusion on the surface.

Figure 1:

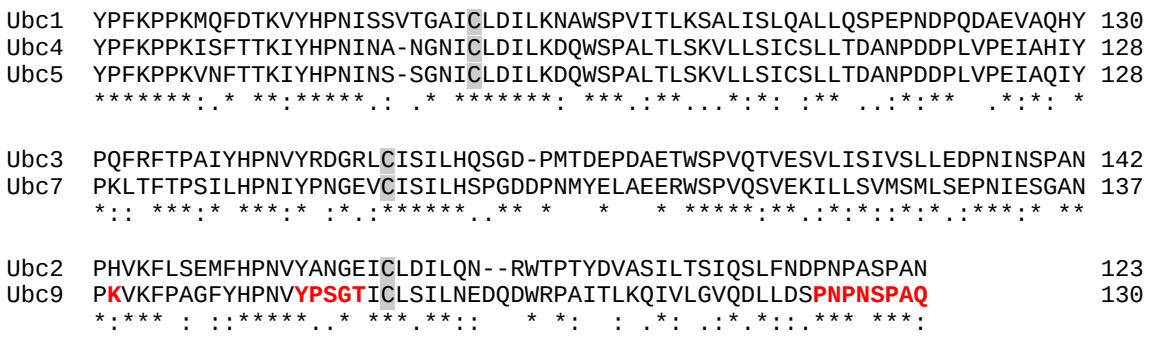

Table 1: Ubiquitin E2 ligases and their intracellular localization

E2	Confirmed Intracellular Localization
Ubc1	Cytosol and nucleus
Ubc2	Cytosol and nucleus
Ubc3	Cytosol and nucleus
Ubc4	Cytosol and nucleus
Ubc5	Cytosol and nucleus
Ubc6	ER and nuclear membrane
Ubc7	ER and nuclear membrane
Ubc8	Cytosol and nucleus
Ubc10	Peroxisomes
Ubc11	Cytosol
Ubc13	Cytosol and nucleus

Table 2: Distribution of ubiquitination sites, based on secondary structure

	ACS2	CCT8	CDC48	CIT2	ERG5	GDH1	GLU1	GPA1	H2B	HSP104	NCP1	PCNA
Strand												
Helix				1	1	1						
Loop	1	1	2	1	1	1	1	1	1	1	1	1

	PDX3	PHO84	PIN3	PMA1	PRE9	RPL40A	RPO21	SAM2	SNC1	SSA1	URA3	
Strand			1	1		2						
Helix				1		2	1	1	1		1	
Loop	1	1		1	1	3		2		2	2	

Table 3: Alignment of ubiquitination sites in nuclear proteins

CDC48	L	R	K	**T**	P
CDC48	I	A	K	A	R
CHD1	**Y**	L	K	N	L
CSR2	Q	D	K	N	H
CTR9	**Y**	Q	K	E	N
ELP3	P	K	K	D	I
HTB2	V	**T**	K	**Y**	**S**
PCNA	I	**T**	K	E	**T**
RPO21	V	**T**	K	E	A
SIC1	P	Q	K	P	**S**
YER067W	P	E	K	I	**S**
YHR097C	D	**S**	K	M	K
YHR097C	**S**	K	K	**S**	**T**
YHR097C	**T**	P	K	Q	Q
YKR041W	A	E	K	R	**Y**

Table 4: Alignment of ubiquitination sites in plasma membrane proteins

GAP1	A	E	K	V	A	
YLR413W	A	E	K	N	F	
YRO2	A	E	K	K	M	
JEN1	A	V	K	A	N	
HXT6	D	D	K	P	L	
HXT7	D	D	K	P	L	
YMR295C	D	D	K	A	R	assoc
JEN1	D	E	K	I	S	
PHO87	D	E	K	A	I	
STE6	D	E	K	H	N	
HXT4	D	M	K	D	F	
YMR295C	D	V	K	I	S	assoc
PDR12	E	D	K	V	P	
YLR414C	E	E	K	P	L	
GSC2	E	G	K	R	T	
HXT7	E	I	K	A	Y	
PHO84	E	K	K	I	H	
AKL1	E	N	K	R	H	assoc
EXO84	E	V	K	L	N	assoc
ENA5	F	G	K	E	E	
TNA1	F	N	K	E	E	
AKL1	K	D	K	D	S	assoc
HXT5	P	P	K	D	E	
GPA1	R	A	K	A	A	assoc
HXT7	S	S	K	T	K	
PDR5	T	E	K	N	A	
TAT1	T	E	K	Q	D	
PHO84	V	N	K	D	T	

Figure 2: