Allosteric Activation of the Ubiquitin Ligase UBR1 by Short Peptides: Molecular Mechanisms and Physiological Functions

Thesis by

Fangyong Du

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2002

(Defended November 21, 2001)

ACKNOWLEDGMENTS

Despite bumpy roads at the beginning of my journey, I have had a wonderful time learning to ask the right questions and design neat experiments in Alex's lab; the thrilling moments of coming up with good ideas are unforgettable.

I am very grateful to Alex for his patience to allow me to grow at my own pace, for the freedom to work independently in his lab, and for driving me to be my best. It has been a great experience to observe closely his creativity, rigor, and above all, his devotion to science. I am thankful for his time and efforts in editing my thesis. I also want to thank my thesis committee— Ray Deshaies, Bill Dunphy, Judith Campbell and Giuseppe Attardi for their advice and encouragement over the years.

I particularly thank Glenn C. Turner, my collaborator and a good personal friend, for his good ideas, in-depth discussions and excellent editorial efforts. His "refuse-to-feel-depressed" attitude has been inspiring to me over the years. In addition, Ailsa Webster helped me get started with protein purification, and her encouragement kept me going for a long way in the early days. Larry Peck showed me how to be meticulous with my experiments, and his patience in correcting my everyday English benefited me greatly. I thank Xiao-Dong Su for his abundant help, he made me understand what "a friend in need is a friend indeed" is all about.

My sincere thanks also go to Drs. Tau-Mu Yi and Scott Stevens for their numerous suggestions and reagents, past and current Varshavsky lab members: Marty Gonzalaz, Hai Rao, Takafumi Tasaki, Yong Tae Kwon, Ilia Davydov, Chris

Byrd, Jun Sheng, Youming Xie, Zan-Xian Xia, Rong-Gui (Cory) Hu – thanks for strains, plasmids, cell lines and many happy hours in the lab.

Finally, I thank my wife Irene Ping Ren and my parents in China, Shi-Wei Du and Chui-Zhu Xu, for their unconditional love and support, and for their faith in me, in good times and bad times.

ABSTRACT

The N-end rule relates the *in vivo* half life of a protein to the identity of its N-terminal residue. UBR1, the E3 of the N-end rule pathway in *Saccharomyces cerevisiae*, targets proteins that bear destabilizing N-terminal residues for Ub-dependent, processive degradation. UBR1 binds protein substrates or dipetides through two distinct sites: the type 1 site, specific for basic residues, and the type 2 site, specific for bulky hydrophobic residues. UBR1 also recognizes an internal degradation signal of the 35 kDa homeodomain protein CUP9, a transcriptional repressor of the di- and tripeptide transporter PTR2.

Here I report that the internal degradation signal of CUP9 is recognized by UBR1 through its third, distinct substrate-binding site. Occupation of the type 1 or type 2 sites of UBR1 by dipeptides allosterically stimulates the UBR1-dependent multi-ubiquitylation of CUP9 in an *in vitro* system, which consists of purified components of the yeast N-end rule pathway. UBR1 is the first E3 shown to be allosterically regulated by small compounds. This regulation underlies, *in vivo*, the accelerated UBR1-dependent degradation of CUP9 in the presence of dipeptides with destabilizing N-terminal residues. The result is a positive feedback circuit that controls the peptide import in *S. cerevisiae*. Specifically, the imported dipeptides bind to UBR1 and accelerate the UBR1-dependent degradation of CUP9, thereby derepressing the transcription of *PTR2* and increasing the cell's capacity to import peptides.

I also describe a new, autoinhibition-based molecular mechanism underlying the activation of UBR1 by dipeptides. UBR1 is an autoinhibited protein, in that the binding of dipeptides to the type 1 and type 2 sites of UBR1 enhances the dissociation of the C-terminal autoinhibitory domain of UBR1 from

its substrate-binding N-terminal region. Moreover, this dissociation, which allows the interaction between UBR1 and CUP9, is strongly increased only if both type 1 and type 2 sites of UBR1 are occupied by dipeptides. An autoinhibitory mechanism discovered in the *S. cerevisiae* UBR1 is likely to recur in metazoan homologs of UBR1, and may also be involved in controlling the activity of other Ub-dependent pathways.

TABLE OF CONTENTS

Chapter 1

The ubiquitin system and the N-end rule pathway

Introduction

Intracellular protein homeostasis involves both protein synthesis and degradation. Protein degradation eliminates damaged or misfolded proteins and controls the concentrations of many regulatory proteins. In eukaryotes, the ATP-dependent ubiquitin (Ub)/proteasome system is responsible for the bulk of selective protein degradation [1, 2]. As a result, the Ub system is essential for the control of many cellular processes, including cell cycle progression, transcriptional regulation, oxygen sensing and stress responses.

Ub is an abundant 76-residue protein that is highly conserved in eukaryotes. The essentially unchanged amino acid sequence and spatial structure of Ub from fungi to humans underscore its functional importance. In early 1980s, through biochemical dissection of the ATP-dependent proteolytic system in rabbit reticulocyte lysate, Hershko and co-workers [3-5] [6] discovered that the heat–stable Ub (initially named APF-1) was an essential component of the ATP-dependent protein degradation system. They elucidated the conserved enzymatic reactions that Ub is involved (Figure 1.1):

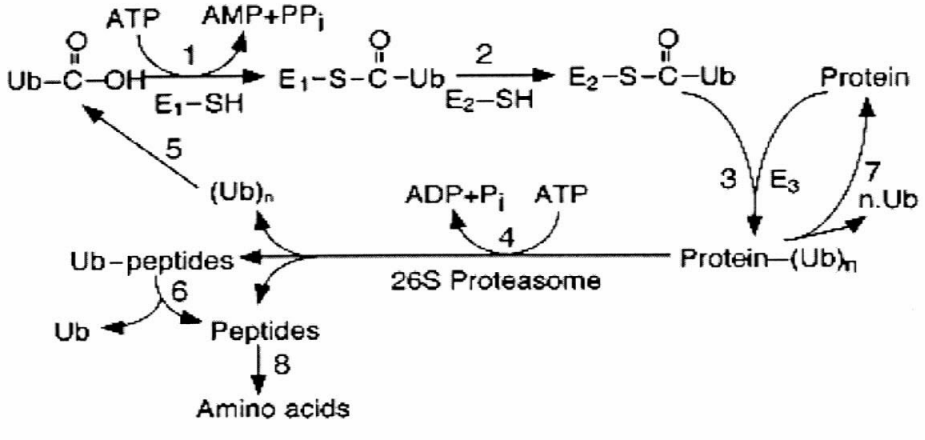


Figure 1.1 Enzymatic reactions of the ubiquitin system (See text for details) [2]

Ub is first activated by ubiquitin-activating enzyme (E1), with the C-terminal glycine of Ub forming, in an ATP-dependent reaction, a high energy thioester bond with a conserved Cys residue of E1. The activated Ub is then transferred to one of the ubiquitin-conjugating enzymes (E2) through the transesterification reaction. An E2 enzyme usually forms a complex with a specific ubiquitin ligase (E3) which recognizes the degradation signal (degron) of a protein substrate. Through coordinated actions of E2 and E3, the carboxyl group of the C-terminal Gly residue of Ub forms an isopeptide bond with the ε-amino group of a lysine residue of a protein substrate. In most cases, after the conjugation of the first Ub moiety, a substrate-linked multi-Ub chain is built processively on the substrate. In a multi-Ub chain, Ub moieties are linked through isopeptide bond between the C-terminal Gly of one Ub and the Lys48 of an adjacent Ub [7]. In at least one Ub-dependent pathway, an additional distinct component, termed E4, is required for efficient Ub chain elongation [8]. The multiubiquitylated substrate is recognized, unfolded and finally degraded to short peptides through the action of a ~2,000 kDa, ATP-dependent multicatalytic protease termed 26S proteasome. It consists of a 20S core particle and two 19S regulatory particles that cap both ends of a 20S core particle [9, 10].

Physiological functions of the Ub-dependent proteolysis were left unaddressed by the early work with the cell-free reticulocyte lysate system. In 1984, Varshavsky and co-workers discovered that the temperature-dependent cell cycle arrest of the mouse cell line ts85 is caused by a temperature-sensitive E1 [11, 12]. These results provided the first evidence that the Ub system is essential for the cell-cycle progression. They further discovered that Ub is required for the bulk of selective protein degradation in living cells [11, 12].

One of the most striking features of Ub system is its exquisite selectivity. Thus, some of the fundamental questions are: What constitutes a degron? How is it recognized? How are the recognition and subsequent degradation regulated? What are specific functions of the Ub-mediated proteolysis of a specific protein substrate? My work, described in this thesis, has produced several direct answers to the above questions in one pathway of the Ub system — the N-end rule pathway in the yeast *S. cerevisiae*.

The Ub/proteasome system in *S. cerevisiae*

Molecular genetic analysis of the physiological functions of the Ub/proteasome system, using *S. cerevisiae* as a model organism, was pioneered by Varshavsky's lab in mid-1980s. Their groundbreaking research laid foundations for subsequent development of the entire Ub field [13]. The yeast *S. cerevisiae* continues to be a favorite model organism, because of the powerful genetic tools that are available with this fungus, and because its relatively small genome has been fully sequenced. I will briefly review our current understanding of the components of the Ub/proteasome system in *S. cerevisiae* (Figure 1.2).

Ub genes

Ub is expressed in *S. cerevisiae* as fusion proteins encoded by four genes: *UBI1*, *UBI2*, *UBI3* and *UBI4* [14, 15]. *UBI1* and *UBI2* encode identical proteins. These Ub fusion proteins are efficiently (cotranslationally) cleaved *in vivo* to yield Ub and "tail" proteins through the action of Ub-specific processing proteases (UBPs). The "tails" of UBI1, UBI2 and UBI3 are ribosomal proteins, which are

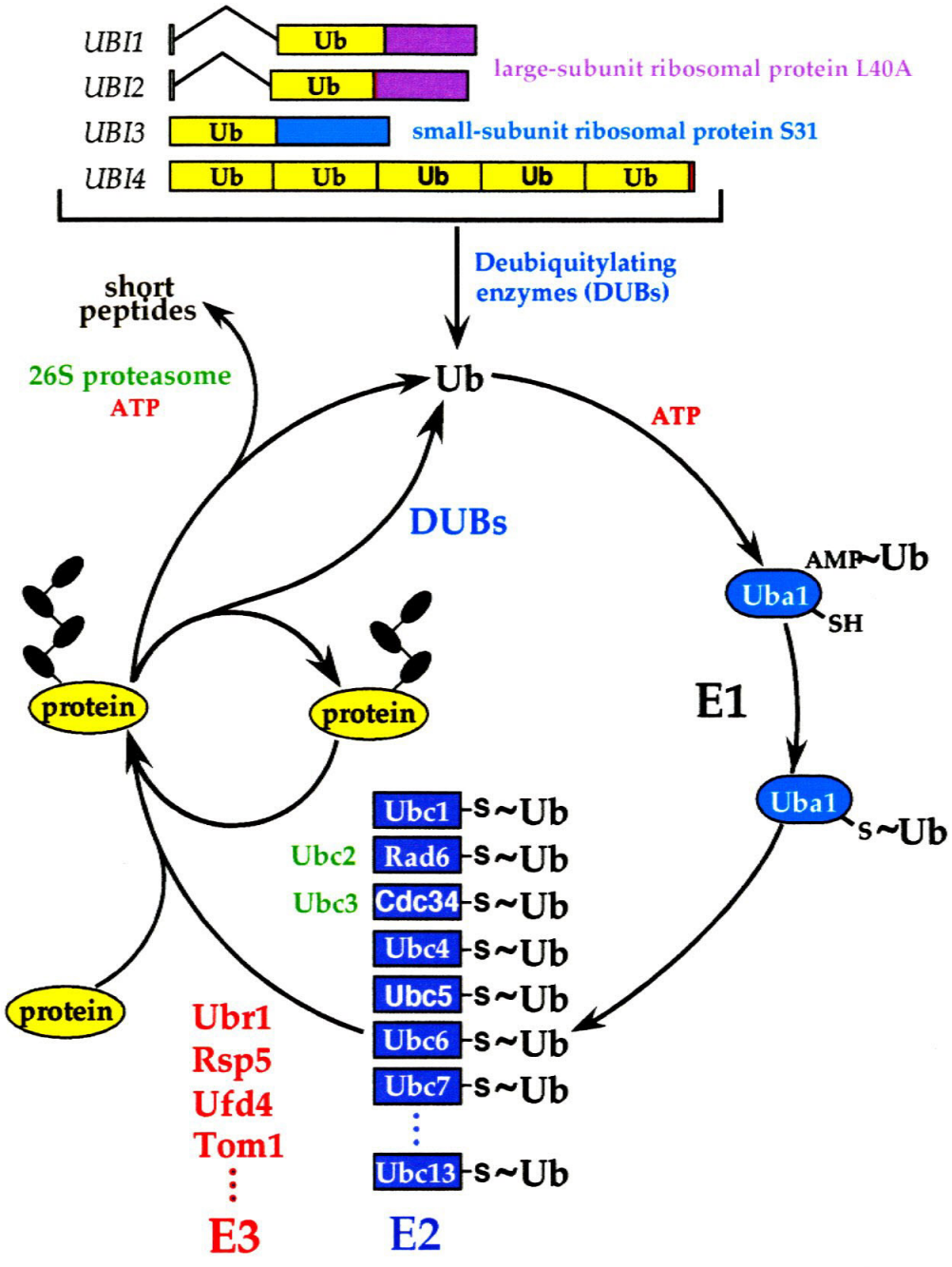


Figure 1.2 The Ubiquitin System in S. *cerevisiae* (see text for discussion)

incorporated into nascent ribosomes more efficiently when they are expressed as Ub fusions than when they are expressed as "free" proteins. In this setting, Ub appears to function as a chaperonin instead of a proteolytic tag [16]. UBI1, UBI2 and UBI3 are expressed primarily in growing cells. In contrast, the poly-Ub UBI4, a head-to-tail fusion of five Ub moieties, is expressed under conditions of stress, including stationary phase. The function of UBI4 is to maintain the pool of free Ub under conditions where this pool decreases as a result of enhanced Ub conjugation to other proteins. *S. cerevisiae* cells lacking *UBI4* are hypersensitive to a broad range of physiological stresses, such as high temperature, starvation and the presence of amino acid analogs [17].

Ubiquitin activating enzyme (E1)

Ubiquitin activating enzyme (UBA), or E1 enzyme, activates Ub by forming a thioester bond between its active-site Cys residue and the C-terminal Gly residue of Ub. The sole E1 enzyme of *S. cerevisiae* is encoded by the essential *UBA1* gene [18]. There are four *UBA1*-related genes in *S. cerevisiae* genome: *AOS1, UBA2, ULA1* and *UBA3*. The sequences of AOS1 and ULA1 are similar to the N-terminal region of UBA1, while UBA2 and UBA3 resemble the C-terminal region of UBA1. Both UBA2 and UBA3 also bear a Cys residue at a position similar to the active-site Cys residue of UBA1 [19] [20]. UBA2 and AOS1 form a complex that activates the Ub-like protein SMT3 (homologous to SUMO-1 in mammals) through the formation of a thioester bond with the C-terminal Gly98 residue of SMT3 [19]. UBA3 and ULA1 form a complex that activates RUB1 (homologous to NEDD8 in mammals) by forming a thioester bond with the C-terminal Gly76 residue of RUB1 [20].

Ubiquitin conjugating enzymes (E2s)

A yeast Ub-conjugating enzyme (UBC), also called E2 enzyme, accepts activated Ub from UBA1 to form a thioester bond with its active-site Cys residue through the transesterification reaction. An E2 enzyme often (if not always) functions as a complex with a specific E3. It is an E2-E3 complex, referred to as a Ub ligase, that determines the substrate specificity of the Ub system [21]. (In the somewhat unrigorous terminology current in the field, the term "Ub ligase" can be used to denote either an E2-E3 complex or E3 alone.) All E2s contain a highly conserved (35-40% identity) catalytic UBC domain where the active-site Cys residue resides.

13 E2-related enzymes (UBC1-UBC13) can be identified in the *S. cerevisiae* genome. Not all of them are Ub-conjugatng enzymes. Specifically, UBC9 accepts activated SMT3 from UBA2 while UBC12 accepts activated RUB1 from UBA3 [20, 22]. The rest — 11 E2s — have either been demonstrated or are presumed to form thioester bond with Ub [23]. Loss-of-function mutations in UBCs result in distinct phenotypes, indicating that E2s have, in general, different functions. For example, RAD6 (UBC2) contains the UBC domain and a highly acidic 23-residue C-terminal region, which contains 20 acidic residues (Asp or Glu) [24]. Yeast cells lacking *RAD6* display pleiotropic defects, including reduced growth rate, enhanced sensitivity to UV and chemical mutagens [25], increased spontaneous mutagenesis, diminished silencing at telomeric loci [26], inability to sporulate, and the failure to degrade N-end rule substrates [27-29]. RAD6 forms separate complexes with RING-finger-containing E3 enzymes RAD18 and UBR1. The RAD6-RAD18 complex is required for postreplicative DNA repair [30] [31], while

the RAD6-UBR1 complex is essential for the activity of the N-end rule pathway [27, 32]. Another E2 enzyme, CDC34 (UBC3) is a subunit of SCF (SKP1, CDC53/Cullin, F box receptor) complexes (see below), which are responsible for the degradation of a variety of cell-cycle regulators, such as SIC1, an S phase cyclin/cyclin-dependent kinase (CDK) inhibitor, G1 cyclin CLN2, and of other regulatory proteins as well, including the GCN4 transcriptional activator [33, 34] [35, 36]. CDC34 is the only Ub-thioester forming E2 that is essential for cell viability—UBC9 is also essential, but it forms thioester with SMT3, not Ub. At least two UBCs are membrane-bound: UBC6 is an integral membrane protein of the endoplasmic reticulum (ER) membrane, with its catalytic domain facing the cytosol. UBC6 is a part of the "ER degradation" pathway for proteins retrotransported from the ER, including integral membrane proteins, such as, for example, misfolded SEC61 [37]. PAS2 (UBC10) is a peripheral membrane protein of yeast peroxisomes. It is required for biogenesis of this organelle [38].

Ubiquitin ligase (E3)

The first discovered E3, named E3α, was defined operationally as a third protein component required, in addition to E1 and E2, for the ligation of Ub to protein substrates in a cell-free system [39]. E3α turned out to be an E3 for the mammalian N-end rule pathway [40]. An E3 generally possesses the following properties: first, it binds to a protein substrate; second, it binds to a specific E2 or E2s; and third, it stimulates substrate ubiquitylation (typically, but not always, multi-ubiquitylation) in the presence of E1 and E2. E3s are the specificity-determining factors of the Ub system [2, 41].

Currently all known E3s fall into two broad classes based on the motifs present in their amino acid sequences: the HECT-domain E3s (homologous to E6-AP carboxy terminus), and the RING-finger E3s [42]. The HECT-domain E3s share a conserved (35-45% identity) ~350 residue region that was originally found near the C-terminus of E6-AP. The conserved Cys residue in the HECT domain accepts activated Ub from its cognate E2 to form a thioester bond which is essential for the activity of HECT-domain E3 [43]. The prototype of HECT-domain E3 is E6-AP. In papilloma virus-infected cells, E6-AP forms a complex with the viral protein E6. The resulting complex, but not E6-AP alone, targets the tumor suppressor p53 for Ub-dependent degradation [44]. There are five genes in the *S. cerevisiae* genome that encode HECT-domain E3s: *RSP5*, *UFD4*, *TOM1*, *HCT4* and *HCT5*. The functions of RSP5, an essential E3, is partially understood. RSP5 catalyzes the multiubiquitylation and degradation of RPB1, the large subunit of RNA polymerase II, in response to DNA damage. The N-terminal WW domains of RSP5 specifically recognize the heptapeptide SPTSPSY present in multiple copies at the C-terminus of RPB1 [45, 46]. RPS5 also recognizes the plasma membrane protein uracil permease encoded by *FUR4*, possibly through its N-terminal C2 lipid-binding domain. However, RPS5-mediated ubiquitylation of FUR4 leads to its endocytosis rather than degradation by 26S proteasome [47].

RING-finger E3s have in common a zinc-binding domain which contains a series of characteristically spaced His and Cys residues [48]. They are thought to function as molecular scaffolds to bring the specific E2s close to their protein substrates, and together with E2, to mediate the synthesis of multi-Ub chains. The RING-finger motif does not seem to have any catalytic functions, based on

the cocrystal structure of a fragment of the RING-finger E3 c-CBL and its cognate E2 UBCH7 [49]. There are dozens of RING-finger proteins in the *S. cerevisiae* genome and hundreds in mammals. At present, it is unclear how many of them may function as E3s.

RING-finger E3s can be either single-subunit or multisubunit [42]. UBR1, the RING-H2 finger E3 that mediates the N-end rule pathway in *S. cerevisiae*, remains one of the best understood single-subunit E3s (see below). The RING-H2 finger is a variant of the RING-finger motif with two specific Cys residues replaced with His residues. There are two major types of multisubunit RING-type E3s in *S. cerevisiae*: the SCF E3s and the cyclosome (another name for the cyclosome is APC, for anaphase-promoting complex). (The choice of the term "APC" to denote the cyclosome E3 was unfortunate, since "APC" was employed earlier to denote a tumor suppressor, termed adenomatous polyposis coli, that is inactivated in most colorectal cancers [50]. It is also an E3, which binds to β-catenin [51] [52] and promotes its degradation [53]. Hence the use of "cyclosome" below.) Both multisubunit E3s contain a small RING-H2-finger protein, HRT1 (RBX1/ROC1) in the SCF-family E3s and APC11 in the cyclosome [36, 54, 55].

An SCF-family E3 is a heterotetrameric complex containing SKP, Cullin, F-box family protein plus HRT1 (RBX1/ROC1). F-box family proteins directly recognize protein substrates by binding to their degrons. 17 genes in the *S. cerevisiae* genome encode F-box proteins, and 2 genes encode CDC53-like proteins, so there are potentially dozens of SCF-family E3s in *S. cerevisiae* [56]. All of the characterized substrates of SCF-family E3s, including CLN1, CLN2 and SIC1, must be phosphorylated to be targeted by the F-box proteins [56, 57].

Interestingly, the activity of the SCF-family E3s can be regulated through reversible modification of the cullin subunit by the Ub-like protein RUB1 (NEDD8 in humans). The conjugation of RUB1 to a specific Lys residue of the cullin subunit in each SCF E3 strongly stimulates substrate multiubiquitylation, through an unknown mechanism [58] [56] [59].

The cyclosome contains at least 12 subunits, including the RING-H2 finger protein APC11 and a cullin homolog APC2. The cyclosome is required for degradation of substrates controlling the metaphase-to-anaphase transition. In addition, the cyclosome-mediated destruction of cyclin B is essential for exit from mitosis [60]. Two classes of WD repeat-containing proteins, CDC20 and CDH1/HCT1, have been identified as substrate-specific cyclosome activators. Both of them bind to cyclosome to form the cyclosome^{Cdc20} and cyclosome^{Cdh1} [60]. It is unclear whether CDC20 and CDH1/HCT1 serve to bring the substrates close to cyclosome, or to provide another activity required for substrate ubiquitylation. A degron recognized by the cyclosome^{Cdc20} includes the 9-residue destruction box (or D box), with the consensus sequence RXXLSSSSN. The D box is present in such substrates as CLB2 and PDS1, whereas the cyclosome^{Cdh1} can target protein substrates containing the 7-residue KEN box, with consensus sequence KENXXXN [61]. The activity of cyclosome is also controlled through cell-cycle regulated phosphorylation of several cyclosome subunits; one of the functions of phosphorylation is to regulate the interaction between cyclosome and Cdc20 [2].

The 26S proteasome in *S. cerevisiae*

26S proteasome is a "self-compartmentalized," ATP-dependent threonine protease. It recognizes a substrate-linked multi-Ub chain, unfolds the substrate, and processively degrades it to short peptides. The multi-Ub chain linked through Lys48 of Ub is disassembled by deubiquityting enzymes (DUBs; they are also called UBPs), and the Ub monomer is recycled [62] [63]. The 26S proteasome in *S. cerevisiae* is composed of a 670 kDa, 20S-core particle (CP), and two 900 kDa, 19S regulatory particles (RP), with one RP capping either end of the cylinder-shaped 20S CP. Purified 26S proteasome can dissociate into CP and RP in the absence of ATP *in vitro* [64].

The archaeon *Thermoplasma acidophilum* has a 20S CP–like structure which was easier to study than the 20S CP from *S. cerevisiae,* because it is consisted of only two types of related α and β subunits. The 14 α and 14 β subunits assemble into four seven-subunit rings, with α subunits forming the outer two heptameric rings, and β subunits forming the inner two rings. The four rings ($\alpha_7\beta_7\beta_7\alpha_7$) assemble into a hollow cylinder with a central channel composed of three chambers formed by the neighboring rings ($\alpha_7\beta_7$, $\beta_7\beta_7$, $\beta_7\alpha_7$). The N-terminal threonine residue of β subunit is the proteolytically active residue, and all the 14 N-terminal threonine residues from the 14 β subunits face the central chamber, comprising the chymotrypsin-like protease activity. This and related evidence indicate that a substrate has to be translocated into the chamber to be degraded. X-ray structure of this 20S bacterial proteasome reveals that at each end of the 20S proteasome from *Thermoplasma acidophilum*, there is a 13 Å opening which is only big enough to allow the passage of an unfolded polypeptide chain [65, 66].

The sequestration of the protease active sites in an internal chamber implies that substrate specificity of the 20S proteasome is controlled by the access of substrate to the interior active sites.

The overall structure of the yeast 20S CP is similar to that of *Thermoplasma acidophilum*, but the yeast 20S CP is composed of 2 sets of 14 related but distinct subunits that can be grouped into seven α-type and seven β-type subunits. Three β-subunits encode three Thr-proteases which are responsible for three protease activities: chymotrypsin-like (cleaves after bulky hydrophobic residues), trypsin-like (cleaves after basic residues) and caspase-like (cleaves after acidic residues) [41]. The X-ray structure of the 20S proteasome from *S. cerevisiae* revealed that the N-terminal tails of several α subunits seal the opening to the central channel so that there is no obvious axial pore [67]. This finding was consistent with the fact that the purified 20S proteasome from *S. cerevisiae* exhibits at most weak peptidase activities, unless it is first "activated" by treatment such as incubation with SDS [68]. Crystal structure of the complex formed between the 20S proteasome from *S. cerevisiae* and a heptameric 11S regulator from trypanosomes has been solved, and it clearly shows that the major function of REG regulator is to keep the gate to the central channel open [69]. Thus, one of the functions of RP is to reverse the closed state of the axial pore, mediated by the N-terminal tails of several α subunits, so that unfolded substrates can be translocated into the central channel of CP [67, 70].

The 900 kDa RP is composed of 17 subunits. It can be further dissociated into two subcomplexes that are aptly named the "base" and the "lid" [64, 71]. The base is consisted of 9 subunits, including all six ATPases (RPT1-RPT6) and

RPN1, RPN2, RPN10. The base directly contacts CP, and activates it for degradation of peptides or unfolded substrates, but not of multiubiquitylated substrate. The ATPases are responsible for unfolding and translocation of protein substrates into CP [72]. The ATPase domain of RPT2 mediates the opening of the central channel of 20S proteasome [73]. The functions of the lid remain largely unknown. Both the lid and base are required to recognize and unfold Ub-protein conjugates both *in vivo* and *in vitro* [71].

It remains unclear how multiubiquitylated substrates are recognized and delivered to the 26S proteasome. A multi-Ub chain is believed to be the principal targeting signal, and the tetra-Ub repeat comprises the minimal proteasomal targeting signal [74] [75]. RPN10, a subunit in the base of RP, was shown to bind to synthetic multi-Ub chains *in vitro* [76]. However, *S. cerevisiae* cells lacking *RPN10* are viable, grow at near-wild-type rates, and degrade the bulk of short-lived proteins normally. The *rpn10Δ* cells exhibit relatively mild degradation defects with the N-end rule substrates, but more severe defect with Ub-Pro--βgal, a substrate of the UFD (Ub Fusion Degradation) pathway [77]. Thus, RPN10 is not the only ligand, and possibly not the major ligand of a multi-Ub chain in 26S proteasome. Recently, the ATPase RPT5 at the base of RP was found to interact with multi-Ub chains in the presence of ATP but not ADP, suggesting that, in addition to substrate unfolding and translocation, the binding of multi-Ub chain to the 19S complex may also require ATP hydrolysis (C. M. Pickart, personal communication). Another distinct route is that E3s themselves may help deliver multiubiquitylated substrates to 26S proteasome. Several E3s were found to interact with subunits in the base of the 19S RP. Examples include UBR1, the

single-subunit RING-H2-finger E3 of the N-end rule pathway; UFD4, the HECT-domain E3 of the UFD (Ub fusion degradation) pathway [78], and KIAA10, a HECT-domain mammalian E3 [79]. Specific subunits of SCF complex and cyclosome are co-purified with 19S RP in the absence of ATP or in the presence of ATP-γ-S, a poorly hydrolyzable ATP analog [80]. However, the physiological significance of the interactions between E3s and 26S proteasome remains to be understood.

The N-end rule pathway

The N-end rule was discovered through the invention of the Ub fusion technique [81], in which the DNA sequence encoding the 76-residue Ub was ligated in-frame to the DNA sequence encoding a reporter protein β-galactosidase (βgal). The construct was anticipated to produce the linear Ub fusion protein Ub-X-βgal (X denoting the first residue of βgal) in *S. cerevisiae*. However, the N-terminal Ub was cotranslationally cleaved off after its C-terminal residue Gly76 by DUBs (UBPs), irrespective of the identity of residue X, proline being the only exception [81]. At least one mammalian UBP enzyme was discovered later to be able to cleave the Ub-Pro peptide bond [82]. The Ub fusion technique overcame the limitation imposed by the specificity of Met-aminopeptidases, and made it possible to generate any of the 20 residues at the N-terminus of a protein of interest.

Remarkably, depending on the identity of the N-terminal residue X, the reporter protein X-βgal showed vastly different metabolic stability in *S. cerevisiae*: Arg-βgal was quickly degraded, in a Ub-dependent manner, with a half-life of

~2 min ($t_{1/2}$ ~ 2 min) while Met-βgal was stable ($t_{1/2}$ > 20 hr). The residues that confer short half-lives to reporter proteins are called destabilizing residues, while those that do not are called stabilizing residues. The relation thus revealed, between the *in vivo* half-life of a protein and the identity of its N-terminal residue, was termed the N-end rule [81] (Figure 1.3).

Distinct versions of the N-end rule pathways are present in all organisms examined (Figure 1.3), from single-cell eukaryotes such as *S. cerevisiae* [81] to the amphibian *Xenopus laevis* [83], mammals (rabbit reticulocyte lysate and mouse cell lines) [84] [85] [86] and plants (*Arabidopsis thaliana*) [87]. Moreover, a version of the N-end rule pathway is present in prokaryotes, which lack Ub and Ub-specific enzymes [88]. In bacteria, a reporter protein bearing a destabilizing N-terminal residue is degraded by the ATP-dependent protease ClpAP, which shares structural and functional features with the eukaryotic 26S proteasome. It is likely (but remains to be demonstrated) that ClpAP directly recognizes a destabilizing N-terminal residue of a substrate.

The N-end rule is organized hierarchically [84] [89]. In eukaryotes, the destabilizing N-terminal residues are consisted of three groups: primary, secondary and tertiary destabilizing residues. Primary destabilizing N-terminal residues are directly recognized by E3s of the N-end rule pathway. In *S. cerevisiae* , they include basic or "type 1" residues (Arg, Lys, His) and bulky hydrophobic or "type 2'" residues (Leu, Trp, Phe, Tyr, Ile). Secondary destabilizing N-terminal residues include the acidic residues Asp and Glu, which have to be conjugated at their N-termini to Arg, a type 1 destabilizing residue, by the enzyme arginyl-tRNA-protein transferase (R-transferase). Tertiary

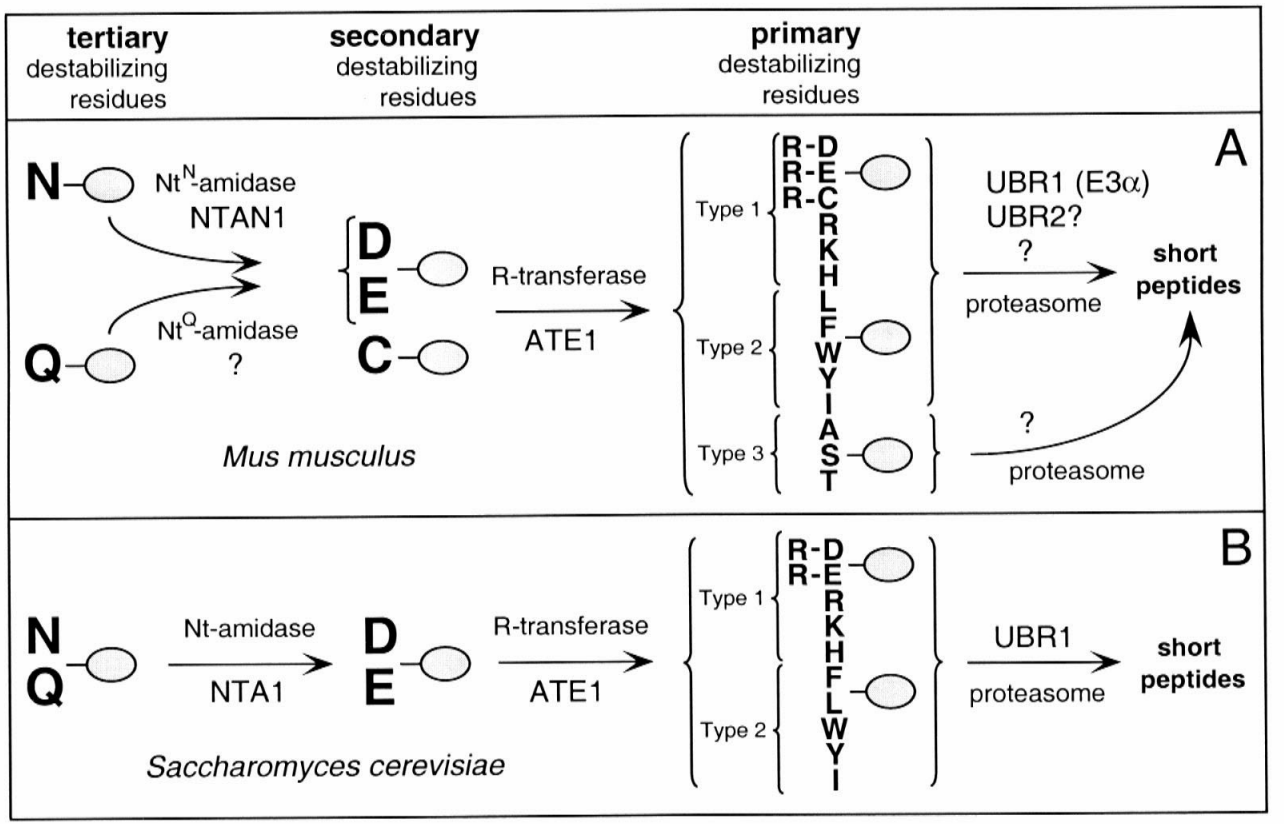


Figure 1.3 The N-end rule pathways in mouse and yeast (see text for discussion)

destabilizing N-terminal residues include Asn and Gln, which have to be deamidated by N-terminal amidase, before conjugated with Arg by R-transferase [86, 90]. In mammals, the hierarchical organization of the N-end rule is essentially the same as in yeast, except that Cys is a secondary destabilizing residue (it is stabilizing in yeast). It is possible that the sulfhydryl side chain of the N-terminal Cys residue has to be oxidized to the negatively charged cysteine sulfinic acid (Cys(O2)) or cysteic acid (Cys(O3)) to become a substrate for arginylation by R-transferase (see below).

N-degron and its applications

Soon after the discovery of the N-end rule in *S. cerevisiae* , it was found [91] that a destabilizing N-terminal residue is necessary but not sufficient to target a reporter protein for Ub-dependent degradation. A mouse DHFR (dihydrofolate reductase) derivative with a destabilizing N-terminal residue remained stable. An additional 45-residue N-terminal extension present in the original reporter protein βgal was required to destabilize DHFR in *S. cerevisiae.* This extension (named e^K) was originated from the *E. coli* LacI repressor and contains two lysine residues at positions 15 and 17. Through extensive permutations of e^K, it was found that at least one of these two lysines in e^K, either Lys15 or Lys17, was essential for X-e^K-DHFR degradation [91]. In parallel, Lys15, 17 were identified as alternative attachment sites for multi-Ub chain [7]. Thus, the degradation signal that is recognized by the N-end rule pathway, termed the N-degron, consists of at least two determinants: a destabilizing N-terminal residue recognized directly by UBR1 (E3 of the N-end rule pathway)

and a spatially "appropriate" lysine residue that acts as multi-Ub attachment site [91]. More recently, the strength of the original N-degron has been markedly increased without loss of its specificity, through the addition of lysine residues or lysine-containing random sequences to the e^K extension [92].

The bipartite organization of N-degron recurs in many other degrons present in short-lived proteins that are targeted for degradation by the Ub/proteasome system. For example, IκB is the inhibitor of the transcription factor NFκB that is central in immune and inflammatory response. Phosphorylation of IκB at residues Ser32 and Ser36, in response to a variety of external stimuli, induces its degradation by the Ub/proteasome system. The degron of IκB is composed of a short phosphopeptide DS*GLDS* (S* denotes phosphoserine), as well as Lys21 and Lys22. The latter are the major (alternative) sites for Ub conjugation. The phosphopeptide DS*GLDS* is directly recognized by an F-box protein β-TrCP in SCF $^{\beta\text{-TrCP}}$, a multisubunit E3 belonging to the SCF family [57].

Trans-targeting and subunit-specific degradation

In a monomeric N-end rule substrate, the two components of N-degron reside in the same polypeptide chain, and thus are recognized in *cis*. However, in an oligomeric substrate these two determinants may reside in different subunits, and were shown to be recognized in *trans* [93]. This was demonstrated in reticulocyte lysate with purified heterotetrameric X-βgal test proteins that contained X-βgal chains of different designs. The recognition, by the N-end rule pathway, of a βgal subunit bearing a destabilizing N-terminal residue but

lacking the Lys15 and Lys17, can lead to *trans*-targeting of a βgal subunit in the same tetramer that bears a stabilizing N-terminal residue but contains Lys15, 17 for ubiquitylation. It was shown that, under these conditions, the ubiquitylated X-βgal subunit was destroyed, whereas the initial subunit, bearing a destabilizing N-terminal residue, was left intact [93]. The subunit-specific degradation by the Ub/proteasome system is made possible, in part, by the properties of 26S proteasome: first, it recognizes the multi-Ub chain as the principal targeting signal; second, it unfolds the multiubiquitylated subunit, so that it dissociates from the other subunits of the same complex before it is translocated into the internal chamber of 26S proteasome. Surprisingly, *trans*-targeting and subunit-specific degradation have also been discovered in bacteria, which do not have the Ub system. The umuD′ subunit in the UmuD-umuD′ dimer is *trans*-targeted by the N-terminal 24 residues of UmuD (UmuD′ is the same as umuD except it lacks the N-terminal 24 residues), and degraded by the ATP-dependent protease ClpXP [94]. The molecular mechanism of this *trans*-targeting remains unclear.

The subunit-specific degradation enables the Ub/proteasome system to destroy specifically only its target subunit of a protein complex, while leaving intact the rest of the complex's subunits. For example, some E3s appear to bind tightly to their multiubiquitylated substrates; subunit-specific degradation by 26S proteasome destroys only the multiubiquitylated protein substrate, releasing the associated E3s [95]. *Trans*-recognition and subunit-specific degradation may be exploited to construct a new class of dominant negative mutants, in which a protein could be specifically destabilized by *trans*-targeting it for the

Ub/proteasome-dependent degradation through another protein or even small molecules [96] [97].

N-degron and conditional mutants

The N-degron has also been used to construct a portable, heat-inducible degron. Arg-DHFR^{ts} is long-lived at 23°C, but is short-lived at 37°C [98]. Fusing this portable, 21 kDa conditional degron in-frame to the N-terminus of a protein of interest yields a new class of *ts* mutants, called *td* (temperature-activated degron). The *td* method bypasses the often laborious and unsuccessful search for *ts* mutants; it only requires that a protein of interest can tolerate N-terminal extensions. The *td* method also eliminates or reduces the undesirable phenotypic lag that is common with traditional *ts* mutants, since a *td* protein is rapidly destroyed at nonpermissive temperature. A recent improvement of the original *td* method [99] reduced the phenotypic lag further by speeding up the degradation of a *td* protein at nonpermissive temperature. This was accomplished by subjecting UBR1 to the transcriptional control by the *GAL1-10* promoter so that in the presence of galactose, the intracellular UBR1 concentration is elevated. At permissive temperature, however, the *td* protein remains stable. The improved *td* method has produced temperature-sensitive (*ts*) mutants for more than 80 percent of the genes that have been tried [99]. This modification of the *td* technique underscored several earlier findings. First, UBR1 is the rate-limiting component of the N-end rule pathway [100]. Second, the high specificity of UBR1 toward N-end rule substrates is retained even when it is overexpressed [100, 101]. And third, a strong increase in the intracellular

concentration of UBR1 is not toxic, unless the cognate E2 enzyme RAD6 is overexpressed as well [100] [101].

In principle, a conditional degron can be constructed so that it is controlled by small membrane-permeable compounds. Levy and colleagues [102] attempted to make Arg-DHFR^{ts} responsive to methotrexate (MTX), a low-M_r ligand of DHFR, both in *S.cerevisiae* and in mammalian cells. This induction system needs to be refined before it can be used similarly to the *td* technique.

Other uses of N-degron

The N-degron was the first portable degron discovered [81] [91], and it has been the platform for many other discoveries. For example, the N-degron has been used to probe the mechanism of protein degradation by 26S proteasome. It was shown that a DHFR-based N-end rule substrate (X-e^K-DHFR) was stabilized in ATP-supplemented reticulocyte lysate after binding to MTX (which stabilizes the folding of DHFR), although the ubiquitylation of X-e^K-DHFR was not inhibited. Thus the unfolding of DHFR must precede, and is required for, its degradation by 26S proteasome [103]. Indeed, it was demonstrated recently that protein degradation by ATP-dependent proteases, including 26S proteasome, is preceded by the local unfolding of protein substrate starting from the degradation signal. Thus, a multidomain protein with independently folded domains is unfolded sequentially prior to its degradation in ATP-dependent proteases [104].

It has been a long-standing question whether a substantial fraction of nascent polypeptide is cotranslationally degraded. Recently, the N-degron was employed to detect and measure cotranslational degradation *in vivo* [105]. Through a novel Ub sandwich technique, it was shown that more than 50% of nascent proteins bearing an N-degron could be degraded cotranslationally (and processively) by the Ub/proteasome system in *S. cerevisiae*.

Components of the N-end rule pathway

In *S. cerevisiae*, the targeting part of the N-end rule pathway (before 26S proteasome) comprises the proteins encoded by *UBR1, RAD6, ATE1* and *NTA1* (Figure 1.3). UBA1 (E1, the Ub-activating enzyme) is a part of every pathway that requires the activated Ub, including the N-end rule pathway.

N-recognin

By definition, N-recognin is an E3 enzyme that recognizes at least some of its substrates through their destabilizing N-terminal residues. In *S. cerevisiae*, the sole N-recognin (UBR1) is a 1,950-residue protein [100]. Several lines of evidence indicate that UBR1 is the sole N-recognin in *S. cerevisiae*. First, the *S. cerevisiae* cells lacking *UBR1* could not degrade any N-degron-bearing protein substrates. Second, UBR1 was found to be specifically crosslinked to Leu-e^K-DHFR, an N-end rule substrate with a type 2 destabilizing residue; but not to Met-e^K-DHFR, which bears a stabilizing residue. And third, UBR1-ha (a UBR1 variant with a C-terminal ha epitope tag), was found to coimmunoprecipitate with Arg -e^K-βgal, an N-end rule substrate bearing a type 1 destabilizing residue; but not

with Val-e^K-βgal, which bears a stabilizing residue [100]. These and other results demonstrated that UBR1 is the sole N-recognin in *S. cerevisiae*, and that it specifically interacts with the type 1 or type 2 destabilizing N-terminal residues.

The binding sites of UBR1 for the type 1 or type 2 destabilizing residues are termed the type 1 site or type 2 site, respectively. Either of the two sites is capable of binding either a dipeptide or a polypeptide chain, but not both at the same time, because dipeptides with type 1 or type 2 destabilizing residues act as competitive inhibitors of the degradation of N-end rule substrates carrying the same type of destabilizing residues [84, 106] [107]. There is an allosteric interaction between the conformations of the type 1 and type 2 sites of UBR1, because type 1 dipeptides (dipeptides with basic N-terminal residues) accelerate the degradation of type 2 N-end rule substrates (the N-end rule substrates bearing bulky hydrophobic N-terminal residues). In contrast, type 2 dipeptides (dipeptides with bulky hydrophobic N-terminal residues) do not appear to affect the degradation of type 1 N-end rule substrates (the N-end rule substrates bearing basic N-terminal residues) [84] [107]. All these observations can be explained by the recently proposed mechanism for structural transitions in UBR1 (Figure 5 of Chapter 4) (see below).

Both type 1 and type 2 sites of UBR1 have been mapped to its N-terminal ~600-residue region through genetic screens (A. Webster, M. Ghislain and A. Varshavsky, personal communication). In these screens, randomly mutagenized *UBR1* gene was introduced into the *ubr1Δ* cells, which expressed a pair of N-end rule reporter proteins: Arg-e^K-Ura3 & Leu-e^K-βgal, or Leu-e^K-Ura3 & Arg-e^K-βgal. UBR1 variants that targeted either type 1 or type 2, but not both

types of the N-end rule substrates were recovered, and the missense mutations in the *UBR1* ORF were identified through DNA sequencing. The five type 1 UBR1 mutants, each with one or two substituted residues: G173R, G173N, D176E, D176N, and C145YV146M, are only able to target type 2, but not type 1, N-end rule substrates for Ub-dependent degradation. These mutations have been assumed to disrupt the integrity of the type 1 site, but not of the type 2 site (A. Webster, M. Ghislain and A. Varshavsky, unpublished data). All of these mutated residues are located in a highly conserved region of UBR1 (see below). Another group of five type 2 UBR1 mutants, each with one substituted residue: V313I, D318N, H321Y, P406S, and E560K, are only able to target type 1, but not type 2, N-end rule substrates for Ub-dependent degradation. These five mutations have been assumed to disrupt the integrity of the type 2 site without affecting the type 1 site, of UBR1. These mutated residues are more dispersed (A. Webster, M. Ghislain and A. Varshavsky, unpublished data).

More features of the *S. cerevisiae* UBR1 have been revealed through sequence alignments with its homologs from other organisms, including the mouse E3α [40], and the UBR1 homologs from the nematode *C. elegans* and the yeast *Kluyveromyces lactis*. Three regions of the *S. cerevisiae* UBR1 are highly conserved among the UBR family members (Figure 1.4) [40]: the N-terminal UBHC (UBR/His/Cys) region (residues 123 –193), in which the His and Cys residues are among the most conserved; the essential RING-H2 finger (residues 1219 -1333) [108]; and the C-terminus-proximal, 108-residue UBLC (UBR/Leu/Cys) region (residues 1,702 - 1,809), in which the Leu and Cys residues are the most conserved. All of the residues that are essential for the integrity of the type 1 site of UBR1 (Cys145, Val146, Gly173, and Asp176), and

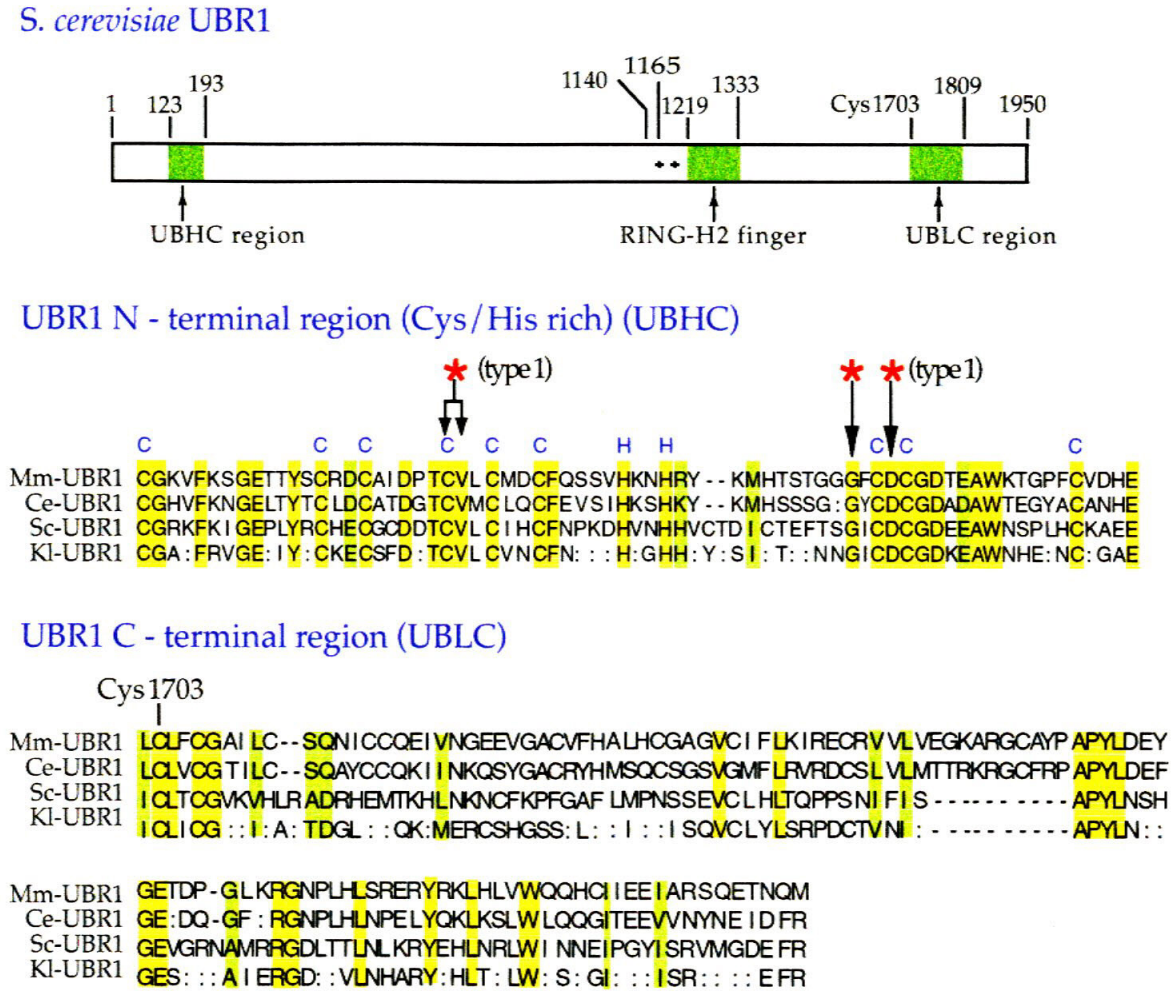


Figure 1.4 The conserved regions of UBR1 homologs (see text for discussion)

some of the residues that are essential for its type 2 site (Asp318, His312, and Glu560), are conserved among the UBR1 homologs from the yeast *S. cerevisiae*, the mouse *mus musculus* and the nematode *C. elegans* [40]. The BRR (basic/residues/rich) region (residues 1,165 -1,175), which mediates the high-affinity interaction between UBR1 and the polyacidic tail of the yeast RAD6 [108], is only present in *S. cerevisiae* UBR1. This is consistent with the fact that, among RAD6 homologs, only the *S. cerevisiae* RAD6 possesses the polyacidic tail. However, it was demonstrated that the high-affinity interaction between the *S. cerevisiae* UBR1 and RAD6, mediated largely through the BRR region of UBR1 and the polyacidic tail of RAD6, is not essential for the activity of the N-end rule pathway in *S. cerevisiae* [108].

UBR1 also possesses a third substrate binding site, termed site "i," for the internal degron of the 35 kDa homeodomain protein CUP9, a transcriptional repressor of the di- and tripeptide transporter PTR2 [109] (Chapter 2). This has been established through the *in vitro* ubiquitylation assays. CUP9 was found to be efficiently conjugated with multi-Ub chains in an ATP-supplemented *in vitro* system, which consisted of Ub and the purified components of the yeast N-end rule pathway, including $UBA1^{H}$ (a UBA1 variant with a C-terminal hexahistidine tag), RAD6 and ^{F}UBR1 (a UBR1 derivative with an N-terminal FLAG epitope tag). Remarkably, addition of the type 1 or type 2 dipeptides, not the dipeptides of the same composition but with a stabilizing N-terminal residue, nor the amino acid components of these dipeptides, substantially stimulated the UBR1-dependent ubiquitylation of CUP9 [109]. These results led to two important conclusions. First, UBR1 directly interacts with CUP9 through a third substrate

binding site that is distinct from its type 1 and type 2 sites. Second, both the type 1 and type 2 sites of UBR1 are not only the substrate-binding sites, but also allosteric sites, and the allosteric effectors for these two sites are type 1 or type 2 dipeptides. In *S. cerevisiae*, the UBR1-mediated allostery underlies the positive feedback circuit governing the di- or tripeptide import [109].

What are the molecular mechanisms underlying the stimulation of UBR1 activity by type 1 or type 2 dipeptides? The answers are only beginning to emerge [110]. As illustrated in the autoinhibition model of UBR1 (Figure 5 of Chapter 4), the *S. cerevisiae* UBR1 is an autoinhibitory protein, its binding to the transcriptional repressor CUP9 is autoinhibited by its C-terminal region, which folds back and interacts, in part through the UBLC domain, with a region that encompasses (or is adjacent to) the site "i" (the internal degron of CUP9). In this "closed" conformation, a steric hindrance by the C-terminal region of UBR1 precludes the binding of UBR1 to CUP9. A pair of type 1 and type 2 dipeptides (red rectangle and green bullet, respectively) synergistically relieves the autoinhibition, by causing an opening of the closed UBR1 conformation, thus abolishing the occlusion of the CUP9-binding site. In the absence of pairs of type 1 and type 2 dipeptides, the probability of the closed $\rightarrow$ open transition is low, as a result, the open, CUP9-binding conformation of UBR1 is a minor species in the largely closed-conformation UBR1 ensemble. The binding of a pair of type 1 and type 2 dipeptides to the type 1 and type 2 sites of UBR1 shifts the equilibrium toward an ensemble where the open-conformation UBR1 becomes the predominant species. The model also posits that the binding of type 1 dipeptides to the (always accessible) type 1 site of UBR1 induces a conformational change that increases the accessibility (and/or affinity) of the type 2 binding site to type

2 dipeptides, and it is the latter binding that induces the closed → open transition in UBR1 and the unmasking of CUP9-binding site. This model also explains two earlier findings: Type 1 dipeptides accelerate the degradation of type 2 N-end rule substrates, whereas type 2 dipeptides do not accelerate the degradation of type 1 N-end rule substrates. How type 1 dipeptides induce the conformational alteration of UBR1 remains to be investigated.

It is possible that some other covalent or non-covalent modifications of the *S. cerevisiae* UBR1, including a protein interaction, phosphorylation or modification with Ub-like proteins, may shift the conformation of the UBR1 ensemble toward one way or the other. For example, amino acid-induced dipeptide import has been attributed to the enhanced degradation of CUP9 through the N-end rule pathway (G.C. Turner, C. C. Byrd and A. Varshavsky, personal communication). It is possible that the amino acid-derived signal(s), transduced through the SSY1 and PTR3-dependent pathway, functions to shift the conformation of the UBR1 ensemble toward its "open" conformation.

The *S. cerevisiae* UBR1 is the first Ub ligase shown to be regulated through autoinhibition. Mammalian cells express at least three distinct homologs of the *S. cerevisiae* UBR1 that contain the conserved UBHC, RING-H2, and UBLC domains [40] (Y. T. Kwon and A. Varshavsky, personal communication). These metazoan E3s are also likely to be regulated through autoinhibition, but their allosteric effectors may not be confined to dipeptides.

Arginyl-tRNA-protein Transferase

Aminoacyl-tRNA-protein transferases catalyze posttranslational incorporation of specific amino acids to the N-termini of acceptor proteins [111].

Its physiological functions remained unknown until the discovery of the N-end rule [81] [112] [84]. In prokaryotes, Leu, Phe-tRNA-protein transferase (L/F-transferase) catalyzes the conjugation of mainly Leu, a primary destabilizing N-terminal residue, to the N-terminal Arg or Lys of an acceptor protein. *E.coli* mutants lacking the gene encoding L/F-transferase are unable to degrade the N-end rule substrates with the N-terminal Arg or Lys residue, but appear normal otherwise [113] [88]. No physiological substrates or functions have been discovered for L/F-transferase, as yet.

Arginyl-tRNA-protein transferase (R-transferase) catalyzes the conjugation of Arg to the N-terminal Asp, Glu or Cys residue in metazoans [84] [114] (Y. T. Kwon and A. Varshavsky, personal communication), but only to Asp or Glu residue in *S. cerevisiae* [115]. The yeast mutants lacking *ATE1*, the only R-transferase gene in *S. cerevisiae*, appear normal except for their inability to degrade the N-end rule substrates bearing tertiary and secondary destabilizing residues [115]. ATE1, expressed from the fully induced *GAL1-10* promoter on a high-copy plasmid, is toxic to *S. cerevisiae* cells. It results in growth arrest of haploid cells, but only slows down the growth of diploid cells. Notably, the toxicity is enhanced in the yeast cells lacking *UBR1*, suggesting that the toxicity is unrelated to the proteolytic activity of the N-end rule pathway (F. Du, H. R. Wang and A. Varshavsky, unpublished data). Growth arrest of haploid cells is relieved by expressing additional copies of any of the six $tRNA^{Arg}$ from *S. cerevisiae*, suggesting that the toxicity caused by overexpression of ATE1 may be due to the sequestration of endogenous $tRNA^{Arg}$ by ATE1, so that there is not

enough $tRNA^{Arg}$ left for protein synthesis (F. Du, H. R. Wang and A. Varshavsky, unpublished data).

The mammalian R-transferase has been partially purified from rabbit reticulocyte lysate, and shown to co-fractionate with arginyl-tRNA synthetase [116]. The cDNA encoding the mouse R-transferase has been isolated; its deduced amino acid sequence is similar to the *S. cerevisiae* ATE1 [117]. Two murine cDNA species, generated through alternative splicing, were found to be present in different ratios in different mouse tissues, and the corresponding two proteins, ATE1-1 and ATE1-2, differ in their intracellular localization [117]. When expressed in yeast cells, both ATE1-1 and ATE1-2 are able to arginylate the N-terminal Asp or Glu residues of an N-end rule substrate, albeit with different efficiency. But neither of these two enzymes, expressed alone or co-expressed in yeast cells, is able to arginylate the N-terminal Cys residue of an otherwise identical test protein [117]. These results appeared to suggest that there exists, in addition to the ATE1-1 and ATE1-2, at least one other distinct R-transferase that is specific for N-terminal Cys residue, and is present in mammals but not in yeast [117]. Recently, however, the mouse cells lacking both ATE1-1 and ATE1-2 were found to be deficient in degrading the N-end rule substrates bearing any of the three secondary destabilizing N-terminal residues: Asp, Glu or Cys, indicating that ATE1-1 and/or ATE1-2 are required for the arginylation of N-terminal Cys residue of an acceptor protein (Y. T. Kwon and A. Varshavsky, personal communication).

Independently of the above advances, I came up with a new explanation regarding the secondary destabilizing residue Cys in mammals, based on the

structural similarity between Asp residue and the oxidized forms of Cys: sulfinic acid (Cys(O2)) residue or cysteic acid (Cys(O3)) residue. Specifically, I suggested that the N-terminal Cys residue has to be oxidized, most likely by a specific enzyme(s), to either Cys(O2) or Cys(O3) residue (denoted as Cys(O2)/Cys(O3) hereafter) to become a substrate of R-transferase and thus a secondary destabilizing residue in mammals. According to this conjecture, R-transferases for the N-terminal Asp and Glu residues (yeast ATE1 and its mammalian homologs ATE1-1 and ATE1-2) can arginylate the N-terminal Cys(O2)/Cys(O3) residue as well, because Cys(O2)/Cys(O3) (particularly Cys(O2)) is structurally similar to Asp. According to this interpretation, the reason for arginylation of N-terminal Cys in mammals but not in yeast may be not the absence of a Cys-specific R-transferase from yeast, but rather the absence of yeast oxidase(s) for N-terminal Cys residue. These new ideas led to several predictions, some of which have already been confirmed in this laboratory.

For example, the above hypothesis predicts that in mammals (but not in yeast) the N-terminal Cys is oxidized to the Cys(O2)/Cys(O3), before conjugation to Arg. However, sulfinic acid residue (Cys(O2)) tends to be unstable, and becomes oxidized to the cysteic acid (Cys(O3)) residue. Thus, one prediction of the hypothesis was that a protein whose N-terminal Cys was arginylated *in vivo* or *in vitro*, would be expected to have either sulfinic acid or cysteic acid as the penultimate residue. This prediction was confirmed with mouse RGS4 as the N-end rule substrate in mouse L cells (I. V. Davydov and A. Varshavsky, personal communication).

RGS4 was identified as an arginylation substrate through an expression-cloning screen in reticulocyte lysate [114]. The deduced second residue Cys of

RGS4 is expected to be exposed at its N-terminus, because the initiator Met residue is cleaved off by Met-aminopeptidases [118]. The N-terminus of RGS4 was found to be conjugated to Arg residue before being targeted through the N-end rule pathway [114]. The RGS4 ^{FH} protein (an RGS4 variant with a FLAG epitope and a hexahistidine tag at its C-terminus) was overexpressed and purified from mouse L cells. It was then digested with CNBr, and the peptides generated were subsequently analyzed by tandem mass spectrometry (I. V. Davydov and A. Varshavsky, personal communication). About half of the recovered N-terminal fragments of the RGS4 protein were discovered to possess an N-terminal Arg residue, followed by an unidentified penultimate residue, whose M_r was 47.3 (±0.1) Da, more than that of a Cys residue. The rest of N-terminal fragments started with the same unidentified residue from their N-termini (I. V. Davydov and A. Varshavsky, personal communication). Most likely, this unidentified residue is the predicted cysteic acid (Cys(O3)), whose side chain, in contrast to the sulfhydryl of Cys, is ionized, which explains that the gained M_r is ~47 Da, rather than ~48 Da derived from the three extra oxygen atoms present in Cys(O3). More recently, re-evaluation of these mass spectrometry data confirmed that the above species was indeed cysteic acid (I. Davydov, personal communication).

Yet another line of recent evidence supports the conjecture that the modified form of Cys residue is indeed Cys(O3) (I. V. Davydov and A. Varshavsky, personal communication). It was noticed, but unexplained initially, that when the purified RGS4 protein was digested first with the AspN protease, instead of CNBr, all the recovered N-terminal fragments started (after alkylation

to protect the unmodified Cys residue, prior to sequencing) with the same unidentified derivative Cys residue, and none of them possessed an N-terminal Arg residue. These results suggest that the N-terminal Arg residue was cleaved off by the AspN protease, which is consistent with the prediction that the above unidentified residue is Cys(O3), because AspN protease is known to be able to cleave the peptide bond N-terminal to the Cys(O3) residue [119]. The observed Cys(O3) residue did not seem to result from the sample manipulations *in vitro*, since the other Cys residue (Cys12) in the same peptide was not oxidized.

This result, and also the failure of ATE1-lacking mouse cells to arginylate a Cys-bearing test protein *in vivo* (see above; Y. T. Kwon and A. Varshavsky, personal communication), strongly suggest that mouse ATE1-1 and/or ATE1-2 can arginylate N-terminal Cys(O2) and/or Cys(O3) residue (this remains to be investigated), in addition to the N-terminal Asp and Glu residues. Several classes of enzymes have been demonstrated to exhibit relaxed specificity to Asp and Glu versus Cys(O2)/Cys(O3) residues. The AspN protease described above is one example. Yet another example is the large family of aspartate aminotransferases from various organisms, which accept free Cys(O2) amino acid, in addition to Asp, as their substrate [120].

The Cys-oxidation hypothesis suggests that *S. cerevisiae* ATE1-encoded R-transferase, if it is expressed in the mouse ATE1(-/-) cells, might also be able to conjugate Arg to the N-terminal Cys(O2)/Cys(O3) residues, in addition to its already known substrates, the N-terminal Asp and Glu residues. Very recently, this prediction has been confirmed with ATE1(-/-) fibroblasts and *S. cerevisiae* ATE1 (Y. T. Kwon and A. Varshavsky, personal communication). In summary, there are, by now, several independent lines of evidence, which, if taken

together, prove that N-terminal Cys must be oxidized prior to its arginylation by R-transferase. A number of important questions, made possible by this advance, remain to be answered. For example, we don't know, as yet, the identity of Cys-specific (possibly N-terminal Cys-specific) oxidase involved, except that it should be present in mammalian cells but absent from S. cerevisiae (at least under normal conditions). (It is very likely, but remains to be proven, that this oxidation is, in fact, catalyzed by a specific enzyme or enzymes.) A putative candidate is the known enzyme, called cysteine dioxygenase (CDO) in mammals, which oxidizes free Cys amino acid to Cys(O2) [121]. It remains to be tested whether CDO can also oxidize N-terminal Cys residue in proteins or short peptides.

As had already been found with mouse RGS4, some of the mammalian proteins bearing a penultimate Cys residue might be *bona fide* N-end rule substrates under some conditions, after the (cotranslational) removal of their N-terminal Met by Met-aminopeptidases. These enzymes are able to cleave the initiator Met off if the second residue is Cys [118]. Identification of Cys-based arginylation substrates is complicated by the fact that N-terminal Cys residue may undergo alternative modifications that include acetylation and palmitoylation [122], in addition to oxidation as proposed here. It is possible that the Cys-specific enzymes involved may compete to modify a given substrate, and that this competition may be physiologically relevant.

Less than 0.7 percent proteins in the Swissprot database have Cys as the penultimate residue. The list of mouse proteins include the ubiquitin ligase MDM2, and the transcriptional factors JUN-B and LEX1. It will be interesting to test whether some of these proteins are stabilized in the mouse ATE1(-/-) cells.

The discovery that arginylation of N-terminal Cys is preceded by its (most likely, enzymatic) oxidation expands the set of covalent modifications that some N-terminal residues must undergo before becoming substrates of R-transferase. The previously known modification of this class was the deamidation of N-terminal Asn or Gln residue, discussed below.

N-terminal Amidase

The *S. cerevisiae* N-terminal amidase (NTA1) converts the N-terminal Asn and Gln residues to Asp and Glu, respectively. *S. cerevisiae* lacking *NTA1* cannot degrade N-end rule substrates that bear the tertiary destabilizing N-terminal Asn and Gln residues, but retain the rest of the N-end rule pathway, and appear normal otherwise [90]. The *S. cerevisiae* cells lacking *NTA1* grow as fast as the wild-type cells in all types of yeast medium tested, including the medium with the non-fermentable lactate, pyruvate or glycerol as the sole carbon source at 37°C, 30°C and 23°C (F. Du and A. Varshavsky, unpublished data).

Two in-frame ATG start codons in the *NTA1* ORF, which are 30 bp apart from each other, encode two initiator Met residues for two NTA1 species. The larger one, translated from the ATG(+1), is 457 residues long; the smaller one lacks the N-terminal 10 residues of the larger species. Both of the predicted isoforms of NTA1 are probably present in *S. cerevisiae* cells [90]. The 29-residue region, between the Asp4 and Asp34 in the larger NTA1 species, resembles the mitochondrial translocation signal. Indeed, an NTA1-GFP fusion protein, with the ATG(+31)-encoded Met residue replaced with Ala residue, is localized predominantly in the mitochondria when expressed in *S. cerevisiae*, but it is also present in the cytosol (H. R. Wang and A. Varshavsky, personal communication).

In contrast, when the smaller NTA1 species was fused to GFP, and the fusion expressed in *S. cerevisiae*, it was localized mainly in the cytosol (H. R. Wang and A. Varshavsky, personal communication). No physiological substrates and functions of yeast NTA1 in either the mitochondria or the cytosol have been discovered, thus far.

Through extensive amino acid sequence alignments and similarity analyses, the N-terminal amidases from the yeast *S. cerevisiae* and *S. pombe* were found to constitute one family of the nitrilase superfamily [123]. Every enzyme in the nitrilase superfamily appears to possess a novel Glu-Lys-Cys catalytic triad, and functions through a thiol acylenzyme intermediate. The residues of the Glu-Lys-Cys catalytic triad for the *S. cerevisiae* NTA1 are Glu63, Lys136 and Cys187. Only Cys187 has been experimentally demonstrated to be essential for the amidase activity of *S. cerevisiae* NTA1 [90].

Early (unpublished) observations (M. Ghislain and A. Varshavsky, personal communication) suggested that *S. cerevisiae* NTA1 and ATE1 physically interact. However, later findings contradicted that preliminary conclusion (F. Du and A. Varshavsky, unpublished data). Specifically, in the initial experiments, anti-FLAG beads were found to co-immunoprecipitate the $ATE1^F$ (an ATE1 variant with a C-terminal FLAG epitope tag) and the untagged NTA1, but not $NTA1^F$ (an NTA1 variant with a C-terminal FLAG epitope tag) and the untagged ATE1, when each pair of the two proteins were co-overexpressed in yeast cells (M. Ghislain and A. Varshavsky, unpublished data). Later it was demonstrated that anti-FLAG beads immunoprecipitated similar amount of NTA1 when it was co-overexpressed with either the $ATE1^F$, or (unexpectedly) the untagged ATE1

— a critical negative control for the specificity of co-immunoprecipitation assay, which was missing in the earlier assays. In addition, close inspection of the NTA1 sequence revealed at least one FLAG-like sequence (the sequence EYKDMDEPD starting from the 286th residue conforms to the consensus sequence of the FLAG epitope DYKXXD, where X can be any residue) [124] (F. Du and A. Varshavsky, unpublished data). These results suggested that the initial observation of an apparent interaction between the NTA1 and ATE1 may have resulted from non-specific interactions between anti-FLAG beads and the untagged NTA1. This interpretation still does not explain why anti-FLAG beads did not immunoprecipitate the untagged NTA1 when it was overexpressed alone in yeast cells. Several other approaches also failed to detect any physical interactions between the *S. cerevisiae* NTA1 and ATE1. For example, no interactions between the purified, enzymatically active NTA1 and ATE1^{H} (an ATE1 variant with a C-terminal hexahistidine tag) could be detected, using such methods as gel filtration and affinity chromatography (F. Du and A. Varshavsky, unpublished data). Yeast two-hybrid assays, with a variety of combinations of NTA1 and ATE1, also failed to detect any interactions (F. Du and A. Varshavsky, unpublished data). Taken together, these results strongly suggest that the yeast NTA1 and ATE1 do not form a high-affinity complex, at least under conditions examined.

Interestingly, the deamidation branch of the N-end rule pathway, which consists of one enzyme, NTA1, in yeast, consists of at least two distinct N-terminal amidases (Nt-amidases) in mammals. One enzyme, Nt^{N}-amidase, is specific for N-terminal Asn, and has been characterized genetically and

biochemically. The other, Nt^{Q}-amidase, is specific for N-terminal Gln, and remains to be characterized. cDNAs encoding the porcine and murine N-terminal amidases that are specific for the N-terminal Asn, but not Gln, have been cloned [125] [126]. The Asn-specific enzyme, termed NTAN1, shares no sequence similarity to the *S. cerevisiae* NTA1 or any other members of the nitrilase superfamily. The *NTAN1(-/-)* mouse strains have been constructed. These mice are viable, outwardly normal, and fertile [86]. As expected, *NTAN1(-/-)* mouse cells are unable to deamidate the N-terminal Asn residue, but remain competent in deamidating the N-terminal Gln residue. As a result, test N-end rule substrates bearing the N-terminal Asn, but not Gln, is stabilized in the mouse *NTAN1 (-/-)* cells [86]. These results indicated that there is at least one distinct N-terminal amidase specific for N-terminal Gln. This enzyme(s) remains to be characterized.

Physiological substrates and functions of the N-end rule pathway

The N-end rule was discovered with artificial (engineered) substrates, produced through the Ub-fusion technique [81]. The genes encoding the components of the N-end rule pathway in *S. cerevisiae* were isolated through genetic screens that utilized these engineered test proteins [100] [90, 115]. It is only recently that the physiological substrates and functions of the N-end rule pathway in *S. cerevisiae* have begun to emerge.

The difficulty of identifying physiological substrates and functions of the N-end rule pathway stemmed mainly from the following two obstacles.

1) Ablation of the N-end rule pathway in *S. cerevisiae*, through a deletion of the *UBR1* gene, resulted in only mild phenotypic changes, such as the slightly

(~3%) slower growth rate on standard media, and the occurrence of higher proportion of two-spored asci. These mild defects could not be exploited directly to set up the genetic screens for the physiological substrates of the N-end rule pathway. To augment the phenotypic differences between the wild-type and the *ubr1Δ* yeast cells, a synthetic lethal screen, which aimed to find the yeast mutants whose viability were dependent on *UBR1*, was carried out, and yielded one yeast mutant, termed *sln1* (synthetic lethal of N-end rule) [127]. The cloned *SLN1* gene was found to encode a eukaryotic homolog of the bacterial two-component regulators [127]. However, the underlying molecular mechanism of the synthetic lethality remains unclear. In another attempt to search for the phenotypes that involve the N-end rule pathway, K. Madura discovered that expression of both UBR1 and RAD6 from fully induced *GAL1-10* promoter inhibited the growth of haploid yeast cells, but not diploid yeast cells. This ploidy-dependent toxicity was traced to the enhanced degradation of the GPA1-encoded Gα, a subunit of the yeast heterotrimeric G protein. Gα was demonstrated to be degraded through the N-end rule pathway [101]. Physiological function of this degradation remains to be understood. The first clear-cut physiological defect of yeast cells lacking UBR1 was discovered by the laboratory of J. Becker (Univ. of Tennessee, Memphis, TN). It was found that *ubr1Δ* cells are unable to import di- or tripeptides from the medium [128]. This defect led to the discovery of CUP9, the first physiological substrate of the N-end rule pathway in *S. cerevisiae* [129] (see below).

2) There is no obvious way to predict *a priori* whether a specific ORF encodes an N-end rule substrate of *S. cerevisiae*, partly because it is still not

feasible, in most cases, to determine whether a protein is endoproteolytically cleaved by a protease that would expose a destabilizing N-terminal residue. Moreover, because of the bipartite organization of the N-degron, a protein bearing a destabilizing N-terminal residue may not be a *bona fide* N-end rule substrate.

The known Met-aminopeptidases in *S. cerevisiae* are unlikely to generate destabilizing N-terminal residues, because they remove the initiator Met residue only if the second residue is one of the seven residues with the smallest radii of gyration [118], all of which are stabilizing residues in *S. cerevisiae*. However, endoproteolytic cleavage by some proteases was predicted to be able to generate destabilizing N-terminal residues [81]. Indeed, the first N-end rule substrate discovered, the Sindbis RNA polymerase nsP4, which bears a type 2 destabilizing N-terminal residue Tyr, is generated through the endoproteolytic cleavage of its viral precursor polyprotein P1234 [130]. Another N-end rule substrate, produced in a similar way, is HIV-1 integrase, whose N-terminal residue is Phe, a type 2 destabilizing N-terminal residue [131]. Recently, SCC1, a subunit of the *S. cerevisiae* cohesin complex, was found to be cleaved by the *ESP1*-encoded cysteine protease separase during the metaphase-anaphase transition; its 33 kDa C-terminal fragment bears an N-terminal Arg residue [132], a type 1 destabilizing N-terminal residue. This C-terminal fragment of SCC1 is targeted through the N-end rule pathway [133] (see below). Moreover, the elimination of the SCC1 fragment by the N-end rule pathway was shown to be essential for chromosome stability [133].

Protein translocation, from the extracellular space or various compartments such as ER or Golgi into the cytosol or nucleus, was also predicted

to be a route that yields physiological N-end rule substrates [81]. Many of the extracellular or compartmentalized proteins have destabilizing N-terminal residues, exposed through endoproteolytic cleavages. Some of these proteins, if retrotranslocated or leaked into the cytosol or nucleus, may become N-end rule substrates [81]. The two known N-end rule substrates generated through this route are p60 [134] and ActA [135]; both are secreted by the intracellular bacterium *Listeria monocytogenes* into the cytosol of the infected mammalian cells.

The N-end rule pathway and peptide import in *S. cerevisiae*

In a genetic screen aimed at identifying the genes that are important in the amino acid-inducible peptide transport in *S. cerevisiae*, Becker and colleagues discovered that the yeast mutants deficient in dipeptides import in the presence of amino acid inducers fell into three PTR (peptide transport) complementation groups: *PTR1*, *PTR2* and *PTR3* [136]. The *PTR2*-encoded protein is predicted to encompass 12 hydrophobic membrane spanning segments, and it shares significant amino acid sequence similarities with the CHL1 nitrate transporter of *Arabidopsis thaliana*, so PTR2 is most likely a di- and tripeptide permease [137]. PTR3 is a peripheral plasma membrane protein, which forms a complex with two other proteins, SSY1 and SSY5. This complex mediates an amino acid-sensing signal transduction pathway of unknown mechanism [138] [139] [140]. *PTR1* turned out to be *UBR1*, and yeast cells lacking *PTR1/UBR1* had very low level of *PTR2* mRNA [128]. In addition to *UBR1*, *RAD6* was found later to be required for peptide import [129]. These results firmly established the requirement of the N-end rule pathway for peptide import in *S. cerevisiae*.

Since the N-end rule pathway was known to be involved in Ub-dependent protein degradation, it was postulated that there exists a short-lived negative regulator of *PTR2*, which is targeted by this pathway [129]. In the absence of *UBR1*, this negative regulator accumulates to high levels, thus shutting off the transcription of *PTR2*. As a result, yeast cells lacking *UBR1* are unable to import dipeptides. Mutational inactivation of this negative regulator were predicted to enable yeast cells to bypass the requirement for *UBR1* in peptide import. A genetic screen for such "bypass" yeast mutants, which could import dipeptides in the absence of *UBR1*, was carried out. All of the mutants obtained fell into three complementation groups, and the genes that were inactivated turned out to be *CUP9*, *TUP1* and *SSN6* [129] (G. C. Turner, C. C. Byrd and A. Varshavsky, unpublished data). CUP9 is exactly the short-lived negative regulator of *PTR2* that the genetic screen was designed to identify (see below).

CUP9 was originally discovered as a protein that is involved, in a way that remains to be understood, in copper resistance when yeast cells grow on lactate, a non-fermentable carbon source [141]. CUP9 is a 35 kDa, homeodomain protein. It specifically binds to the *PTR2* promoter. In addition, it interacts with the SSN6-TUP1 repressor complex. Thus, CUP9 functions as a specific transcriptional repressor of *PTR2*, most likely through recruiting the SSN6-TUP1 repressor to the *PTR2* promoter [129] (G. C. Turner, C. C. Byrd and A. Varshavsky, unpublished data).

CUP9 was found to be short-lived, with half life of ~5 min ($t_{1/2} \approx 5$ min). Its degradation required the N-end rule pathway: the half-life was increased from ~5 min to at least 50 min in *S. cerevisiae* lacking either *RAD6* or *UBR1* [129].

In contrast to the engineered N-end rule substrates, which are targeted through the N-degron, CUP9 was found to be targeted through an internal degron. This has been firmly established by three lines of evidence. First, the $CUP9^{F}$ protein (a CUP9 variant with a FLAG epitope tag at its C-terminus), purified from the yeast cells lacking *UBR1*, was found to have a blocked (presumably acetylated) N-terminus, since it could not be sequenced through Edman degradation. Second, the GST-CUP9 fusion protein, in which CUP9 is fused in-frame to the C-terminus of GST, is also degraded through the N-end rule pathway, with similar half-life to CUP9. The degradation of the GST-CUP9 fusion is not accompanied by the appearance of a fragment containing the N-terminal GST moiety, strongly suggesting that endoproteolytic cleavage of CUP9 is not a prerequisite to its processive degradation by the 26S proteasome. Third, $^{F}CUP9^{H}$ (a CUP9 derivative with an N-terminal FLAG epitope tag and a C-terminal hexahistidine tag) expressed in, and purified from, *E. coli* is efficiently multiubiquitylated in an *in vitro* system, which is composed of the purified components of the yeast N-end rule pathway, including $UBA1^{H}$ (a UBA1 variant with a C-terminal hexahistidine tag), RAD6, ^{F}UBR1 (a UBR1 derivative with an N-terminal FLAG epitope tag) and Ub in the ATP-supplemented buffer (Chapter 2). These results establish that CUP9 is directly recognized by UBR1, rather than, for example, *trans*-targeted to UBR1 through an interacting protein that bears a destabilizing N-terminal residue.

The internal degron of CUP9 has been mapped to its C-terminal region through a genetic screen. In this screen, the *CUP9* ORF in the *URA3-CUP9* fusion was randomly mutagenized, and the stabilized URA3-CUP9 mutants in the wild-

type yeast cells were selected. It was found that all of the single-residue substitutions, which stabilized both the URA3-CUP9 fusion and CUP9, are located within the last 33 residues of the 306-residue CUP9 (F. Navarro-Garcia, G. C. Turner and A. Varshavsky, personal communication). These single-residue substitutions were demonstrated to weaken the affinity between UBR1 and CUP9, and the extent of affinity decrease correlates with the degree of the metabolic stabilization of the corresponding CUP9 mutants in *S. cerevisiae* [110]. These results suggest that these single-residue substitutions directly perturb the region through which CUP9 binds to UBR1.

The 16-residue fragment between the residues 286 to 301 of CUP9 is predicted to form an amphipathic helix (F. Du and A. Varshavsky, unpublished data). Most of the mutations that stabilized CUP9 *in vivo* are mapped to the hydrophobic surface of the amphipathic helix. For example, Leu294 is central to the amphipathic helix, and its substitution with Pro, a known helix breaker, is predicted to destroy the internal degron of CUP9. Indeed, CUP9(L294P), a CUP9 variant with its Leu294 mutated to Pro, was found to be the most stabilized Cup9 mutant *in vivo* (F. Navarro-Garcia, G. C. Turner and A. Varshavsky, personal communication). The amphipathic helix model of the internal degron of CUP9 remains to be verified directly. How the exposure of the internal degron of CUP9 might be regulated remains to to be understood (see below).

What is the physiological rationale of the targeting of CUP9 by UBR1, the E3 of the N-end rule pathway? What are the physiological functions of the N-end rule pathway in peptide import in *S. cerevisiae*? The answers to these questions did not emerge until the seminal pulse-chase experiments by Turner [109]. He discovered that either type 1 or type 2 dipeptides, but not dipeptides of

the same composition but with stabilizing N-terminal residues, accelerate CUP9 degradation through the N-end rule pathway. This accelerated CUP degradation leads to the derepression of PTR2 transcription, and consequently to an increase in the yeast cell's capability to import di- and tripeptides. In summary, dipeptides with destabilizing N-terminal residues stimulate their own uptake, as well as uptake of other di- and tripeptides. The molecular mechanisms underlying this UBR1-mediated positive feedback circuit, which regulates the dipeptide import in *S. cerevisiae*, are discussed above (see the section on N-recognin) [109] [110].

The positive feedback regulation governing peptide import in *S. cerevisiae* is vividly demonstrated through colony growth assays. Yeast strain auxotrophic for leucine grows robustly on SHM plates supplemented with the dipeptide Leu-Ala (230 μM), which bears a destabilizing N-terminal residue Leu, and thus stimulates its own uptake. In contrast, the same yeast strain does not form visible colonies in the presence of the dipeptide Ala-Leu (230 μM), which is of the same amino acid composition as Leu-Ala, but bears a stabilizing N-terminal residue Ala [109].

Food sources that *S. cerevisiae* encounters in the wild are likely to contain mixtures of type 1 and type 2 peptides. The cooperativity between the type 1 and type 2 sites of yeast UBR1 (discussed in the section on N-recognin) may enable *S. cerevisiae* to sense even lower concentration of mixtures of peptides than when only one type peptide is present. Indeed, it was demonstrated that yeast cells start importing dipeptides at lower dipeptide concentration in the medium when both type 1 and type 2 dipeptides are present, in comparison to a setting where only one type of dipeptide is present [110].

The N-end rule pathway and chromosome stability

In eukaryotes, replicated DNA molecules remain attached to each other until the onset of anaphase. Cohesion of sister chromatids involves a protein complex called cohesin. In *S. cerevisiae*, the cohesin is composed of four subunits: SMC1, SMC3, SCC1 and SCC3 [142]. Cleavage of SCC1 during metaphase-anaphase transition is essential for the initiation of sister chromatids separation [132]. The 566-residue SCC1 is cleaved by a cysteine protease [143], termed ESP1, which is activated by the cyclosome-dependent degradation of its inhibitor PDS1 [144]. SCC1 is cleaved by ESP1 at two sites: the peptide bonds C-terminal to the residues 180 and 268, with the latter cleavage site being the major one [132]. The C-terminal fragment thus generated starts with the residue Arg269, a type 1 destabilizing residue in the N-end rule This fragment, Arg-SCC1$^{269-566}$, is short-lived ($t_{1/2} \approx 2$ min) in the wild-type yeast cells but long-lived ($t_{1/2} > 2$ h) in the yeast cells lacking *UBR1*. In contrast, Met-SCC1$^{269-566}$, which is identical to Arg-SCC1$^{269-566}$ except for the N-terminal residue, is long-lived in the yeast cells with or without *UBR1*. So Arg-SCC1$^{269-566}$ is the first physiological N-end rule substrate that is targeted through the N-degron in *S. cerevisiae* [133]. The accumulation of this C-terminal fragment of SCC1 is toxic to yeast cells: high expression level of Arg-SCC1$^{269-566}$ is lethal to yeast cells lacking *UBR1*, but results in slow growth of wild-type S. cerevisiae (which can degrade Arg-SCC1$^{269-566}$). However, high expression level of Met-SCC1$^{269-566}$ (which is stable in both wild-type and *ubr1Δ* cells) is lethal to both the wild-type or the *ubr1Δ* yeast cells [133].

The chromosome-loss defect of *ubr1Δ* cells was traced to their inability to degrade Arg-SCC1$^{269-566}$, as shown through the use of a colony sectoring assay [145] [133]. *ubr1Δ* yeast cells, which do not have the N-end rule pathway, lose an artificial chromosome at least 100 times more frequently than their congenic *UBR1* cells. Moreover, the bulk of the increased chromosome loss is caused by the metabolic stabilization of Arg-SCC1$^{269-566}$ in *ubr1Δ* cells [133]. This dramatic defect escaped earlier scrutiny largely because yeast cells lacking *UBR1* did not differ significantly from their wild-type counterparts in the rate and kinetics of cell-cycle progression.

The human SCC1 homolog is also cleaved by separase at two sites [146], exposing either Glu (a secondary destabilizing residue) or Leu (a type 2 primary destabilizing residue) as the N-terminal residues. It will be interesting to determine whether one or both of these fragments are physiological substrates of the mammalian N-end rule pathway.

ESP1 is a cysteine protease, similarly to caspases involved in apoptosis. Several cleavage products of the caspase substrates have been found to expose destabilizing N-terminal residues, and the number of such substrates continues to increase. It will not be surprising if some of the caspase-produced cleavage products are *bona fide* physiological substrates of the N-end rule pathway.

Future directions

Over the past several years, we have made significant progress in understanding both physiological functions and molecular mechanisms of the N-end rule pathway in *S. cerevisiae*. These advances made possible new

approaches to understanding the mammalian N-end rule pathway. In the meantime, the cloning of cDNAs encoding components of the mammalian N-end rule pathway, including mouse NTAN1, ATE1, UBR1, UBR2, and UBR3, led to the construction of mouse strains lacking each of these components. This systematic dissection of the mammalian N-end rule pathway in under way in the Varshavsky laboratory (Y. T. Kwon and A. Varshavsky, personal communication).. It is hoped that analogous analyses of the N-end rule pathway in other genetically tractable multicellular eukaryotes, including the nematode *C. elegance*, the fruit fly *Drosophila melanogaster*, and the plant *Arabidopsis thaliana* will begin soon enough. Mentioned below are some of the important questions about the N-end rule pathway that remain to be addressed.

1) Do mammalian UBR1 homologs have similar autoinhibitory architectures and regulatory mechanisms as the yeast UBR1? The answers to this question will help understand the molecular mechanisms underlying the regulation of UBR-family Ub ligases. All UBR1 homologs share at least three conserved regions, including the N-terminal UBHC region, the essential RING-H2 finger and the C-terminal UBLC region; it is likely that they possess similar autoinhibitory architectures, although their allosteric effectors may not necessarily be dipeptides. An informative experiment is to test whether the N-terminal fragments of these UBR1 homologs interact with their own C-terminal fragments, and whether these interactions can also be disrupted by type 1 and type 2 dipeptides, analogous to what has been done with the yeast UBR1 [110].

2) What are physiological substrates of the mammalian N-end rule pathway? How to find them? How is their degradation regulated? What are

physiological functions of the degradation of a given N-end rule substrate? Identifying physiological substrates is the key to our understanding of the N-end rule pathway in any organism. Knowing the substrates will help to understand the mechanistic and functional underpinnings of a variety of defects observed in mice lacking specific components of the N-end rule pathway (Y. T. Kwon and A. Varshavsky, personal communication). To find N-end rule substrates that are targeted through their internal degrons, one of the approaches is to search for the proteins that are physically associated with the mouse UBR1 homologs, or rather, with their N-terminal fragments if mammalian UBR are found to be autoinhibitory. One way to identify such interacting proteins: N-terminal fragment of a UBR protein that contains substrate-binding sites is fused to two independent affinity tags, and is expressed from a strong promoter in mammalian cells derived from the mouse strain lacking the corresponding UBR homolog. The double-tagged protein is then purified through two affinity steps specific for the two affinity tags, and the co-purified proteins are identified through mass spectroscopy. To inhibit protein binding to the possible type 1 or type 2 sites of these UBR1 homologs, type 1 and type 2 dipeptides may be included in the buffers throughout the purification processes.

3) How does the N-terminal Cys residue become a secondary destabilizing residue in mammals? Is the actually arginylated derivative of N-terminal Cys a sulfinic acid ((Cys(O2)) or cysteic acid (Cys(O3))? What is the identity of oxidase(s) that converts the N-terminal Cys residue to the substrate of R-transferase? And what are physiological substrates (other than RGS4) that are targeted through this branch of the N-end rule pathway? Finally, how did the cysteine branch of the mammalian N-end rule pathway evolve? (It appears to be

absent from fungi.) For more details, please refer to the section on Arg-tRNA-protein transferase.

4) What is physiological rationale of the cooperativity between the type 1 and type 2 sites of yeast UBR1? What are the underlying molecular mechanisms? Living cells need to integrate multiple input signals to generate coordinated responses. Coordination at the molecular level is often achieved through the synergism between multiple recognition domains of signaling proteins, with each of these domains responsive to one specific signal [147]. Cooperativity between the multiple domains of a signaling protein is often employed to amplify the coincident input signals [148] [147]. Yet another powerful way to amplify the input signals is through the positive feedback regulation, which underlies many "all-or-none" responses observed in a variety of signal transduction pathways [149] [150].

Both strategies, the positive feedback regulation, and the cooperativity between different recognition domains of a signaling protein, are employed to regulate the import of peptides by *S. cerevisiae* [109] [110]. While we have understood relatively well the molecular mechanisms underlying the positive feedback regulation of peptide import, we know little about physiological functions and molecular mechanisms that underlie the cooperativity between type 1 and type 2 sites of the yeast UBR1 Ub ligase. One possibility is that the yeast UBR1-based positive feedback circuit, coupled with the cooperativity between the type 1 and type 2 sites of UBR1 protein, enable yeast cells to switch from one state — PTR2 transcription shut off, to another state — PTR2 transcription is maximally induced, over a narrow range of dipeptide concentrations. This model is not necessarily incompatible with the observed

gradual increase of the level of PTR2 mRNA in response to the addition of dipeptide(s), [109], because the "switch-like" response of the PTR2 transcription may occur in individual yeast cells, instead of a population of cells as a whole. In other words, PTR2 transcription in a single yeast cell might be largely an "all-or-none" event: it is either completely off, or is maximally induced. The observed gradual increase of PTR2 mRNA in a population of yeast cells is caused largely by the gradual increase of the number of the yeast cells whose PTR2 transcription is maximally induced, instead of the gradual increase of the PTR2 transcription in each individual yeast cells.

A way to verify the above conjecture is to fuse the *GFP* ORF to the 3′ of the chromosomal *PTR2 ORF*, and then to examine the intensity of GFP fluorescence of individual yeast cells, in the presence of multiple combinations, and concentrations, of different dipeptides. A properly designed PTR2-GFP fusion is likely to remain competent in importing dipeptides, since PTR2 can tolerate epitope tags at its C-terminus. If the number of yeast cells with maximally expressed PTR2 (high fluorescence levels) is plotted against the concentration of specific dipeptide pairs, it might be possible to determine the Hill coefficient (degree of cooperativity of the response) for each type 1/type 2 dipeptide pair. By comparing the Hill coefficient for these dipeptide pairs, for example, Arg-Ala & Leu-Ala, Arg-Ala & Ala-Leu, and Ala-Arg & Leu-Ala, it might be possible to determine the relative contributions from the positive feedback regulation circuit (Figure 4 of Chapter 3) and cooperativity between the type 1 and type 2 sites of UBR1. A quantitative analysis of this kind has recently been carried out with the mitogen-activated protein kinase (MAPK) cascade of *Xenopus* oocytes [150].

5) What constitutes the internal degron of CUP9? How is it recognized by yeast UBR1? How might the exposure of this internal degron be regulated? While it is certain that the internal degron of CUP9 encompasses the C-terminal region of CUP9, including the predicted amphipathic-helix-forming region between the residues 286 to 301, a detailed mapping of this degron remains to be done.

The internal degron of CUP9 is currently thought to be constitutively exposed, since the ectopically expressed (overexpressed) CUP9 is constantly degraded in the wild-type yeast cells. It is possible that degradation of CUP9 expressed at wild-type levels is actually conditional rather than constitutive. For example, the degron of CUP9 might be shielded in its complex with the SSN6-TUP1 repressor. One precedent for a scenario of this kind: another homeodomain protein, MATα2, a cell-type specific transcriptional repressor, is constitutively degraded through the Ub system in haploid yeast cells, but is dramatically stabilized in diploid cells. This stabilization of MATα2 is due to the masking of its amphipathic-helix-forming degron by MATa1, another cell-type specific transcriptional repressor that interacts with MATα2 [151].

The above conjecture predicts that overexpression of the SSN6-TUP1 repressor, but not their missense mutants with diminished affinity for CUP9, may stabilize CUP9 *in vivo*. A version of this model predicts that long-lived CUP9 mutants may bind to the SSN6-TUP1 repressor less tightly than wild-type CUP9. Stabilization of CUP9 by the SSN6-TUP1 repressor may also be physiologically relevant, because it ensures that CUP9 is stabilized when it binds to its cognate promoter, and functions as a specific transcriptional repressor in a complex that

also includes the SSN6-TUP1 repressor. CUP9 may become short-lived only when it dissociates from the SSN6-TUP1 repressor. In other words, the earlier observation that CUP9 was constantly degraded in wild-type cells might be an artifact of CUP9 overexpression. An alternative model, which is also not precluded by the current data, is that CUP9 remains short-lived even when it is bound to DNA.

Finally, the established *in vitro* ubiquitylation system, which consists of the purified components of the yeast N-end rule pathway, is both specific and simple in composition. Future utilization of this system should illuminate the mechanism of substrate recognition and multiubiquitylation in the N-end rule pathway. Since the discovery of the N-end rule pathway in 1986, its studies repeatedly uncovered features of this pathway that are directly relevant to the understanding of the entire Ub system.

References

1. Varshavsky, A., *The ubiquitin system.* Trends Biochem. Sci., 1997. **22**: p. 383-387.
2. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu. Rev. Biochem., 1998. **67**: p. 425-479.
3. Hershko, A., A. Ciechanover, and I.A. Rose, *Resolution of the ATP-dependent proteolytic system from reticulocytes: a component that interacts with ATP.* Proc. Natl. Acad. Sci. USA, 1979. **76**(7): p. 3107-10.
4. Hershko, A., et al., *Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis.* Proc. Natl. Acad. Sci. USA, 1980. **77**(4): p. 1783-6.
5. Hershko, A. and A. Ciechanover, *Mechanisms of intracellular protein breakdown.* Annu. Rev. Biochem., 1982. **51**: p. 335-64.
6. Hershko, A., *Lessons from the discovery of the ubiquitin system.* Trends Biochem. Sci., 1996. **21**(11): p. 445-9.
7. Chau, V., et al., *A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein.* Science, 1989. **243**: p. 1576-1583.
8. Koegl, M., et al., *A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly.* Cell, 1999. **96**(5): p. 635-644.
9. Coux, O., K. Tanaka, and A.L. Goldberg, *Structure and functions of the 20S and 26S proteasomes.* Annu. Rev. Biochem., 1996. **65**: p. 801-817.
10. Rechsteiner, M., *The 26S proteasome.*, in *Ubiquitin and the Biology of the Cell*, J.M. Peters, J.R. Harris, and D. Finley, Editors. 1998, Plenum Press: New York. p. 147-189.

11. Finley, D., A. Ciechanover, and A. Varshavsky, *Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85.* Cell, 1984. **37**(1): p. 43-55.

12. Ciechanover, A., D. Finley, and A. Varshavsky, *Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85.* Cell, 1984. **37**(1): p. 57-66.

13. Hershko, A., A. Ciechanover, and A. Varshavsky, *The ubiquitin system.* Nature Med., 2000. **10**: p. 1073-1081.

14. Ozkaynak, E., D. Finley, and A. Varshavsky, *The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein.* Nature, 1984. **312**(5995): p. 663-666.

15. Ozkaynak, E., et al., *The yeast ubiquitin genes: a family of natural gene fusions.* EMBO J., 1987. **6**(5): p. 1429-39.

16. Finley, D., B. Bartel, and A. Varshavsky, *The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis.* Nature, 1989. **338**: p. 394-401.

17. Finley, D., E. Özkaynak, and A. Varshavsky, *The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses.* Cell, 1987. **48**(6): p. 1035-46.

18. McGrath, J.P., S. Jentsch, and A. Varshavsky, *UBA1: an essential yeast gene encoding ubiquitin-activating enzyme.* EMBO J., 1991. **10**: p. 227-236.

19. Johnson, E.S., et al., *The ubiquitin-like protein Smt3p is activated for conjugation to other proteins by an Aos1p/Uba2p heterodimer.* EMBO J, 1997. **16**(18): p. 5509-19.

20. Liakopoulos, D., et al., *A novel protein modification pathway related to the ubiquitin system.* EMBO J., 1998. **17**(8): p. 2208-14.

21. Jentsch, S., *The ubiquitin-conjugating system.* Annu. Rev. Genet., 1992. **26**: p. 179-207.

22. Johnson, E.S. and G. Blobel, *Ubc9p is the conjugating enzyme for the ubiquitin-like protein Smt3p.* J. Biol. Chem., 1997. **272**(43): p. 26799-802.

23. Scheffner, M., S. Smith, and S. Jentsch, *The ubiquitin conjugation system.*, in *Ubiquitin and the Biology of the Cell*, J.-M. Peters, J.R. Harris, and D. Finley, Editors. 1998, Plenum Press: New York. p. 65-98.

24. Jentsch, S., J.P. McGrath, and A. Varshavsky, *The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme.* Nature, 1987. **329**: p. 131-134.

25. Tuite, M.F. and B.S. Cox, *RAD6+ gene of Saccharomyces cerevisiae codes for two mutationally separable deoxyribonucleic acid repair functions.* Mol. Cell. Biol., 1981. **1**(2): p. 153-7.

26. Huang, H., et al., *The ubiquitin-conjugating enzyme Rad6 (Ubc2) is required for silencing in Saccharomyces cerevisiae.* Mol. Cell. Biol., 1997. **17**(11): p. 6693-9.

27. Dohmen, R.J., et al., *The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 7351-7355.

28. Sung, P., et al., *Yeast RAD6 encoded ubiquitin conjugating enzyme mediates protein degradation dependent on the N-end-recognizing E3 enzyme.* EMBO J., 1991. **10**(8): p. 2187-93.

29. Watkins, J.F., et al., *The extremely conserved amino-terminus of RAD6 ubiquitin-conjugating enzyme is essential for N-end rule-dependent protein degradation.* Genes Dev., 1993. **7**: p. 250-261.

30. Bailly, V., et al., *Specific complex formation between yeast RAD6 and RAD18 proteins: a potential mechanism for targeting RAD6 ubiquitin-conjugating activity to DNA damage sites.* Genes Dev., 1994. **8**(7): p. 811-20.

31. Bailly, V., et al., *Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities.* J. Biol. Chem., 1997. **272**(37): p. 23360-5.

32. Madura, K., R.J. Dohmen, and A. Varshavsky, *N-recognin/Ubc2 interactions in the N-end rule pathway.* J. Biol. Chem., 1993. **268**: p. 12046-12054.

33. Goebl, M.G., et al., *The yeast cell cycle gene CDC34 encodes a ubiquitin-conjugating enzyme.* Science, 1988. **241**: p. 1331-1335.

34. King, R.W., et al., *How proteolysis drives the cell cycle.* Science, 1996. **274**(5293): p. 1652-1659.

35. Skowyra, D., et al., *Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1 [see comments].* Science, 1999. **284**: p. 662-665.

36. Seol, J.H., et al., *Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34.* Genes Dev., 1999. **13**: p. 1614-1626.

37. Sommer, T. and S. Jentsch, *A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum.* Nature, 1993. **365**(6442): p. 176-9.

38. Wiebel, F.F. and W.H. Kunau, *The Pas2 protein essential for peroxisome biogenesis is related to ubiquitin-conjugating enzymes.* Nature, 1992. **359**(6390): p. 73-6.

39. Hershko, A., et al., *Components of ubiquitin-protein ligase system. Resolution, affinity purification, and role in protein breakdown.* J. Biol. Chem., 1983. **258**(13): p. 8206-14.

40. Kwon, Y.T., et al., *The mouse and human genes encoding the recognition component of the N-end rule pathway.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 7898-7903.

41. Hochstrasser, M., *Ubiquitin-dependent protein degradation.* Annu. Rev. Genet., 1996. **30**: p. 405-439.

42. Pickart, C., *Mechanisms underlying ubiquitination.* Annu. Rev. Biochem., 2001. **70**: p. 503-533.

43. Scheffner, M., U. Nuber, and J.M. Huibregtse, *Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade.* Nature, 1995. **373**: p. 81-83.

44. Scheffner, M., et al., *The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53.* Cell, 1993. **75**: p. 495-505.

45. Huibregtse, J.M., J.C. Yang, and S.L. Beaudenon, *The large subunit of RNA polymerase II is a substrate of the Rsp5 ubiquitin-protein ligase.* Proc. Natl. Acad. Sci. USA, 1997. **94**(8): p. 3656-61.

46. Wang, G., J. Yang, and J.M. Huibregtse, *Functional domains of the Rsp5 ubiquitin-protein ligase.* Mol. Cell. Biol., 1999. **19**: p. 342-352.

47. Hicke, L., *Getting down with ubiquitin: turning off cell surface receptors, transporters and channels.* Trends Cell Biol., 1999. **9**: p. 107-112.

48. Saurin, A.J., et al., *Does this have a familiar RING?* Trends Biochem. Sci., 1996. **21**: p. 208-214.

49. Zheng, N., et al., *Structure of a c-Cbl-UbcH7 complex: RING domain function in ubiquitin- protein ligases.* Cell, 2000. **102**(4): p. 533-9.

50. Kinzler, K.W. and B. Vogelstein, *Lessons from hereditary colorectal cancer.* Cell, 1996. **87**: p. 159-170.

51. Rubinfeld, B., et al., *Association of the APC gene product with beta-catenin.* Science, 1993. **262**(5140): p. 1731-4.

52. Su, L.K., B. Vogelstein, and K.W. Kinzler, *Association of the APC tumor suppressor protein with catenins.* Science, 1993. **262**(5140): p. 1734-7.

53. Rubinfeld, B., et al., *Stabilization of beta-catenin by genetic defects in melanoma cell lines.* Science, 1997. **275**(5307): p. 1790-2.

54. Ohta, T., et al., *Roc1, a homolog of Apc11, represents a family of cullin partners with an associated ubiquitin ligase activity.* Mol. Cell, 1999. **3**: p. 535-541.

55. Zachariae, W., et al., *Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins.* Science, 1998. **279**(5354): p. 1216-1219.

56. Deshaies, R.J., *SCF and cullin/RING-H2-based ubiquitin ligases.* Annu. Rev. Cell Dev. Biol., 1999. **15**: p. 435-467.

57. Laney, J.D. and M. Hochstrasser, *Substrate targeting in the ubiquitin system.* Cell, 1999. **97**(4): p. 427-430.

58. Lammer, D., et al., *Modification of yeast Cdc53p by the ubiquitin-related protein rub1p affects function of the SCFCdc4 complex.* Genes Dev., 1998. **12**(7): p. 914-26.

59. Lyapina, S., et al., *Promotion of NEDD-CUL1 conjugate cleavage by COP9 signalosome.* Science, 2001. **292**(5520): p. 1382-5.

60. Zachariae, W. and K. Nasmyth, *Whose end is destruction: cell division and the anaphase-promoting complex.* Genes Dev., 1999. **13**(16): p. 2039-2058.

61. Jackson, P.K., et al., *The lore of the RINGs: substrate recognition and catalysis by ubiquitin ligases.* Trends Cell Biol., 2000. **10**(10): p. 429-39.

62. Hilt, W. and D.H. Wolf, *Proteasomes: destruction as a programme.* Trends Biochem. Sci., 1996. **21**(3): p. 96-102.

63. Baumeister, W., et al., *The proteasome: paradigm of a self-compartmentalizing protease.* Cell, 1998. **92**: p. 367-380.

64. Glickman, M.H., et al., *The regulatory particle of the Saccharomyces cerevisiae proteasome.* Mol. Cell. Biol., 1998. **18**(6): p. 3149-3162.

65. Lowe, J., et al., *Crystal structure of the 20S proteasome from the archaeon T. acidophilum at 3.4 A resolution.* Science, 1995. **268**(5210): p. 533-9.

66. Seemuller, E., et al., *Proteasome from Thermoplasma acidophilum: a threonine protease.* Science, 1995. **268**(5210): p. 579-82.

67. Groll, M., et al., *Structure of 20S proteasome from yeast at 2.4 Å resolution.* Nature, 1997. **386**: p. 463-471.

68. DeMartino, G.N. and C.A. Slaughter, *The proteasome, a novel protease regulated by multiple mechanisms.* J. Biol. Chem., 1999. **274**: p. 22123-22126.

69. Whitby, F.G., et al., *Structural basis for the activation of 20S proteasomes by 11S regulators.* Nature, 2000. **408**(6808): p. 115-20.

70. Groll, M., et al., *A gated channel into the proteasome core particle.* Nat. Struct. Biol., 2000. **7**(11): p. 1062-7.

71. Glickman, M.H., et al., *A subcomplex of the proteasome regulatory particle required for ubiquitin-conjugate degradation and related to the COP9-signalosome and eIF3.* Cell, 1998. **94**(5): p. 615-623.

72. Braun, B.C., et al., *The base of the proteasome regulatory particle exhibits chaperone-like activity.* Nat. Cell Biol., 1999. **1**(4): p. 221-6.

73. Kohler, A., et al., *The axial channel of the proteasome core particle is gated by the Rpt2 ATPase and controls both substrate entry and product release.* Mol. Cell, 2001. **7**(6): p. 1143-52.

74. Thrower, J.S., et al., *Recognition of the polyubiquitin proteolytic signal.* EMBO J., 2000. **19**(1): p. 94-102.

75. Pickart, C., *Ubiquitin in chains.* Trends Biochem. Sci., 2000. **25**: p. 544-548.

76. Deveraux, Q., et al., *A 26S protease subunit that binds ubiquitin conjugates.* J. Biol. Chem., 1994. **269**(10): p. 7059-7061.

77. van Nocker, S., et al., *The multiubiquitin-chain-binding protein Mcb1 is a component of the 26S proteasome in Saccharomyces cerevisiae and plays a nonessential, substrate-specific role in protein turnover.* Mol. Cell. Biol., 1996. **16**(11): p. 6020-6028.

78. Xie, Y. and A. Varshavsky, *Physical association of ubiquitin ligases and the 26S proteasome.* Proc. Natl. Acad. Sci. USA, 2000. **97**(6): p. 2497-2502.

79. You, J. and C.M. Pickart, *A HECT domain E3 enzyme assembles novel polyubiquitin chains.* J. Biol. Chem., 2001. **276**(23): p. 19871-8.

80. Verma, R., et al., *Proteasome proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes.* Mol. Biol. Cell, 2000. **11**: p. 3425-3439.

81. Bachmair, A., D. Finley, and A. Varshavsky, *In vivo half-life of a protein is a function of its amino-terminal residue.* Science, 1986. **234**: p. 179-186.

82. Gilchrist, C.A., D.A. Gray, and R.T. Baker, *A ubiquitin-specific protease that efficiently cleaves the ubiquitin-proline bond.* J. Biol. Chem., 1997. **272**: p. 32280-32285.

83. Davydov, I.V., D. Patra, and A. Varshavsky, *The N-end rule pathway in Xenopus egg extracts.* Arch. Biochem. Biophys., 1998. **357**: p. 317-325.

84. Gonda, D.K., et al., *Universality and structure of the N-end rule.* J. Biol. Chem., 1989. **264**: p. 16700-16712.

85. Lévy, F., et al., *Using ubiquitin to follow the metabolic fate of a protein.* Proc. Natl. Acad. Sci. USA, 1996. **93**: p. 4907-4912.

86. Kwon, Y.T., et al., *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., 2000. **20**(11): p. 4135-4148.

87. Potuschak, T., et al., *PRT1 of Arabidopsis thaliana encodes a component of the plant N-end rule pathway.* Proc. Natl. Acad. Sci. USA, 1998. **95**(14): p. 7904-7908.

88. Tobias, J.W., et al., *The N-end rule in bacteria.* Science, 1991. **254**: p. 1374-1377.

89. Varshavsky, A., et al., *The N-end rule pathway.*, in *Ubiquitin and the Biology of the Cell*, J.-M. Peters, J.R. Harris, and D. Finley, Editors. 1998, Plenum Press: New York. p. 223-278.

90. Baker, R.T. and A. Varshavsky, *Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway.* J. Biol. Chem., 1995. **270**(20): p. 12065-12074.

91. Bachmair, A. and A. Varshavsky, *The degradation signal in a short-lived protein.* Cell, 1989. **56**: p. 1019-1032.

92. Suzuki, T. and A. Varshavsky, *Degradation signals in the lysine-asparagine sequence space.* EMBO J., 1999. **18**(21): p. 6017-6026.

93. Johnson, E.S., D.K. Gonda, and A. Varshavsky, *Cis-trans recognition and subunit-specific degradation of short-lived proteins.* Nature, 1990. **346**: p. 287-291.

94. Gonzalez, M., et al., *Subunit-specific degradation of the UmuD/D' heterodimer by the ClpXP protease: the role of trans recognition in UmuD' stability.* EMBO J., 2000. **19**(19): p. 5251-8.

95. Verma, R., et al., *Selective degradation of ubiquitinated Sic1 by purified 26S proteasome yields active S phase Cyclin-Cdk.* Mol. Cell, 2001. **8**: p. 439-448.

96. Varshavsky, A., *The N-end rule.* Cell, 1992. **69**: p. 725-735.

97. Sakamoto, K.M., et al., *Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation.* Proc. Natl. Acad. Sci. USA, 2001. **98**(15): p. 8554-9.

98. Dohmen, R.J., P. Wu, and A. Varshavsky, *Heat-inducible degron: a method for constructing temperature-sensitive mutants.* Science, 1994. **263**: p. 1273-1276.

99. Labib, K., J.A. Tercero, and J.F. Diffley, *Uninterrupted MCM2-7 function required for DNA replication fork progression.* Science, 2000. **288**: p. 1643-7.

100. Bartel, B., I. Wünning, and A. Varshavsky, *The recognition component of the N-end rule pathway.* EMBO J., 1990. **9**: p. 3179-3189.

101. Madura, K. and A. Varshavsky, *Degradation of Gα by the N-end rule pathway.* Science, 1994. **265**: p. 1454-1458.

102. Lévy, F., J.A. Johnston, and A. Varshavsky, *Analysis of a conditional degradation signal in yeast and mammalian cells.* Eur. J. Biochem., 1999. **259**(1-2): p. 244-52.

103. Johnston, J.A., et al., *Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway.* J. Biol. Chem., 1995. **270**(14): p. 8172-8.

104. Lee, C., et al., *ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal.* Mol. Cell, 2001. **7**(3): p. 627-37.

105. Turner, G.C. and A. Varshavsky, *Detecting and measuring cotranslational protein degradation in vivo.* Science, 2000. **289**: p. 2117-2120.

106. Reiss, Y., D. Kaim, and A. Hershko, *Specificity of binding of N-terminal residues of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites.* J. Biol. Chem., 1988. **263**: p. 2693-2698.

107. Baker, R.T. and A. Varshavsky, *Inhibition of the N-end rule pathway in living cells.* Proc. Natl. Acad. Sci. USA, 1991. **87**: p. 2374-2378.

108. Xie, Y. and A. Varshavsky, *The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain.* EMBO J., 1999. **18**: p. 6832-6844.

109. Turner, G.C., F. Du, and A. Varshavsky, *Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway.* Nature, 2000. **405**(6786): p. 579-583.

110. Du, F., F. Navarro-Garcia, and A. Varshavsky, *pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain.* manuscript in preparation, 2001.

111. Soffer, R.L., *Biochemistry and biology of aminoacyl-tRNA-protein transferases.*, in *Transfer RNA: Biological Aspects.*, D. Söll, J. Abelson, and P.R. Schimmel, Editors. 1980, Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY. p. 493-505.

112. Ferber, S. and A. Ciechanover, *Role of arginine-tRNA in protein degradation by the ubiquitin pathway.* Nature, 1987. **326**(6115): p. 808-11.

113. Shrader, T.E., J.W. Tobias, and A. Varshavsky, *The N-end rule in Escherichia coli: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene aat.* J. Bacteriol., 1993. **175**: p. 4364-4374.

114. Davydov, I.V. and A. Varshavsky, *RGS4 is arginylated and degraded by the N-end rule pathway in vitro.* J. Biol. Chem., 2000. **275**: p. 22931-22941.

115. Balzi, E., et al., *Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of Saccharomyces cerevisiae.* J. Biol. Chem., 1990. **265**: p. 7464-7471.

116. Ciechanover, A., et al., *Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. Its involvement in post-translational modification and degradation of acidic NH2 termini substrates of the ubiquitin pathway.* J. Biol. Chem., 1988. **263**(23): p. 11155-67.

117. Kwon, Y.T., A.S. Kashina, and A. Varshavsky, *Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway.* Mol. Cell. Biol., 1999. **19**(1): p. 182-193.

118. Bradshaw, R.A., W.W. Brickey, and K.W. Walker, *N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families.* Trends Biochem. Sci., 1998. **23**(7): p. 263-7.

119. Tetaz, T., et al., *Relaxed specificity of endoproteinase Asp-N: this enzyme cleaves at peptide bonds N-terminal to glutamate as well as aspartate and cysteic acid residues.* Biochem. Int., 1990. **22**(3): p. 561-6.

120. Yagi, T., H. Kagamiyama, and M. Nozaki, *Cysteine sulfinate transamination activity of aspartate aminotransferases.* Biochem. Biophys. Res. Commun., 1979. **90**(2): p. 447-52.

121. Parsons, R.B., et al., *Cysteine dioxygenase: regional expression of activity in rat brain.* Neurosci. Lett., 1998. **248**(2): p. 101-4.

122. Tu, Y., et al., *Palmitoylation of a conserved cysteine in the regulator of G protein signaling (RGS) domain modulates the GTPase-activating activity of RGS4 and RGS10.* J. Biol. Chem., 1999. **274**: p. 38260-38267.

123. Pace, H.C. and C. Brenner, *The nitrilase superfamily: classification, structure and function.* Genome Biology, 2001. **2**(1): p. 1-9.

124. Miceli, R.M., M.E. DeGraaf, and H.D. Fischer, *Two-stage selection of sequences from a random phage display library delineates both core residues and permitted structural range within an epitope.* J. Immunol. Methods, 1994. **167**(1-2): p. 279-87.

125. Stewart, A.E., S.M. Arfin, and R.A. Bradshaw, *The sequence of porcine protein NH2-terminal asparagine amidohydrolase. A new component of the N-end rule pathway.* J. Biol. Chem., 1995. **270**: p. 25-28.

126. Grigoryev, S., et al., *A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway.* J. Biol. Chem., 1996. **271**: p. 28521-28532.

127. Ota, I.M. and A. Varshavsky, *A yeast protein similar to bacterial two-component regulators.* Science, 1993. **262**: p. 566-569.

128. Alagramam, K., F. Naider, and J.M. Becker, *A recognition component of the ubiquitin system is required for peptide transport in Saccharomyces cerevisiae.* Mol. Microbiol., 1995. **15**: p. 225-234.

129. Byrd, C., G.C. Turner, and A. Varshavsky, *The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor.* EMBO J., 1998. **17**: p. 269-277.

130. deGroot, R.J., et al., *Sindbis virus RNA polymerase is degraded by the N-end rule pathway.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 8967-8971.

131. Mulder, L.C.F. and M.A. Muesing, *Degradation of HIV-1 integrase by the N-end rule pathway.* J. Biol. Chem., 2000. **275**: p. 29749-29753.

132. Uhlmann, F., F. Lottspeich, and K. Nasmyth, *Sister-chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1.* Nature, 1999. **400**(6739): p. 37-42.

133. Rao, H., et al., *Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability.* Nature, 2001. **410**: p. 955-960.

134. Sijts, A.J., I. Pilip, and E.G. Pamer, *The Listeria monocytogenes-secreted p60 protein is an N-end rule substrate in the cytosol of infected cells. Implications for major histocompatibility complex class I antigen processing of bacterial proteins.* J. Biol. Chem., 1997. **272**: p. 19261-19268.

135. Moors, M.A., V. Auerbuch, and D.A. Portnoy, *Stability of the Listeria monocytogenes ActA protein in mammalian cells is regulated by the N-end rule pathway.* Cell. Microbiol., 1999. **1**(3): p. 249-57.

136. Island, M.D., et al., *Isolation and characterization of S. cerevisiae mutants deficient in amino acid-inducible peptide transport.* Curr. Genet., 1991. **20**: p. 457-463.

137. Perry, J.R., et al., *Isolation and characterization of a Saccharomyces cerevisiae peptide transport gene.* Mol. Cell. Biol., 1994. **14**(1): p. 104-15.

138. Barnes, D., et al., *PTR3, a novel gene mediating amino acid-inducible regulation of peptide transport in Saccharomyces cerevisiae.* Mol. Microbiol., 1998. **29**(1): p. 297-310.

139. Klasson, H., G.R. Fink, and P.O. Ljungdahl, *Ssy1p and Ptr3p are plasma membrane components of a yeast system that senses extracellular amino acids.* Mol. Cell. Biol., 1999. **19**(8): p. 5405-16.

140. Forsberg, H. and P.O. Ljungdahl, *Genetic and biochemical analysis of the yeast plasma membrane Ssy1p- Ptr3p-Ssy5p sensor of extracellular amino acids.* Mol. Cell. Biol., 2001. **21**(3): p. 814-26.

141. Knight, S.A.B., et al., *Identification and analysis of a Saccharomyces cerevisiae copper homeostasis gene encoding a homeodomain protein.* Mol. Cell. Biol., 1994. **14**: p. 7792-7804.

142. Uhlmann, F. and K. Nasmyth, *Cohesion between sister chromatids must be established during DNA replication.* Curr. Biol., 1998. **8**(20): p. 1095-1101.

143. Uhlmann, F., et al., *Cleavage of cohesin by the CD-clan protease separin triggers anaphase in yeast.* Cell, 2000. **103**: p. 375-386.

144. Cohen-Fix, O., et al., *Anaphase initiation in Saccharomyces cerevisiae is controlled by the APC-dependent degradation of the anaphase inhibitor Pds1p.* Genes Dev., 1996. **10**(24): p. 3081-3093.

145. Spencer, F., et al., *Mitotic chromosome transmission fidelity mutants in Saccharomyces cerevisiae.* Genetics, 1990. **124**(2): p. 237-249.

146. Hauf, S., I.C. Waizenegger, and J.M. Peters, *Cohesin cleavage by separase required for anaphase and cytokinesis in human cells.* Science, 2001. **293**(5533): p. 1320-3.

147. Fawcett, J. and T. Pawson, *Signal transduction. N-WASP regulation--the sting in the tail.* Science, 2000. **290**(5492): p. 725-6.

148. Prehoda, K.E., et al., *Integration of multiple signals through cooperative regulation of the N- WASP-Arp2/3 complex.* Science, 2000. **290**(5492): p. 801-6.

149. Koshland, D.E., Jr., A. Goldbeter, and J.B. Stock, *Amplification and adaptation in regulatory and sensory systems.* Science, 1982. **217**(4556): p. 220-5.

150. Ferrell, J.E., Jr. and E.M. Machleder, *The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes.* Science, 1998. **280**(5365): p. 895-8.

151. Johnson, P.R., et al., *Degradation signal masking by heterodimerization of MATalpha2 and MATa1 blocks their mutual destruction by the ubiquitin-proteasome pathway.* Cell, 1998. **94**(2): p. 217-27.

Chapter 2

Establishment of an *in vitro* ubiquitylation system with purified components of the *S. cerevisiae* N-end rule pathway

Fangyong Du, Ailsa Webster, Glenn C. Turner and Alexander Varshavsky*

Introduction

Biochemical dissection of the ATP-dependent proteolytic system in rabbit reticulocyte lysate in the early 1980s led to the discovery of the role of Ub in protein degradation, and the enzymes responsible for Ub conjugation to protein substrates [1] [2]. Since then, reconstitutions of the ATP-dependent ubiquitylation and degradation, using cell-free systems, including reticulocyte lysate, yeast extract and systems composed of purified proteins, have continued to yield insights into the Ub system. For example, E6-AP, the prototype of HECT-domain E3s, was found to form an obligatory thioester bond with Ub before it transfers Ub to its substrate [3]. Reconstitution of the *S. cerevisiae* UFD pathway with purified components revealed that, in conjugation with E1, E2 and E3, a Ub chain assembly factor, termed E4, binds to the Ub moieties of preformed Ub conjugates, and is required for efficient multiubiquitylation of UFD substrates [4]. An *in vitro* ubiquitylation system composed of purified recombinant proteins that were identified genetically to be important in the G1/S transition in *S. cerevisiae*, illuminated the essential role of the Ub system in the G1/S transition, and established a paradigm of how an F-box protein of an SCF-family E3 specifically recognizes a phosphorylated protein substrate [5]. An essential RING-H2 finger protein HRT1 was discovered later to be required for efficient multiubiquitylation of phosphorylated substrates by SCF-family E3s [6] [7].

In this chapter, I describe reconstitution of *in vitro* ubiquitylation of the N-end rule substrates lysozyme and CUP9, with purified components of the *S. cerevisiae* N-end rule pathway consisting of E1 (UBA1), E2 (RAD6) and the RING-H2 finger E3 (UBR1). In Chapter 3, the reconstituted *in vitro* ubiquitylation system was employed to demonstrate that UBR1-dependent

multi-ubiquitylation of CUP9 was stimulated by type 1 or type 2 dipeptides. UBR1 was therefore the first E3 of the Ub system shown to be allosterically regulated by small compounds [8].

Materials and methods

Yeast and bacterial strains

The *S. cerevisiae* strain SC295 (*MATa GAL4 GAL80 ura3-52 leu2-3,112 reg1-501 gal1 pep4-3*) was a gift of Dr. S. A. Johnston (Univ. of Texas, SW Medical Center, Dallas, TX, USA). It was used to overexpress *S. cerevisiae* $UBA1^{h}$ and ^{f}UBR1 because of its low protease activity [9]. The *E. coli* strain BL21(DE3) (Novagen, Madison, WI) was used to overexpress all other proteins, except for yeast UBC4, which was expressed in the *E. coli* AR58 strain [10].

Plasmids

The high copy yeast plasmid pJD325 overexpressing *S. cerevisiae* $UBA1^{h}$ (a UBA1 variant with a C-terminal hexahistidine tag) from the P_{CUP1} promoter [5] was a gift of Dr. R. J. Dohmen (Institute for Genetics, Univ. of Cologne, Cologne, Germany). The plasmids pT7-His6Ub and pT7-His6UbK48R expressing, respectively, ^{h}Ub (an *S. cerevisiae* Ub variant with an N-terminal hexahistidine tag) and ^{h}UbK48R (a ^{h}Ub variant with Lys48→Arg mutation) in *E. coli*, originated from Dr. J. Callis (Univ. of California, Davis). The plasmid PL_{λ}-Ubc4, which expresses *S. cerevisiae* UBC4 in *E. coli*, originated from Dr. Vince Chau [10]. The plasmid pET-11d-Rad6 expressing *S. cerevisiae* RAD6 in *E. coli* was constructed by inserting the *RAD6 ORF* between the unique *Nco I* and *BamH I*

sites of the pET-11d vector (Novagen, Madison, WI). The *RAD6 ORF* was amplified by PCR using the primers AW17 (5′-GGGACCATGGGGTCCAC ACCAGCTAGAAGAAGGTTG-3′) and AW18 (5′-AATCGGATCCGAATT CATAATATCGGCTCGGCA-3′). The plasmid pNTFlagUBR1 was constructed to overexpress ^{f}UBR1 (a UBR1 variant with an N-terminal FLAG epitope tag) from the p_{ADH1} promoter on a Yeplac181-based high copy plasmid [11]. The N-terminal FLAG tag was added through PCR with the primers AW7 (5′-GCCA GTCGACATG*GACTACAAGGACGACGATGACAAG*GGAGGTTCCTCCGTTG CTGATGATGATTTAGGA-3′) and UBR1PC2 (5′-CCAACTAGTGAATTCAGT ACATATATC-3′), the *Sal I* and *Spe I* sites were underlined, and the DNA sequence encoding the FLAG tag was in *italic* (a small linker Gly-Gly-Ser was inserted after the FLAG tag). This PCR fragment was inserted between the unique *Sal I* and *Spe I* sites of the plasmid pAWUBR1 (A. Webster, M. Ghislain and A. Varshavsky, unpublished data). The plasmid pET-11c-^{f}CUP9^{h}, which expresses ^{f}CUP9^{h} (a CUP9 variant with an N-terminal FLAG tag and a C-terminal hexahistidine tag) from the T7 promoter in *E. coli*, was constructed by inserting the *CUP9 ORF* between the unique *Nde I* and *BamH I* sites of the pET-11c vector (Novagen, Madison, WI). The *CUP9 ORF* was amplified by PCR with the primers GTO113 (5′-GCCGCATATGGGC*GATTATAAAGATGACGATGACA AA*AATTATAAC TGCGAAATACA-3′) and GTO55 (5′-GCCGGGATCCTTA*AT GATGA TGATGATGAT*GAGTACCACC AGAATTCATATCAGGGTTGGATAG-3′), the *Nde I* and *BamH I* sites were underlined, and the DNA sequences encoding the FLAG tag in GTO113 and the hexahistidine tag in GTO55 were in *italic*. A small linker Ser-Gly-Gly-Thr was added before the hexahistidine tag.

Overexpression of the recombinant proteins and their purification

***S. cerevisiae* UBA1^h**—UBA1^h was overexpressed in *S. cerevisiae* SC295 and purified as described [5]. After Ni-NTA and Ub "covalent" affinity chromatography, UBA1^h was loaded onto a calibrated HiLoad 16/60 Superdex 200pg gel-filtration column equilibrated with buffer F (25 mM HEPES (pH 7.6), 15% glycerol, 0.1 M KCl, 0.1 mM DTT). UBA1^h eluted as a symmetrical peak centered at 67.0 ml, suggesting that its apparent molecular weight is ~161 kDa, which is significantly higher than the molecular weight calculated from *UBA1 ORF* (~115 kDa). The purified UBA1^h was concentrated to 1.1 mg/ml, frozen in liquid N_2, and stored at -80°C.

***S. cerevisiae* RAD6** — A single colony of BL21(DE3) transformed with pET-11d-Rad6 was grown in 4 l of LB medium in the presence of carbenicillin (50 μg/ml) to A_{600} ~ 1.0. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM, followed by a 6-hr incubation at 23°C. The cells were harvested by centrifugation, washed once with PBS, and frozen in liquid N_2. The pellet was resuspended in 30 ml of lysis buffer (20 mM K_2HPO_4-KH_2PO_4 (pH 7.4), 10% glycerol, 1 mM EDTA, 0.1 mM DTT, 10 mM β-mercaptoethanol) in the presence of protease inhibitors (1 mM PMSF, 1 mM pepstatin, 50 μM AEBSF). Freshly dissolved chicken egg white lysozyme in lysis buffer was added to a final concentration of 0.5 mg/ml, followed by incubation at 4°C for 30 min, and the cells were further disrupted by sonication. The lysate was clarified by centrifugation at 11,200*g* for 30 min, and the supernatant was loaded onto a DEAE column equilibrated with lysis buffer. The DEAE column was developed with 20-column volumes linear gradient of KCl (0-1 M) in the

lysis buffer, and the fractions containing RAD6 were identified by SDS-15% PAGE, pooled and dialyzed into buffer B (20 mM K_2HPO_4-KH_2PO_4 (pH 7.4), 10% glycerol, 50 mM KCl, 1 mM DTT). The dialyzed sample was loaded onto a 5 ml Q-Sepharose column equilibrated with buffer B, and then developed with a 50 ml linear gradient of KCl (0.05 -1M) in 20 mM K_2HPO_4-KH_2PO_4 (pH 7.4), 10% glycerol, 1 mM DTT. RAD6 eluted at ~0.5 M KCl. The fractions containing RAD6 were pooled, dialyzed into buffer D (20 mM Tris-HCl (pH 7.9), 10% glycerol, 0.5 mM DTT, 0.2 mM EDTA), and loaded onto 1 ml FPLC Mono-Q HR5/5 ion-exchange column equilibrated in buffer D. The Mono-Q column was developed with 10 ml linear gradient of KCl (0.3 - 0.55 M) in buffer D. RAD6 eluted at ~0.42 M KCl. The pooled RAD6 fractions were concentrated to 10 mg/ml, frozen in liquid N_2 and stored at -80°C. RAD6 used in this study was further fractionated through a calibrated Superdex 75 (HR 10/30) gel-filtration column equilibrated in buffer F (20 mM HEPES (pH 7.2), 10% glycerol, 50 mM KCl, 0.2 mM EDTA, 0.1 mM DTT). RAD6 eluted as a symmetrical peak centered ~11.8 ml, suggesting that its apparent molecular weight is ~29 kDa, which is significantly higher than the molecular weight calculated from *RAD6 ORF* (19.8 kDa). The anomalously large apparent molecular weight of RAD6 was reported to be caused by its polyacidic tail [12].

***S. cerevisiae* UBC4** —UBC4 was overexpressed in *E. coli*, and purified as described [10]. The purified UBC4 eluted from a calibrated Superdex 75 (HR 10/30) gel-filtration column as a symmetrical peak centered ~12.6 ml in buffer E (10 mM HEPES (pH 7.2), 10% glycerol, 10 mM NaCl, 0.1 mM EDTA, 0.1 mM DTT), suggesting that its apparent molecular weight is ~21 kDa, which is close to

the molecular weight calculated from *UBC4 ORF* (16.4 kDa). Therefore, UBC4 is most likely a monomer in buffer E.

***S. cerevisiae* ^{f}UBR1**—A single colony of SC295 strain transformed with pNTFlagUBR1 was inoculated into 2 l of plasmid-retaining SD medium, and grown at 30°C to A_{600} ~1.0, followed by addition of an equal volume of YPD medium. The cells were grown for three more generations until A_{600} ~4.0, then harvested by centrifugation, washed once with cold PBS, and frozen in liquid N_2. The frozen pellet was ground to fine powder in liquid N_2 using a mortar and pestle, and resuspended (6 ml of buffer per 1 g of pellet) in the lysis buffer (50 mM HEPES (pH 7.5), 10% glycerol, 0.5% NP40, 0.2 M KCl, 1 mM DTT) containing protease inhibitors (1 protease inhibitors cocktail tablet (Roche, Indianapolis, IN, USA) per 10 ml buffer). The suspension was centrifuged at 11,200*g* for 30 min, and the supernatant was then filtered through 0.80 μm filter. The filtrate was mixed with 2 ml of anti-FLAG M2 affinity gel (Sigma, St. Louis, MO) at 4°C for 1 hr. The beads were collected and washed sequentially with 10 ml of lysis buffer, 10 ml of buffer A (50 mM HEPES (pH 7.5), 10% glycerol, 0.5% NP40, 1 M KCl, 1 mM EDTA, 1 mM DTT) and 10 ml of buffer B (lysis buffer minus 0.5% NP40). ^{f}UBR1 was eluted with buffer C (buffer B with 0.5 mg/ml FLAG peptide). The eluant was loaded onto a RAD6-affinity column, which was prepared by immobilizing purified RAD6 (8 mg) to 1.2 ml of Affi-Gel 15 (Bio-Rad, Hercules, CA) following manufacturer's instructions. The column was washed with 10 ml of buffer D (50 mM HEPES (pH 7.5), 10% glycerol, 0.2 M KCl, 1 mM EDTA 1 mM DTT), and ^{f}UBR1 was eluted with 6 ml of buffer E (buffer D plus 2 M KCl). After concentration with Centricon (MWCO 100K) (Millipore,

Bedford, MA), the eluant was loaded onto a calibrated HiLoad 16/60 Superdex 200pg gel-filtration column equilibrated in buffer F (20 mM HEPES (pH 7.5), 10% glycerol, 0.4 M KCl, 0.2 mM DTT, 0.2 mM EDTA). ^{f}UBR1 eluted at 58.7 ml, suggesting that its apparent molecular weight is ~360 kDa, which is significantly higher than the molecular weight calculated from *UBR1 ORF* (225 kDa). Purified ^{f}UBR1 was concentrated to 1.0 mg/ml, frozen in liquid N_2, and stored at -80°C.

Radiolabeled ^{f}CUP9^{h}— BL21(DE3) was co-transformed with plasmids pET-11c-^{f}CUP9^{h} and pRI952, the latter expresses tRNAs for the codons AGG, AGA, and AUA, which are rare in *E. coli* [13]. A single *E. coli* transformant was resuspended into 1 ml of LB medium, then diluted 100 fold with fresh LB medium, and 150 µl of diluted culture was spread onto an LB plate containing ampicillin (60 µg/ml) and chloramphenicol (25 µg/ml). After incubation at 37°C for ~18 hr, the *E. coli* colonies on LB plate were scraped, resuspended into 10 ml of labeling DMEM medium, and inoculated into 400 ml of labeling DMEM [14]. The labeling DMEM medium was composed of 25 mM HEPES (pH 7.5) and DMEM with 4.5 g/l glucose without glutamine, methionine and cysteine (ICN, Costa Mesa, CA). The culture was grown at 37°C for ~6 hr until A_{600} ~1.0. IPTG was then added to a final concentration of 1 mM. 15 min later, 1 mCi of [^{35}S] methionine/cysteine (EXPRESS, New England Nuclear, Boston, MA) was added. The culture was grown for another 2 hr at 30°C, and the cells were harvested, washed once with cold PBS and resuspended into 15 ml of buffer A (25 mM K_2HPO_4-KH_2PO_4 (pH 7.8), 10% glycerol, 0.3 M NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol). Freshly dissolved chicken egg white lysozyme in buffer A was added to a final concentration of 0.5 mg/ml, followed by incubation at 4°C

for 30 min. The cells were further disrupted by sonication, and NP40 was added to a final concentration of 0.5%. The suspension was centrifuged at 11,200*g* for 30 min, and the supernatant was mixed with 1 ml Ni-NTA resin (Qiagen, Valencia, CA, USA). After gentle rotation at 4°C for 10 min, the resin was transferred to a column, and washed with 10 ml of buffer A. The bound proteins were eluted with 5 ml of buffer B (buffer A with 0.3 M imidazole), and the eluant was gently mixed with 0.5 ml anti-FLAG M2 affinity gel for 30 min. The affinity gel was collected and washed once with 10 ml of buffer C (20 mM HEPES (pH 7.6), 10% glycerol, 0.15 M NaCl, 0.2 mM DTT). Radiolabeled ^{f}CUP9^{h} was eluted with 3 ml of buffer D (buffer C with 0.2 mg/ml FLAG peptide), and FLAG peptide was removed by dialysis with a Slide-A-Lyzer dialysis cassette (MWCO 10K) (Pierce, Rockford, IL). The specific radioactivity of metabolically radiolabeled ^{f}CUP9^{h} was ~ 2.5 X 10^{4} cpm/μg.

Iodination of lysozyme

Chicken egg white lysozyme was radioiodinated using IODO-Beads (Pierce) and NaI^{125} (100 mCi/ml) (Amersham Pharmacia Biotech) following manufacturers' instructions. The iodinated lysozyme was dialyzed into PBS using a Slide-A-Lyzer Dialysis Cassette (MWCO 2K) (Pierce). The specific radioactivity of iodinated lysozyme was ~3.1 X 10^{5} cpm/μg.

In vitro ubiquitylation assay

The *in vitro* ubiquitylation assays were carried out in buffer U (25 mM HEPES/KOH (pH 7.5), 25 mM KCl, 5 mM $MgCl_2$, 0.1 mM DTT, 2 mM ATP). The

ATP regeneration system (10 mM phosphocreatine, 10 IU creatine phosphokinase/ml) was not used because it did not enhance ubiquitylation efficiency in this system (data not shown). The reactions contained the following components: 7 μM Ub or Ub-derivatives, 50 nM $UBA1^h$, 50 nM RAD6 (or 50 nM UBC4), 50 nM $^fUBR1$ and 550 nM ^{35}S-labeled $^fCUP9^h$ (or 2 μM radioiodinated lysozyme), 0.5 mg/ml ovalbumin (Grade VII) (Sigma) as a carrier protein. Protein stocks were diluted with buffer D (10 mM HEPES (pH 7.5), 20 mM KCl, 0.1 mM DTT). All components except $UBA1^h$ were mixed on ice for 10 min. $UBA1^h$ was then added and reactions shifted to 30°C. After the indicated times, the reactions were terminated by adding an equal volume of 2 X SDS loading buffer (100 mM Tris-HCl (pH 6.8), 4% SDS, 20% glycerol, 0.2 M DTT, 0.1% Bromophenol blue), and heating at 95°C for 5 min. The samples were fractionated by SDS-PAGE.

Results and discussion

Overexpression and purification of components of the yeast N-end rule pathway

Specific components of the yeast N-end rule pathway were overexpressed in *S. cerevisiae* or in *E. coli*, and purified as described (see MATERIALS AND METHODS). Briefly, $UBA1^h$ (a UBA1 variant with a C-terminal hexahistidine tag) was overexpressed in yeast from a Cu^{2+}-inducible promoter on a high copy vector; it was purified to apparent homogeneity (Figure 1a, lane 2) through Ni-NTA and Ub "covalent" affinity chromatography [15], followed by gel-filtration column. The enzymatic activity of $UBA1^h$ was monitored through the formation of DTT-sensitive $UBA1^h$-Ub conjugate (data not shown). RAD6

was overexpressed in *E. coli* from the T7 promoter; it was purified to apparent homogeneity through DEAE, Q-Sepharose, FPLC Mono-Q and gel-filtration columns (Figure 1b, lane 3). RAD6 formed a DTT-sensitive thioester bond with Ub, if and only if $UBA1^h$ was also present (data not shown). UBC4 overexpressed in *E. coli* was purified to apparent homogeneity through Q-fast flow, FPLC Mono-Q and gel-filtration columns (Figure 1b, lane 2), and it formed a DTT-sensitive thioester bond with Ub in the presence of $UBA1^h$ (data not shown). ^{f}UBR1 (a UBR1 variant with an N-terminal FLAG epitope tag) was overexpressed in yeast from the P_{ADH1} promoter on a high copy vector; it was purified to apparent homogeneity by fractionation through monoclonal anti-FLAG antibody and RAD6 affinity chromatography, followed by gel-filtration column (Figure 1a, lane 3). The apparent molecular weight of ^{f}UBR1, deduced from the elution volume from a calibrated gel-filtration column, was ~360 kDa, which is significantly higher than the molecular weight calculated from *UBR1 ORF* (~225 kDa). Similarly, the apparent molecular weight of rabbit E3α, the mammalian homolog of yeast UBR1, was estimated to be ~300 kDa, whereas its M_r on SDS-PAGE gel was ~180 kDa [16]. The deduced molecular weight of rabbit E3α is likely to be close to that of mouse E3α (~200 kDa), whose cDNA has been cloned [17]. To determine whether UBR1 forms dimers, co-immunoprecipitation experiments were carried out with extracts of *ubr1Δ* cells overexpressing $UBR1^{ha}$ (a UBR1 variant with a C-terminal HA tag) and ^{f}UBR1. Anti-HA monoclonal antibody could not immunoprecipitate ^{f}UBR1, or vice versa (data not shown), suggesting that UBR1 does not form dimers, at least not tightly-associated dimers under the conditions that have been tested.

Reconstitution of *in vitro* ubiquitylation of lysozyme, an N-end-rule substrate

The chicken egg white lysozyme, which possesses a type 1 destabilizing N-terminal Lys residue, is turned into an N-end rule substrate after it is iodinated, apparently because of perturbation of its tertiary structuure that exposes a destabilizing N-terminal residue of lysozyme (lysine), and/or makes its internal Lys residiue(s) targetable for ubiquitylation [18]. In the *in vitro* ubiquitylation system that consisted of purified components of the yeast N-end rule pathway, including UBA1^{h}, RAD6, ^{f}UBR1, and Ub in ATP-supplemented reaction buffer, the radioiodinated lysozyme was efficiently conjugated to multiple Ub moieties, manifested in the accumulation of lysozyme-derived, HMW species on top of the gel (Figure 2, lanes 4-6). Substitution of Ub with ^{h}Ub (a Ub variant with an N-terminal hexahistidine tag) resulted in slower migrating conjugates, which were particularly evident for the mono- and di-ubiquitylated lysozyme (Figure 2, lane 2 versus lane 1). Substitution of ^{h}Ub with ^{h}UbK48R (a ^{h}Ub variant with Lys48→Arg mutation) dramatically reduced HMW conjugates (Figure 2, lane 3 versus lane 2), indicating that the multiple Ub moieties conjugated to lysozyme formed multi-Ub chains that were linked mainly through Lys48 of Ub. Both ^{f}UBR1 and RAD6 were required for multi-Ub chain formation, since no Ub conjugates were formed in the absence of either of them (Figure 2, lanes 7-8 versus lane 1).

These results demonstrated that purified components of the yeast N-end rule pathway, including UBA1^{h}, RAD6, ^{f}UBR1, were enzymatically active, and were both necessary and sufficient in assembling Lys48-linked multi-Ub chains

on iodinated lysozyme, an artificial N-end rule substrate with a type 1 destabilizing N-terminal (Lys) residue.

Reconstitution of *in vitro* ubiquitylation of CUP9

The 35 kDa homeodomain protein CUP9, a transcriptional repressor of the di- and tripeptide transporter PTR2, was found to be targeted *in vivo* by the yeast N-end rule pathway [19]. However, CUP9 did not expose a destabilizing N-terminal residue prior to its degradation [19]. Thus it was unclear whether CUP9 was directly recognized by UBR1, the RING-H2 finger E3 of the N-end rule pathway [20]. An alternative model posited that CUP9 might be *trans*-targeted by its ligand bearing a destabilizing N-terminal residue. The *trans*-targeting of protein substrates of the Ub system was first demonstrated with engineered N-end rule substrates in reticulocyte lysate [21].

To determine whether CUP9 was directly recognized by UBR1, and multiubiquitylated in the UBR1-dependent *in vitro* ubiquitylation system, we purified ^{f}CUP9^{h} (a CUP9 variant with an N-terminal FLAG epitope tag and a C-terminal hexahistidine tag) that was overexpressed and metabolically radiolabeled in *E. coli* (Figure 1c, lane 2). ^{f}CUP9^{h} was targeted by the N-end rule pathway, and its *in vivo* half life is similar to that of wild-type CUP9 (data not shown). In the *in vitro* ubiquitylation system that consisted of purified components of the yeast N-end rule pathway, ^{f}CUP9^{h} was efficiently multiubiquitylated (Figure 3a, lane 6 versus lane 8). Substitution of Ub with ^{h}Ub resulted in slower migrating conjugates (Figure 3a, lane 7 versus lane 6), while substitution of Ub with chain-terminating K48RUb or MeUb (BostonBiochem,

Cambridge, MA) dramatically reduced HMW conjugates (Figure 3b), indicating that multi-Ub chains assembled on ^{f}CUP9^{h} were linked mainly (if not exclusively) through Lys48 of Ub. While ^{f}CUP9^{h} was not ubiquitylated at all in the absence of ATP, Ub, UBA1^{h} or RAD6 in the *in vitro* ubiquitylation system (Figure 3a, lanes 1-2 and lanes 4-5); notably, in the absence of ^{f}UBR1, ^{f}CUP9^{h} was mono-ubiquitylated, and remained largely so even after 2-hr incubation (Figure 5a). In contrast, in a complete *in vitro* ubiquitylation reaction that contained ^{f}UBR1, ^{f}CUP9^{h} was conjugated with multi- Ub chains. After 1-hr incubation, a large proportion of these multi-Ub chains contained more than 25 Ub moieties (Figure 5b).

Previous studies suggested that besides RAD6, another Ub-conjugating enzyme UBC4 was also involved in the degradation of engineered N-end rule substrates [22], and the physiological substrate CUP9 [23]. In contrast to RAD6, which forms a tight complex with UBR1 [24] [20], UBC4 does not interact with either UBR1 or RAD6 (data not shown). In the ^{f}UBR1-dependent *in vitro* ubiquitylation system, UBC4 could not replace RAD6 in ubiquitylating ^{f}CUP9^{h} (Figure, 4a, lanes 4-5 versus lanes 2-3), even when its concentration was nine fold higher than that of RAD6 (Figure 4a, lanes 8-9 versus lanes 2-3). UBC4 did not have any effect on the ^{f}UBR1-independent, but RAD6-dependent mono-ubiquitylation of ^{f}CUP9^{h} either (Figure 4b).

These results established that purified components of the yeast N-end rule pathway, including UBA1^{h}, RAD6, ^{f}UBR1, together with Ub and ATP in the buffer, were both necessary and sufficient in forming Lys48-linked multi-Ub chains on ^{f}CUP9^{h}, indicating that ^{f}CUP9^{h} overexpressed in, and purified from *E.*

coli, and thus was not post-translationally modified by yeast proteins, was directly recognized by ^{f}UBR (through a third, distinct substrate-binding site of UBR1 (Chapter 3)). The internal degron of CUP9 that is directly recognized by UBR1 has been mapped to its C-terminal region encompassing 16-residue fragment between the residues 286 to 301. This 16-residue fragment was predicted to form an amphipathic helix (Chapter 5.1).

It remains unclear why ^{f}CUP9^{h} is mono-ubiquitylated in the absence of ^{f}UBR1. One possibility is that RAD6, which carries a polyacidic tail, interacts non-specifically with CUP9, a basic protein (pI 9.5), and transfers activated Ub to a Lys residue of CUP9; however, lysozyme, which is similarly basic (pI 8.9), was not ubiquitylated at all in the absence of ^{f}UBR1 (Figure 2, lane 7). In yeast cells, the Ub that is possibly conjugated to CUP9 by RAD6 is likely to be cleaved by deubiquitinating enzymes. UBC4 could not replace RAD6 in the ^{f}UBR1-dependent *in vitro* ubiquitylation of ^{f}CUP9^{h}, nor could it cooperate with RAD6 to form multi-Ub chains on CUP9 in the absence of ^{f}UBR1, suggesting that UBC4 may be involved in the N-end rule pathway indirectly; for example, UBC4 may upregulate the activity of the N-end rule pathway by increasing the transcription levels of *RAD6* or *UBR1*.

UBR1-mediated multi-Ub chain formation is processive.

It was proposed [18] that E3α-mediated multi-Ub chain formation was processive, based on the observation that Ub moieties continued to be linked to substrate that was conjugated with Ub, even in the excess of "free" substrate. To follow the ^{f}UBR1-mediated ubiquitylation process, we did a time course of

^{f}CUP9^{h} ubiquitylation (Figure 5b). The starting concentration of ^{f}CUP9^{h} was ~9 fold higher than that of ^{f}UBR1. After 1-hr incubation, more than half of ^{f}CUP9^{h} were conjugated with multi-Ub chains (Figure 5b, lane 5), indicating that ^{f}UBR1 dissociated from ^{f}CUP9^{h} -Ub conjugates, and initiated multiple rounds of Ub conjugation to ^{f}CUP9^{h}. Over the entire course, short Ub chains linked to ^{f}CUP9^{h} were continuously converted to longer Ub chains (a multi-Ub chain could contain more than 25 Ub moieties). Close inspection revealed that, in ^{f}UBR1-dependent ubiquitylation of ^{f}CUP9^{h}, the amount of ^{f}CUP9^{h} conjugated with short Ub chains continuously decreased, despite the fact that they were also continuously generated at the same time. For example, mono-ubiquitylated ^{f}CUP9^{h} was produced by conjugation of one Ub moiety to the "free" ^{f}CUP9^{h} and was consumed by its conversion to di-ubiquitylated ^{f}CUP9^{h}; both reactions were mediated by ^{f}UBR1. The continuous decline of the amount of mono-ubiquitylated ^{f}CUP9^{h} indicated that the rate of its production was slower than that of its consumption, despite the constant presence of larger amount of "free" ^{f}CUP9^{h}, the precursor of mono-ubiquitylated ^{f}CUP9^{h}. Similar analysis could be applied to other ^{f}CUP9^{h}-Ub conjugates. Therefore, ^{f}CUP9^{h}–Ub conjugates with longer Ub chains were better substrates for ^{f}UBR1-mediated ubiquitylation reaction than ^{f}CUP9^{h}–Ub conjugates with shorter Ub chains, i.e., ^{f}UBR1-mediated multi-Ub chain formation was processive.

The molecular mechanisms of the processive nature of multi-Ub chain formation remains largely unknown. It was suggested that processivity resulted from the higher affinity of E3s for their ubiquitylated substrates than "free"

substrates [25], presumably because E3s directly bind to Ub or Ub chains. However, preliminary data indicated that ^{f}UBR1 does not have high-affinity interactions with Ub, or Ub chains. Specifically, anti-FLAG antibody could not co-immunoprecipitate Ub or Ub chains from a mixture of partially purified ^{f}UBR1 and Ub chains (containing Ub, Ub_2, to Ub_8) (data not shown). The processive assembly of multi-Ub chains on protein substrates of the Ub system may have been evolved to counter the actions of many deubiquitinating enzymes present in eukaryotic cells.

In summary, I describe in this chapter overexpression and purification of components of the yeast N-end rule pathway, and reconstitution of *in vitro* ubiquitylation of lysozyme and CUP9, an artificial and a physiological substrate of the N-end rule pathway. The multi-Ub chains conjugated to both substrates are linked mainly through Lys48 of Ub, although these two substrates bind to distinct sites of UBR1. How UBR1 controls multi-Ub chain configuration remains unknown. The reconstituted, ^{f}UBR1-dependent *in vitro* ubiquitylation system was employed in Chapter 3 to discover that the activity of UBR1 was allosterically regulated by type 1 or type 2 dipeptides.

References

1. Hershko, A. and A. Ciechanover, *Mechanisms of intracellular protein breakdown.* Annu. Rev. Biochem., 1982. **51**: p. 335-64.
2. Hershko, A. and A. Ciechanover, *The ubiquitin system for protein degradation.* Annu. Rev. Biochem., 1992. **61**: p. 761-807.
3. Scheffner, M., U. Nuber, and J.M. Huibregtse, *Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade.* Nature, 1995. **373**: p. 81-83.
4. Koegl, M., et al., *A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly.* Cell, 1999. **96**(5): p. 635-644.
5. Feldman, R.M., et al., *A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p.* Cell, 1997. **91**(2): p. 221-30.
6. Seol, J.H., et al., *Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34.* Genes Dev., 1999. **13**: p. 1614-1626.
7. Skowyra, D., et al., *Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1 [see comments].* Science, 1999. **284**: p. 662-665.
8. Turner, G.C., F. Du, and A. Varshavsky, *Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway.* Nature, 2000. **405**(6786): p. 579-583.
9. Joshua-Tor, L., et al., *Crystal structure of a conserved protease that binds DNA: the bleomycin hydrolase, Gal6.* Science, 1995. **269**(5226): p. 945-50.

10. Cook, W.J., et al., *Tertiary structures of class I ubiquitin-conjugating enzymes are highly conserved: crystal structure of yeast Ubc4.* Biochemistry, 1993. **32**(50): p. 13809-17.

11. Gietz, R.D. and A. Sugino, *New yeast-Escherichia coli shuttle vectors constructed with in vitro mutagenized yeast genes lacking six-base pair restriction sites.* Gene, 1988. **74**(2): p. 527-34.

12. Morrison, A., E.J. Miller, and L. Prakash, *Domain structure and functional analysis of the carboxyl-terminal polyacidic sequence of the RAD6 protein of Saccharomyces cerevisiae.* Mol. Cell. Biol., 1988. **8**(3): p. 1179-85.

13. Del Tito, B.J., Jr., et al., *Effects of a minor isoleucyl tRNA on heterologous protein translation in Escherichia coli.* J. Bacteriol., 1995. **177**(24): p. 7086-91.

14. Giovane, C., G. Schwalbach, and E. Weiss, *In vivo labeling of over-expressed recombinant proteins in E. coli.* Biotechniques, 1997. **22**(5): p. 796-8.

15. Ciechanover, A., et al., *"Covalent affinity" purification of ubiquitin-activating enzyme.* J. Biol. Chem., 1982. **257**(5): p. 2537-42.

16. Hershko, A., et al., *The protein substrate binding site of the ubiquitin-protein ligase system.* J. Biol. Chem., 1986. **261**(26): p. 11992-9.

17. Kwon, Y.T., et al., *The mouse and human genes encoding the recognition component of the N-end rule pathway.* Proc. Natl. Acad. Sci. USA, 1998. **95**: p. 7898-7903.

18. Hershko, A., et al., *Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis.* Proc. Natl. Acad. Sci. USA, 1980. **77**(4): p. 1783-6.

19. Byrd, C., G.C. Turner, and A. Varshavsky, *The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor.* EMBO J., 1998. **17**: p. 269-277.

20. Xie, Y. and A. Varshavsky, *The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain.* EMBO J., 1999. **18**: p. 6832-6844.

21. Johnson, E.S., D.K. Gonda, and A. Varshavsky, *Cis-trans recognition and subunit-specific degradation of short-lived proteins.* Nature, 1990. **346**: p. 287-291.

22. Bartel, B., *Molecular genetics of the ubiquitin system: the ubiquitin fusion proteins and proteolytic targeting mechanisms*, in *Ph.D. Thesis. Department of Biology, MIT*. 1990: Cambridge, MA.

23. Byrd, C., *Regulation of Peptide Transport*, in *Department of Biology*. 1998, Caltech: Pasadena, CA. p. 155.

24. Dohmen, R.J., et al., *The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme.* Proc. Natl. Acad. Sci. USA, 1991. **88**: p. 7351-7355.

25. Reiss, Y., H. Heller, and A. Hershko, *Binding sites of ubiquitin-protein ligase. Binding of ubiquitin-protein conjugates and of ubiquitin-carrier protein.* J. Biol. Chem., 1989. **264**(18): p. 10378-83.

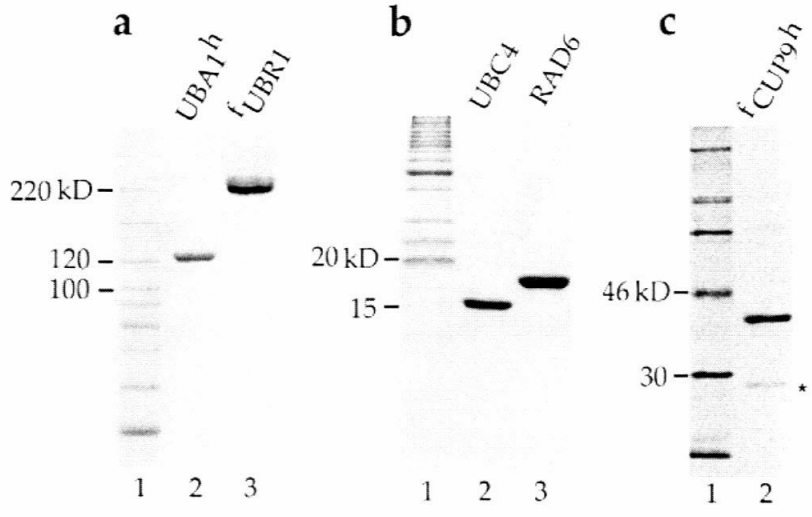


Figure 1 The purified components of the yeast N-end rule pathway, and the radiolabeled N-end rule substrate ^{f}CUP9^{h}. **a-c**, Recombinant proteins were overexpressed in *E. coli* (RAD6, UBC4 and ^{f}CUP9^{h}), or yeast (UBA1^{h} and ^{f}UBR1), and were purified as described in MATERIALS AND METHODS. Purified proteins were displayed on SDS-PAGE gels, and were stained with Coomassie (**a-b**), or visualized by autoradiography (**c**). The molecular weight standards were fractionated in lane 1 of each gel. **a**, 1.6 μg of UBA1^{h} (lane 2) and 3 μg of ^{f}UBR1 (lane 3) were displayed on SDS-8% PAGE. **b**, 5 μg of UBC4 (lane 2) and 5 μg of RAD6 (lane 3) were displayed on SDS-15% PAGE. **c**, 0.2 μg of radiolabeled ^{f}CUP9^{h} (lane 2) was displayed on SDS-10% PAGE, and visualized by autoradiography. An asterisk denotes a contaminating protein.

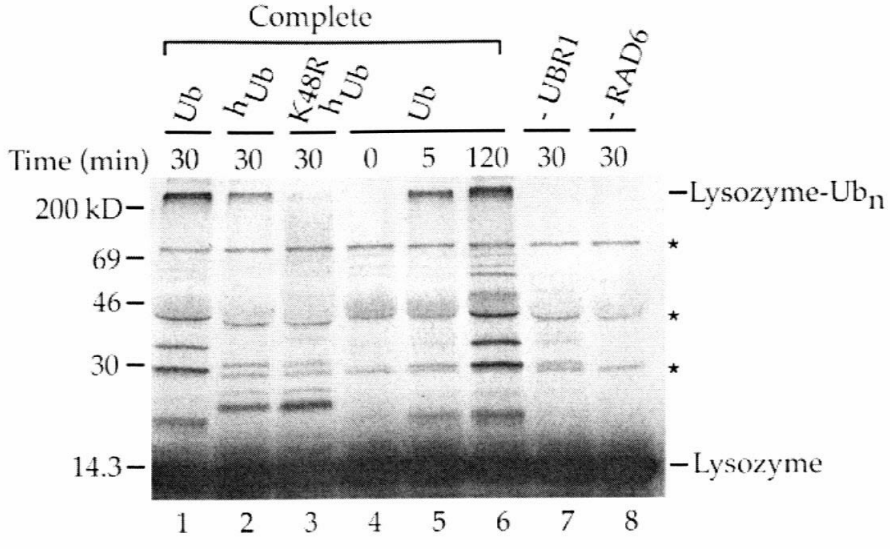


Figure 2 Reconstitution of *in vitro* ubiquitylation of ^{125}I-lysozyme with purified components of the yeast N-end rule pathway. Purified UBA1^{h}, RAD6, ^{f}UBR1 and Ub or Ub derivatives (^{h}Ub for lane 2 and K48R^{h}Ub for lane 3) were mixed with ^{125}I-lysozyme in ATP-supplemented buffer, reaction mixture in lanes 7-8 lacked either ^{f}UBR1 or RAD6 as indicated. Reaction mixtures were incubated at 30°C for the designated times, then fractionated by SDS-12% PAGE, followed by autoradiography. The asterisks denote the contaminants in ^{125}I-lysozyme.

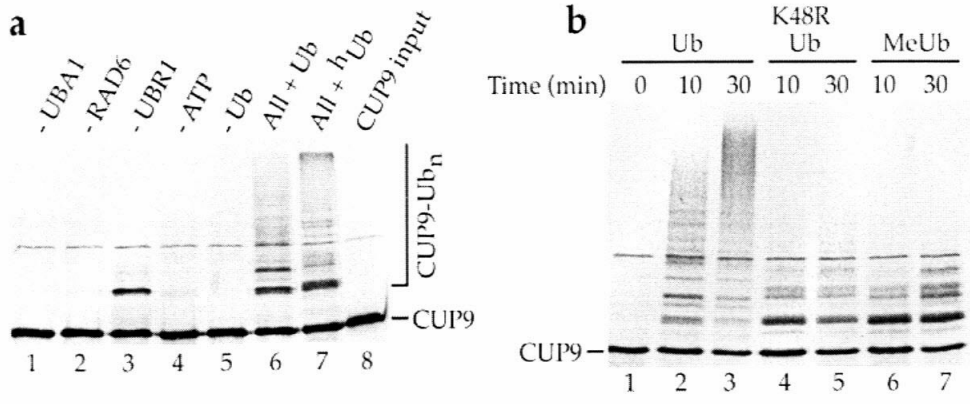


Figure 3 Reconstitution of *in vitro* ubiquitylation of ^{f}CUP9^{h} with purified components of the yeast N-end rule pathway. **a**, Purified UBA1^{h}, RAD6, ^{f}UBR1 and Ub (lane 6) or ^{h}Ub (lane 7) were mixed with radiolabeled ^{f}CUP9^{h} (lane 8) in ATP-supplemented buffer, reactions in lanes 1-5 lacked the indicated components. Reaction mixtures were incubated at 30°C for 10 min, fractionated by SDS-10% PAGE, followed by autoradiography. **b**, The multi-Ub chain formed on ^{f}CUP9^{h} was linked mainly through Lys48 of Ub. Reactions contained UBA1^{h}, RAD6, ^{f}UBR1, ^{f}CUP9^{h} and Ub (lanes 1-3) or K48RUb (lanes 4-5) or methylated Ub (lanes 6-7) in ATP-supplemented buffer. Reaction mixtures were incubated at 30°C for the designated times then fractionated by SDS-8% PAGE. The positions of radiolabeled ^{f}CUP9^{h} were denoted.

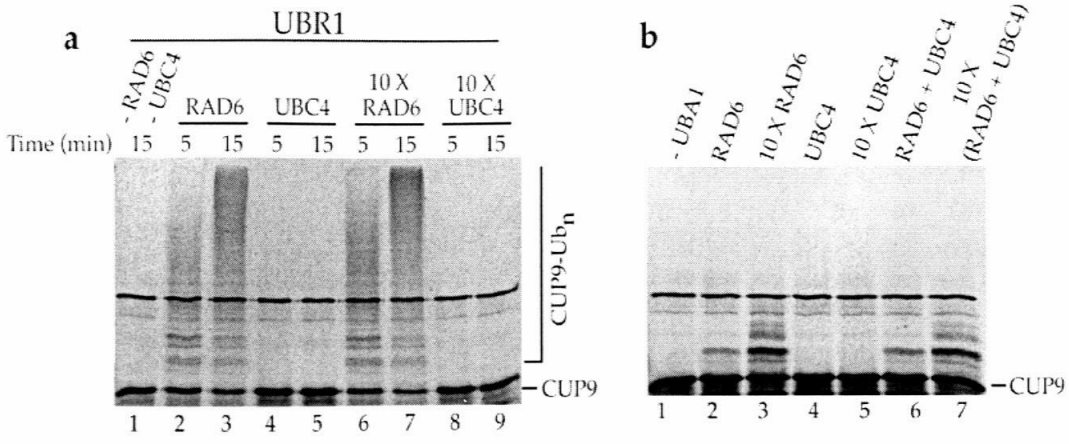


Figure 4 **a**, UBC4 could not substitute RAD6 in supporting the ^{f}UBR1-dependent multi-ubiquitylation of ^{f}CUP9^{h}. Purified UBA1^{h}, ^{f}UBR1, Ub and radiolabeled ^{f}CUP9^{h} were mixed with RAD6 or UBC4 in ATP-supplemented buffer. The reaction in lane 1 contained neither RAD6 nor UBC4. The reactions in lanes 6-7 contained nine fold higher concentration of RAD6 than that in lanes 2-3, while the reactions in lane 8-9 contained nine fold higher concentration of UBC4 than that in lanes 4-5. Reactions were allowed to proceed at 30°C for 20 min. **b**, UBC4 and RAD6 could not support multiubiquitylation of ^{f}CUP9^{h} in the absence of ^{f}UBR1. Purified UBA1^{h}, Ub and radiolabeled ^{f}CUP9^{h} were mixed with either RAD6, or UBC4, or their combinations, in ATP-supplemented buffer. The reaction in lane 1 lacked UBA1^{h}, the reactions in lanes 2-7 contained different concentrations of RAD6, or UBC4, or their combinations, as denoted. Reactions were allowed to proceed at 30°C for 20 min.

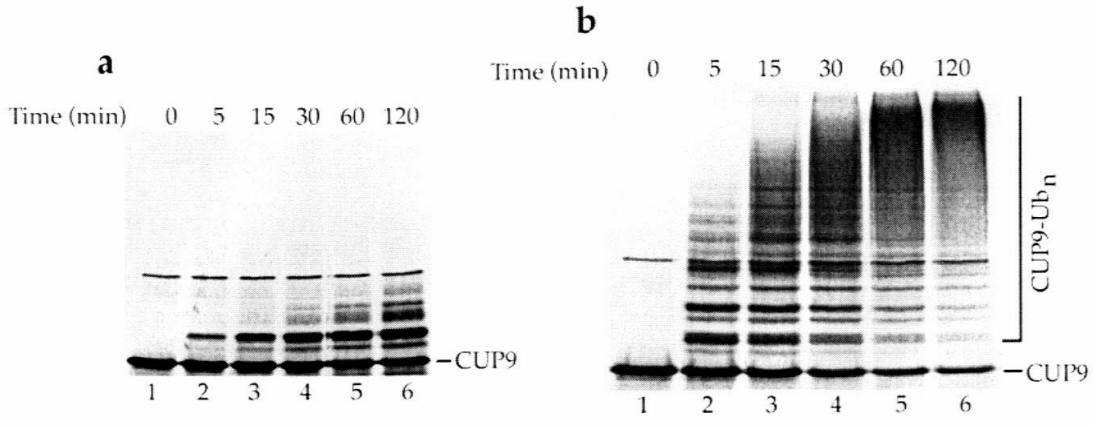


Figure 5 **a**, Time course of ^{f}UBR1-independent, RAD6-catalyzed ubiquitylation of ^{f}CUP9^{h}. Purified UBA1^{h}, RAD6, Ub and radiolabeled ^{f}CUP9^{h} were mixed in ATP-supplemented buffer, reactions were allowed to proceed at 30°C for the designated times. **b,** Time course of ^{f}UBR1-mediated ^{f}CUP9^{h} *in vitro* ubiquitylation. Same as in **a** except that the reactions contained ^{f}UBR1.

Chapter 3

Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway

This work has been published (Turner et al., Nature 405: 579-583, 2000).

Glenn C. Turner*, Fangyong Du*, and Alexander Varshavsky

*These authors contributed equally to this work.

Abstract

Protein degradation by the ubiquitin system controls the intracellular concentrations of many regulatory proteins. A protein substrate of the ubiquitin system is conjugated to ubiquitin through the action of three enzymes, E1, E2 and E3, with the degradation signal (degron) of the substrate recognized by E3[1-3]. The resulting multi-ubiquitylated substrate is degraded by the 26S proteasome[4]. Here we describe the physiological regulation of a ubiquitin-dependent pathway through allosteric modulation of its E3 activity by small compounds. Ubr1, the E3 enzyme of the N-end rule pathway (a ubiquitin-dependent proteolytic system) in *Saccharomyces cerevisiae*, mediates the degradation of Cup9, a transcriptional repressor of the peptide transporter Ptr2[5]. Ubr1 also targets proteins that have destabilizing amino-terminal residues[6]. We show that the degradation of Cup9 is allosterically activated by dipeptides with destabilizing N-terminal residues. In the resulting positive feedback circuit, imported dipeptides bind to Ubr1 and accelerate the Ubr1-dependent degradation of Cup9, thereby de-repressing the expression of Ptr2 and increasing the cell's capacity to import peptides. These findings identify the physiological rationale for the targeting of Cup9 by Ubr1, and indicate that small compounds may regulate other ubiquitin-dependent pathways.

Introduction, results and discussion

The rate of degradation of specific proteins is often regulated by modulating the exposure or the structure of their degrons. For example, the degrons of the cyclin-dependent kinase inhibitors Sic1 and p27 are activated by

phosphorylation, which is timed to bring about their destruction at key transition points in the cell cycle[7]. In other cases, phosphorylation regulates the activity of an E3 itself. For example, the anaphase-promoting complex, a multisubunit E3, is activated only at mitosis[8]. One ubiquitin-dependent proteolytic system, called the N-end rule pathway, targets proteins carrying a degradation signal called the N-degron[9], which comprises a destabilizing N-terminal residue and a lysine residue[10,11]. In *S. cerevisiae*, there are two classes of destabilizing residues, basic, or 'type 1' (Arg, Lys and His) and bulky hydrophobic, or 'type 2' (Phe, Leu, Tyr, Trp and Ile). Ubr1, a RING-H2 finger-containing E3 of relative molecular mass 225,000 (M_r 225K), directly recognizes these N-terminal residues[6,12]. A complex of Ubr1 and the E2 enzyme Rad6 (Ubc2), synthesizes a multi-ubiquitin chain linked to a lysine residue of the substrate[13]. Dipeptides with a destabilizing N-terminal residue of either basic or hydrophobic type act as competitive inhibitors of the degradation of N-end rule substrates carrying the same type of destabilizing residue[14-16]. Thus, Ubr1 contains two distinct N-terminal residue-binding sites that are each capable of binding either a dipeptide or a protein, but not both at the same time.

The N-end rule pathway was discovered through the use of engineered reporter proteins[9]. The first physiological function of Ubr1 in *S. cerevisiae* has recently been shown to be the regulation of peptide uptake[17], through control of degradation of the M_r 35K homeodomain protein Cup9, a transcriptional repressor of the di- and tripeptide transporter Ptr2[5]. Ubr1 targets Cup9 through a degron located in the carboxy-terminal half of Cup9 (F. Navarro-Garcia, G.C.T. and A. V. unpublished data). Despite this unexpected mode of recognition, we

asked whether dipeptides bearing destabilizing N-terminal residues could affect the Ubr1-mediated degradation of Cup9, as dipeptides are able to inhibit degradation of canonical N-end rule substrates[16].

We addressed this question using $Cup9_{NSF}$, a C-terminally Flag-tagged variant of the Cup9 repressor bearing an Asn-215 → Ser substitution. $Cup9_{NSF}$ was degraded indistinguishably from wild-type Cup9 but was predicted to have a much lower affinity for DNA[18,19], so it would not influence the expression of Ptr2 and the uptake of dipeptides. $Cup9_{NSF}$ was expressed as part of a fusion of the form Flag-DHFR-ubiquitin-$Cup9_{NSF}$ (where Flag-DHFR is N-terminally Flag-tgged mouse dihydrofolate reductase; see Methods; Fig. 1). Ubiquitin-specific proteases co-translationally cleave this UPR (ubiquitin-protein-reference) fusion at the ubiquitin-Cup9 junction, yielding the test protein $Cup9_{NSF}$ and the long-lived Flag-DHFR-ubiquitin reference protein, which serves as an internal control for variations in expression levels and immunoprecipitation efficiency[20,21].

Cells expressing $Cup9_{NSF}$ were grown in minimal medium containing allantoin as the nitrogen source to avoid the known effects of nitrogen catabolite repression on *PTR2* expression[22]. Leu-Ala and Arg-Ala, dipeptides bearing either type of destabilizing N-terminal residue (Leu, bulky hydrophobic; Arg, basic), were added to a final concentration of 10 mM (see Methods). This dipeptide concentration results in maximal inhibition of degradation of N-end rule substrates[16]. Notably, the addition of either Leu-Ala or Arg-Ala exerted an *opposite* effect on $Cup9_{NSF}$, strongly accelerating its degradation in wild-type

(*UBR1*) cells. The half-life of $Cup9_{NSF}$ decreased from ~5 min in the absence of dipeptides (Fig. 1c) to less than 1 min in their presence (Fig. 1b). This stimulatory effect was not observed in a *ubr1Δ* strain, indicating that the augmented degradation of $Cup9_{NSF}$ was dependent on Ubr1. The enhancement of degradation required dipeptides with destabilizing N-terminal residues: dipeptides of the same composition but bearing a stabilizing residue (Ala-Leu and Ala-Arg) did not affect the degradation of $Cup9_{NSF}$ ($t_{1/2}$ ~5 min) (Fig. 1b, and data not shown). Similar results were obtained with cells expressing $Cup9_{NSF}$ that was not a part of a UPR fusion (data not shown).

To determine the concentration dependence of the stimulation, we measured the degradation of Cup9 at a range of concentrations of Trp-Ala. The enhancement of $Cup9_{NSF}$ degradation was detectable at 1 μM Trp-Ala, the lowest concentration tested ($t_{1/2}$ ~1 min) (Fig. 1c). In contrast, the degradation of $Cup9_{NSF}$ was not altered either by Ala-Trp or by the constituent amino acids Trp and Ala (Fig. 1c). Similar results were obtained using Leu-Ala and Arg-Ala (data not shown). These results indicate that the relevant signaling molecule in this process is a dipeptide bearing a destabilizing N-terminal residue. The range of dipeptide concentrations that significantly stimulated Cup9 degradation was similar to physiologically active levels of many other nutrients.

Cup9 represses transcription of the transporter-encoding *PTR2* gene[5]. Thus, the dipeptide-induced acceleration of Cup9 degradation would be expected to increase the levels of *PTR2* messenger RNA, ultimately leading to an increase in dipeptide uptake. We tested this by examining the levels of *PTR2*

mRNA in the presence or absence of dipeptides in the medium. At 25 μM, both Trp-Ala and Arg-Ala induced *PTR2* expression in the wildtype (*UBR1*) strain (Fig. 2a). Both Ubr1 and Cup9 were required for these effects, as the expression of *PTR2* was not altered by dipeptides in *ubr1Δ* and *cup9Δ* strains. Induction of *PTR2* mRNA could be observed at 1 μM Trp-Ala, and increased at higher dipeptide concentrations (Fig. 2b), in agreement with the observed changes in the half-life of Cup9 (Fig. 1c).

A plausible mechanism of the enhancement effect is that a dipeptide interacts with either the basic or hydrophobic N-terminal residue-binding sites of Ubr1, and a distinct (third) substrate-binding site of Ubr1 recognizes the internal degron of Cup9. In this model, the interaction of Ubr1 with dipeptides allosterically increases the ability of the Ubr1-Rad6 (E3-E2) complex to ubiquitylate Cup9. To test whether dipeptides act directly through Ubr1, we examined the effect of dipeptides on Cup9 ubiquitylation in an *in vitro* system consisting of the following purified components: Ubr1 (E3), Rad6 (E2), Uba1 (E1), ubiquitin, ATP, and radiolabelled Cup9. Cup9 was significantly multi-ubiquitylated, in a Ubr1/Rad6-dependent reaction (data not shown), in the absence of added dipeptides (Fig. 3). This result was consistent with the relatively rapid *in vivo* degradation of Cup9 ($t_{1/2}$ ~5 min) in the absence of dipeptides (Fig. 1c).

The addition of dipeptides bearing either type of destabilizing N-terminal residue to the *in vitro* system substantially stimulated the Ubr1-dependent multi-ubiquitylation of Cup9 (Fig. 3). Dipeptides of the same composition but bearing a stabilizing N-terminal residue did not stimulate multi-ubiquitylation,

nor did the amino acid components of these dipeptides (Fig. 3). These results demonstrate that dipeptides act directly through Ubr1, without an intermediate signalling pathway. The underlying allosteric mechanism may involve increased affinity of Ubr1 for Cup9, or enhanced ubiquitylation activity of the Ubr1-Rad6 complex towards Cup9, or both.

This work shows that the two binding sites of Ubr1 that interact with destabilizing N-terminal residues can act as allosteric sites that enable Ubr1 to sense the presence of imported dipeptides, and to accelerate degradation of the Cup9 repressor, resulting in an appropriate induction of the Ptr2 transporter (Fig. 4a-d). This model predicts that a dipeptide bearing a destabilizing N-terminal residue, for example Leu-Ala, should stimulate its own uptake, in contrast to Ala-Leu. This prediction was borne out when we tested the ability of these two leucine-containing dipeptides to support the growth of *S. cerevisiae* auxotrophic for leucine (Fig. 4e).

Food sources that *S. cerevisiae* encounters outside the laboratory setting are likely to contain mixtures of short peptides, a subset of which would be capable of activating the Ubr1-based positive feedback circuit. We modeled this situation by providing cells with a mixture of Leu-Ala at a low concentration (2 μM) and Ala-Leu at a high concentration (230 μM). Although neither dipeptide supplement alone could satisfy the strain's requirement for leucine, a mixture of the two dipeptides supported robust growth (Fig. 4e). Moreover, a mixture of Ala-Leu (230 μM) and a dipeptide *lacking* leucine but bearing a destabilizing N-terminal residue (10 μM Arg-Ala or 1 μM Lys-Ala) also rescued growth. In contrast, dipeptides bearing stabilizing N-terminal residues, or the amino-acid

components of these dipeptides, could not rescue the growth of Leu$^-$ cells in the presence of 230 μM Ala-Leu (Fig. 4e).

This work establishes for the first time that the activity of an E3 can be directly linked to the presence of an environmental signal through an allosteric interaction with a small compound. Specifically, dipeptides bearing destabilizing N-terminal residues are shown to act as allosteric activators of Ubr1, enhancing its ability to support the ubiquitylation and degradation of Cup9. Physiologically, this results in a positive feedback circuit governing the uptake of peptides (Fig. 4). Through their binding to Ubr1, the imported dipeptides accelerate degradation of Cup9, thereby derepressing the synthesis of the Ptr2 transporter and enhancing the cell's ability to import di- and tripeptides. As most cells have the capacity to import peptides, the results of this work suggest that peptide import may be regulated similarly by Ubr1 homologues in metazoans[23] and by the ClpAP-dependent N-end rule pathway in *Escherichia coli*[24]. The ubiquitin system is either known or suspected to be important in the control of intermediary metabolism and the transport of small molecules across membranes[2,3]. Our findings suggest that these compounds, or their enzymatically produced derivatives, may modulate the functions of E3s in the ubiquitin system similarly to the effects observed here with dipeptides and Ubr1.

Methods

Yeast strains and plasmids. The *S. cerevisiae* strains used in pulse-chase experiments, JD52 (*MAT***a** *lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*), JD55 (*ubr1Δ::HIS3*), have been described[25]. Flag-DHFR-ubiquitin-Cup9$_{NSF}$ was expressed from the P_{MET25} promoter on the centromeric vector p416MET25[26]. Construction details are available upon request. Strains used for Northern analyses were AVY30 (*MATα leu2-3,112 ubr1Δ::LEU2*), AVY 31 (*MATα leu2-3,112 cup9Δ::LEU2*), and AVY32 (*MATα LEU2*), constructed in the RJD350 background (*MATα leu2-3,112*; a gift from R. Deshaies) using restriction fragments obtained from plasmids pSOB30[6], pCB119[5] and pJJ252[27], respectively. The RJD350 strain was used for the colony formation assay (Fig. 4e). SHM plates[5] were supplemented with dipeptides or amino acids at the following concentrations: 230 μM Leu-Ala (or 2 μM where indicated), 230 μM Ala-Leu, 230 μM each of Leu and Ala, 10 μM Arg-Ala, 10 μM Ala-Arg, 1 μM Lys-Ala, 1 μM Ala-Lys, 10 μM each of Arg and Ala, or 1 μM each of Lys and Ala.

Pulse-chase analysis. Cells were cultured in SHM[5] with auxotrophic supplements. Dipeptides were added to cultures at an absorbance of~0.6 at 600 nm and incubation continued for 2.5 h in the experiments in Fig. 1b, and for 30 min for Fig. 1c. Cells were harvested, washed in 0.8 ml of SHM, resuspended in 0.4 ml of SHM, and labeled for 5 min at 30°C with 0.16 mCi of ^{35}S-EXPRESS (New England Nuclear). Cells were pelleted and resuspended in fresh SHM containing 4 mM L-methionine and 2 mM L-cysteine. Samples (0.1 ml) were taken at the time points indicated and transferred to chilled tubes, each containing 0.5 ml of 0.5-mm glass beads, 0.7 ml of ice-cold lysis buffer (1%

Triton-X100, 0.15 M NaCl, 5 mM EDTA, 50 mM sodium HEPES, pH 7.5), and a mixture of protease inhibitors (final concentrations 1 mM phenylmethylsulfonyl fluoride, 2 μg/ml aprotinin, 0.5 μg/ml leupeptin, and 0.7 μg/ml pepstatin). Extracts were prepared and immuno-precipitations carried out as described[25], using anti-Flag M2 resin (Sigma). Immunoprecipitates were fractionated by 13% SDS-PAGE, and detected by autoradiography.

RNA preparation and northern analysis. Cells were cultured in SHM to an absorbance of~0.6 at 600 nm. The indicated dipeptides were then added to a 25 μM final concentration, and the incubation was continued for an additional 30 min. Total RNA was prepared[28], and 25 μg samples were electrophoresed in 1% formaldehyde-agarose gels, followed by blotting for northern analysis[29].

Ubr1-dependent *in vitro* ubiquitylation system. The components of this system were purified as follows. N-terminally hexahistidine-tagged Uba1 was overexpressed in *S. cerevisiae* and purified by fractionation over Ni-NTA, ubiquitin affinity and Superdex-200 columns. Rad6 was overexpressed in *E. coli*, and purified by fractionation over DEAE, Mono-Q and Superdex-75 columns. N-terminally Flag-tagged Ubr1 was overexpressed in *S. cerevisiae*, and purified by fractionation over anti-Flag M2, Rad6 affinity, and Superdex-200 columns. N-terminally Flag-tagged, C-terminally hexahistidine-tagged Cup9 was expressed and radiolabelled in *E. coli*, and purified by fractionation over Ni-NTA and anti-Flag M2 columns. Details of these expression and purification protocols are available upon request.

The *in vitro* ubiquitylation reactions contained the following components: 7 μM ubiquitin, 50 nM Uba1, 50 nM Rad6, 50 nM Ubr1, 550 nM ^{35}S-labelled

Cup9, 25 mM HEPES/KOH (pH 7.5), 25 mM KCl, 5 mM $MgCl_2$, 2 mM ATP, 0.1 mM dithiothreitol and 0.5 mg ml^{-1} ovalbumin as a carrier protein. Dipeptides or amino acids, as indicated, were added to a final concentration of 2 μM (top panel, Fig. 3) or 10 μM (bottom panel, Fig. 3). All components except Uba1 were mixed on ice for 10 min; Uba1 was then added and reactions shifted to 30°C. After the indicated times, the reactions were terminated by adding an equal volume of 2 x SDS-PAGE loading buffer and heating at 95°C for 5 min, followed by 8% SDS-PAGE.

Acknowledgments

We thank A. Webster for her valuable contributions to establishing of the *in vitro* system, and R. Deshaies, J. Dohmen, L. Prakash, H. Rao and J. Sheng for their gifts of plasmids and strains. We also thank C. Byrd, H. Rao, T. Iverson and especially R. Deshaies for helpful discussions and comments on the manuscript. This work was supported by grants to A. V. from the NIH. Correspondence and requests for materials should be addressed to A. V. (e-mail: avarsh@caltech.edu).

References

1. Laney, J. D. and Hochstrasser, M. Substrate targeting in the ubiquitin system. *Cell* **97**, 427-430 (1999).
2. Hershko, A. and Ciechanover, A. The ubiquitin system. *Annu. Rev. Biochem.* **76**, 425-479 (1998).
3. Varshavsky, A. The ubiquitin system. *Trends Biochem. Sci.* **22**, 383-387 (1997).
4. Baumeister, W., Walz, J., Zühl, F. and Seemüller, E. The proteasome: paradigm of a self-compartmentalizing protease. *Cell* **92**, 367-380 (1998).
5. Byrd, C., Turner, G. C. and Varshavsky, A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. *EMBO J.* **17**, 269-277 (1998).
6. Bartel, B., Wünning, I. and Varshavsky, A. The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179-3189 (1990).
7. King, R. W., Deshaies, R. J., Peters, J. M. and Kirschner, M. W. How proteolysis drives the cell cycle. *Science* **274**, 1652-1659 (1996).
8. Kotani, S., Tanaka, H., Yasuda, H. and Todokoro, K. Regulation of APC activity by phosphorylation and regulatory factors. *J. Cell Biol.* **146**, 791-800 (1999).
9. Bachmair, A., Finley, D. and Varshavsky, A. In vivo half-life of a protein is a function of its amino-terminal residue. *Science* **234**, 179-186 (1986).
10. Bachmair, A. and Varshavsky, A. The degradation signal in a short-lived protein. *Cell* **56**, 1019-1032 (1989).
11. Varshavsky, A. The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142-12149 (1996).

12. Xie, Y. M. and Varshavsky, A. The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain. *EMBO J.* **18**, 6832-6844 (1999).

13. Dohmen, R. J., Madura, K., Bartel, B. and Varshavsky, A. The N-end rule is mediated by the UBC2(RAD6) ubiquitin-conjugating enzyme. *Proc. Natl. Acad. Sci. USA* **88**, 7351-7355 (1991).

14. Reiss, Y., Kaim, D. and Hershko, A. Specificity of binding of N-terminal residues of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.* **263**, 2693-2269 (1988).

15. Gonda, D. K. *et al.* Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700-16712 (1989).

16. Baker, R. T. and Varshavsky, A. Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **87**, 2374-2378 (1991).

17. Alagramam, K., Naider, F. and Becker, J. M. A recognition component of the ubiquitin system is required for peptide transport in Saccharomyces cerevisiae. *Mol. Microbiol.* **15**, 225-234 (1995).

18. Knight, S. A. B., Tamai, K. T., Kosman, D. J. and Thiele, D. J. Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* **14**, 7792-7804 (1994).

19. Wolberger, C., Vershon, A. K., Liu, B. S., Johnson, A. D. and Pabo, C. O. Crystal-Structure of a Matα-2 Homeodomain-Operator Complex Suggests a General-Model For Homeodomain-DNA Interactions. *Cell* **67**, 517-528 (1991).

20. Lévy, F., Johnsson, N., Rumenapf, T. and Varshavsky, A. Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907-4912 (1996).

21. Suzuki, T. and Varshavsky, A. Degradation signals in the lysine-asparagine sequence space. *EMBO J.* **18**, 6017-6026 (1999).

22. Barnes, D., Lai, W., Breslav, M., Naider, F. and Becker, J. M. PTR3, a novel gene mediating amino acid-inducible regulation of peptide transport in Saccharomyces cerevisiae. *Mol. Microbiol.* **29**, 297-310 (1998).

23. Kwon, Y. T. *et al.* The mouse and human genes encoding the recognition component of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **95**, 7898-7903 (1998).

24. Tobias, J. W., Shrader, T. E., Rocap, G. and Varshavsky, A. The N-end rule in bacteria. *Science* **254**, 1374-1377 (1991).

25. Ghislain, M., Dohmen, R. J., Levy, F. and Varshavsky, A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in Saccharomyces cerevisiae. *EMBO J.* **15**, 4884-4899 (1996).

26. Mumberg, D., Muller, R. and Funk, M. Regulatable promoters of Saccharomyces cerevisiae - comparison of transcriptional activity and their use for heterologous expression. *Nucl. Acids Res.* **22**, 5767-5768 (1994).

27. Jones, J. S. and Prakash, L. Yeast Saccharomyces cerevisiae selectable markers in pUC18 polylinkers. *Yeast* **6**, 363-366 (1990).

28. Schmitt, M. E., Brown, T. A. and Trumpower, B. L. A rapid and simple method for preparation of RNA from Saccharomyces cerevisiae. *Nucl. Acids Res.* **18**, 3091-3092 (1990).

29. Ausubel, F. M. *et al.* (eds.) *Current Protocols in Molecular Biology.* (Wiley-Interscience, New York, 1996).

Figure 1 Enhancement of Cup9p degradation by dipeptides with destabilizing N-terminal residues. **a**, The fusion protein used for pulse-chase analysis. The stable Flag-DHFR-ubiquitin reference portion of the fusion is co-translationally cleaved from $Cup9_{NSF}$ by ubiquitin-specific proteases (UBPs). **b**, Pulse-chase analysis of Flag-DHFR-ubiquitin-$Cup9_{NSF}$ in the presence of various dipeptides at a concentration of 10 mM. Dipeptides bearing either basic (Arg-Ala) or bulky hydrophobic (Leu-Ala) destabilizing N-terminal residues strongly enhance $Cup9_{NSF}$ degradation, but only in strains expressing Ubr1. Dipeptides bearing a stabilizing N-terminal residue (Ala-Arg and Ala-Leu) do not alter $Cup9_{NSF}$ degradation. **c**, Effects of different concentrations of Trp-Ala on the enhancement of $Cup9_{NSF}$ degradation in wild-type (*UBR1*) cells. Dash indicates that pulse-chase analysis was performed in the absence of added dipeptides. Enhancement of $Cup9_{NSF}$ degradation was detectable at 1 μM Trp-Ala. Ub^{R48}, ubiquitin containing mutation at Lys48 to Arg (ref. 20).

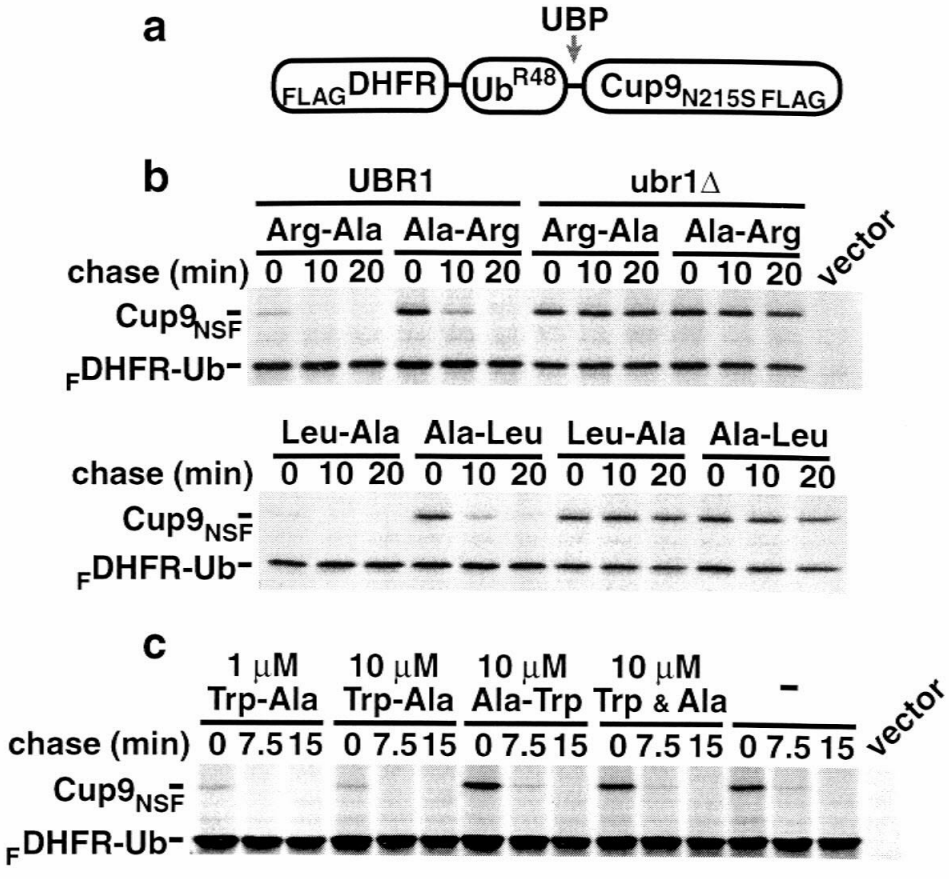

a
UBP
FLAG DHFR
Ub R48
Cup9 N215S FLAG
b
UBR1
ubr1Δ
Arg-Ala
Ala-Arg
Arg-Ala
Ala-Arg
chase (min) 0 10 20 0 10 20 0 10 20 0 10 20
vector
Cup9 NSF
FDHFR-Ub
Leu-Ala
Ala-Leu
Leu-Ala
Ala-Leu
chase (min) 0 10 20 0 10 20 0 10 20 0 10 20
Cup9 NSF
FDHFR-Ub
c
1 μM Trp-Ala
10 μM Trp-Ala
10 μM Ala-Trp
10 μM Trp & Ala
-
chase (min) 0 7.5 15 0 7.5 15 0 7.5 15 0 7.5 15 0 7.5 15
vector
Cup9 NSF
FDHFR-Ub

Figure 2 Effects of dipeptides on expression of the dipeptide transporter gene *PTR2*. **a**, Induction of *PTR2* expression by dipeptides bearing destabilizing N-terminal residues (Trp-Ala and Arg-Ala) required both *UBR1* and *CUP9*. Dipeptides bearing a stabilizing N-terminal residue (Ala-Trp and Ala-Arg) had no effect on *PTR2* expression. *PTR2* mRNA and the *ACT1* mRNA loading control are indicated. **b**, Effect of different concentrations of Trp-Ala on the levels of *PTR2* mRNA.

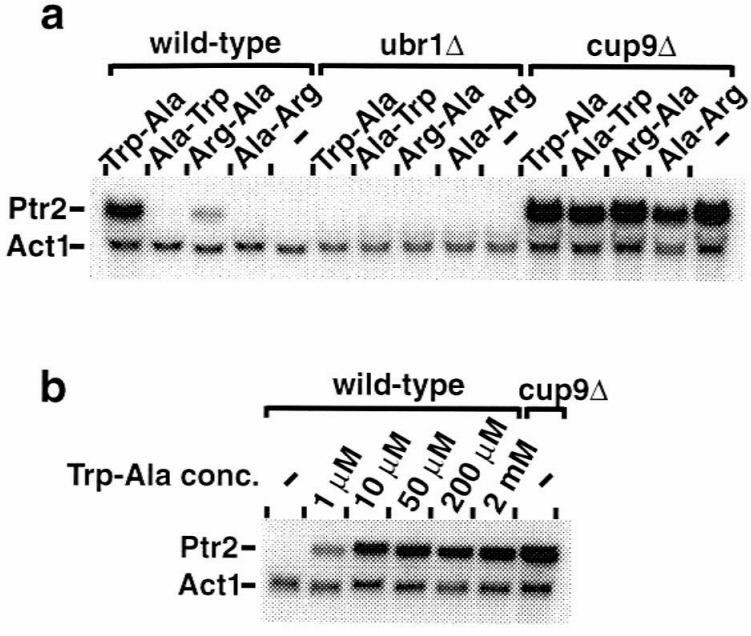

a
wild-type
ubr1Δ
cup9Δ
Trp-Ala
Ala-Trp
Arg-Ala
Ala-Arg
–
Ptr2–
Act1–
b
wild-type
cup9Δ
Trp-Ala conc.
–
1 μM
10 μM
50 μM
200 μM
2 mM
–
Ptr2–
Act1–

Figure 3 *In vitro* ubiquitylation of Cup9 is enhanced by dipeptides bearing destabilizing N-terminal residues. Reactions consisting of purified Uba1 (E1), Rad6 (E2), Ubr1 (E3), ubiquitin, ATP and radiolabelled Cup9 were supplemented with the indicated dipeptides or amino acids (top panel, 2 μM, bottom panel, 10 μM), or left unsupplemented (—), and allowed to proceed at 30°C for the designated times. Radiolabelled input Cup9, and its multi-ubiquitylated derivatives, are indicated.

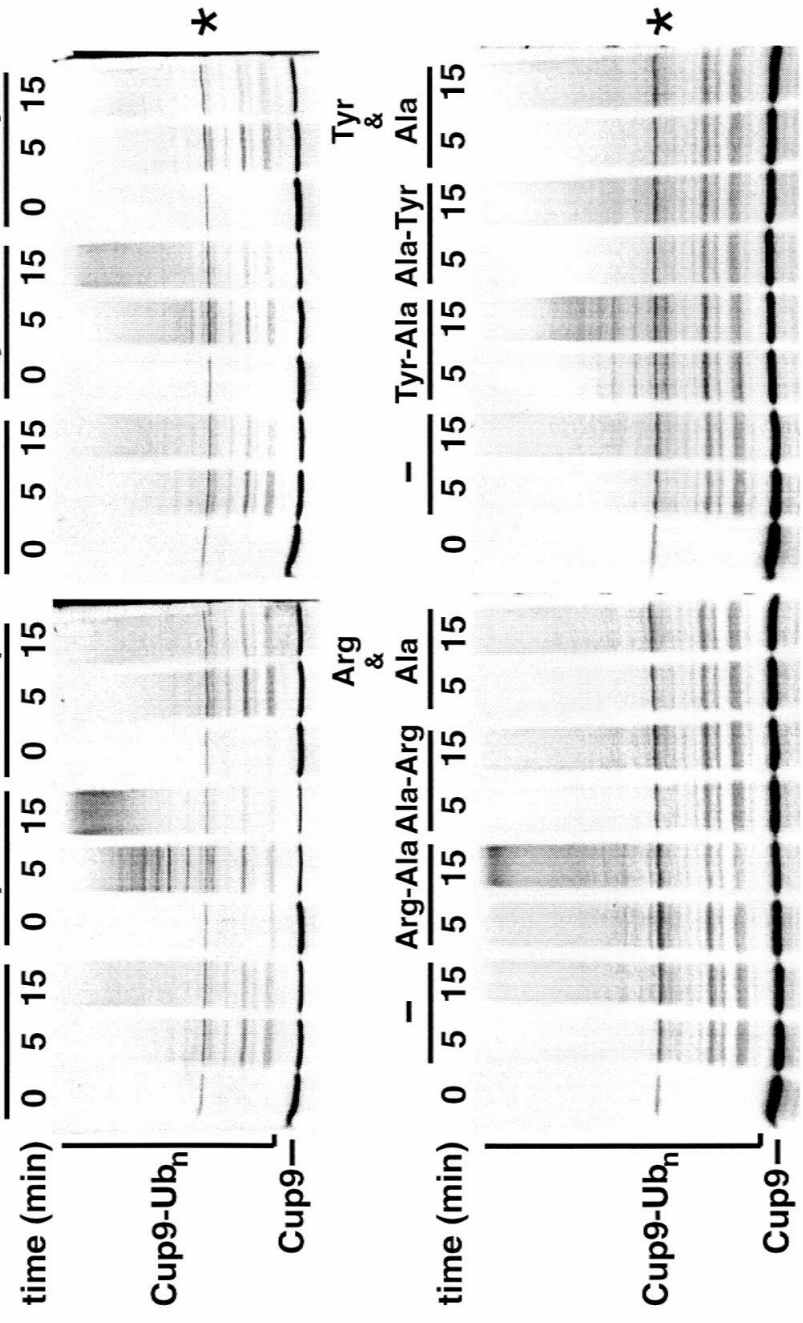
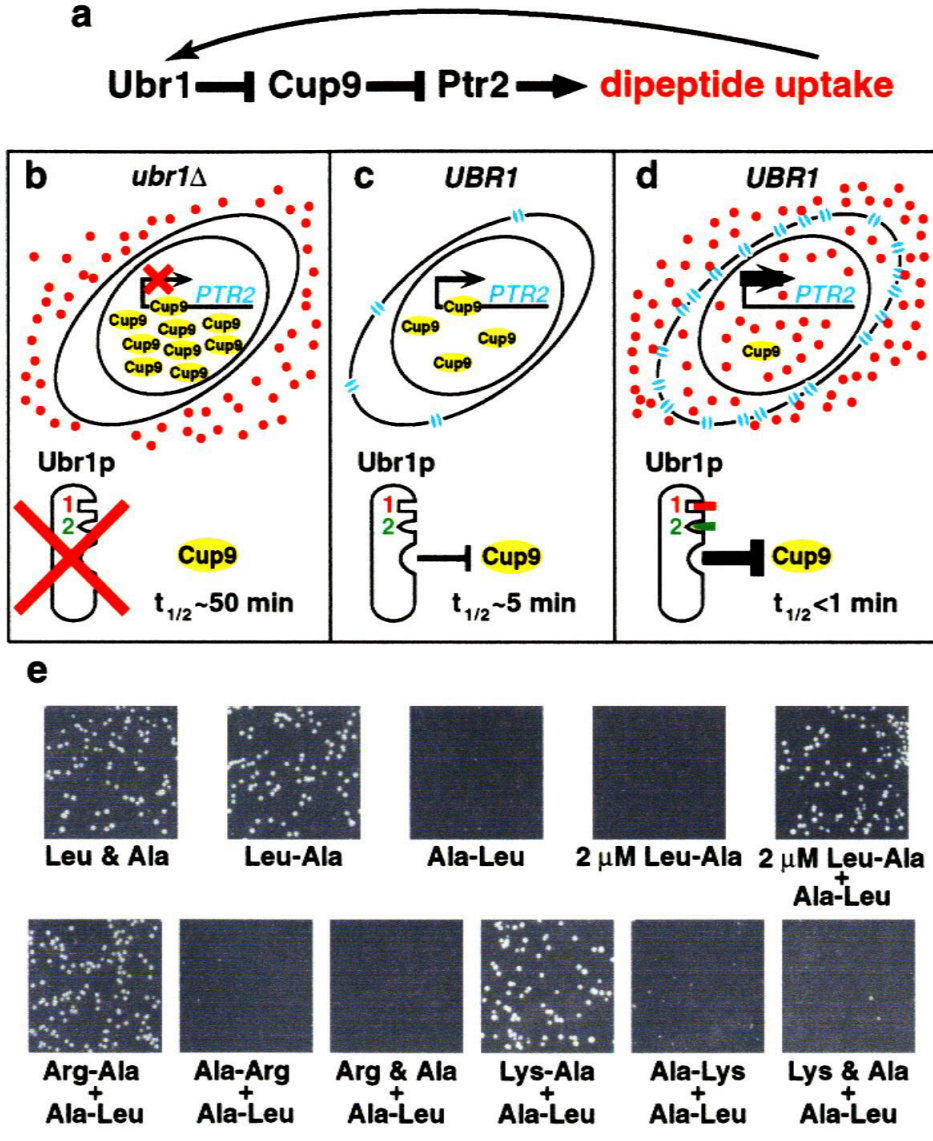

time (min)
–
0 5 15
Trp-Ala
0 5 15
Ala-Trp
0 5 15
Arg-Ala
5 15
Ala-Arg
5 15
Arg & Ala
5 15
Lys-Ala
0 5 15
Ala-Lys
0 5 15
Tyr-Ala
5 15
Ala-Tyr
5 15
Tyr & Ala
5 15
Cup9-Ub$_n$
Cup9
*

Figure 4 Feedback regulation of peptide import in *S. cerevisiae*. **a**, The peptide transport circuit. **b**, Ubr1 is required for dipeptide uptake. In the absence of Ubr1 (*ubr1Δ*), the transcriptional repressor Cup9 is long-lived, accumulates to high levels and extinguishes the expression of the *PTR2* gene. Thus, *ubr1Δ* cells cannot import dipeptides (red dots). **c,** In a wild-type (*UBR1*) cell growing in the absence of extracellular dipeptides, Ubr1 targets Cup9 for degradation ($t_{1/2}$ ~5 min), resulting in a moderate concentration of Cup9 and weak but significant expression of the Ptr2 transporter (blue double ovals). **d,** In wild-type (*UBR1*) cells growing in the presence of extracellular dipeptides some of which bear destabilizing N-terminal residues, the imported dipeptides bind to either the basic (type 1, red) or the hydrophobic (type 2, green) residue-binding site of Ubr1. Binding of either type of dipeptide to Ubr1 allosterically increases the rate of Ubr1-mediated degradation of Cup9. Peptides of both types are shown bound to Ubr1p, but the binding of either peptide accelerates Cup9 degradation. The resulting decrease of the half-life of Cup9 from ~5 min to < 1 min results in a very low concentration of Cup9, and consequently to strong induction of the Ptr2 transporter. **e,** Colony formation assays. A *S. cerevisiae* strain requiring leucine for growth was incubated on plates supplemented with either dipeptides or their amino-acid constituents at the following concentrations: Leu-Ala, 230 μM (2 μM where indicated); Ala-Leu, 230 μM; Arg-Ala, 10 μM; Ala-Arg, 10 μM; Lys-Ala, 1 μM; Ala-Lys, 1 μM; Arg and Ala, 10 μM each, Lys and Ala, 1 μM each.


a
Ubr1
Cup9
Ptr2
dipeptide uptake
b
ubr1Δ
PTR2
Cup9
Ubr1p
1
2
Cup9
t1/2~50 min
c
UBR1
PTR2
Cup9
Ubr1p
1
2
Cup9
t1/2~5 min
d
UBR1
PTR2
Cup9
Ubr1p
1
2
Cup9
t1/2<1 min
e
Leu & Ala
Leu-Ala
Ala-Leu
2 μM Leu-Ala
2 μM Leu-Ala + Ala-Leu
Arg-Ala + Ala-Leu
Ala-Arg + Ala-Leu
Arg & Ala + Ala-Leu
Lys-Ala + Ala-Leu
Ala-Lys + Ala-Leu
Lys & Ala + Ala-Leu

Chapter 4

Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain

(prepared for submission)

Fangyong Du, Federico Navarro-Garcia and Alexander Varshavsky

Abstract

Protein degradation by the ubiquitin (Ub) system controls the concentrations of many regulatory proteins. The degradation signals of these proteins are recognized by Ub ligases. Two binding sites of UBR1, the Ub ligase of the N-end rule pathway in *Saccharomyces cerevisiae*, recognize basic (type 1) and bulky hydrophobic (type 2) N-terminal residues of proteins or short peptides. A third substrate-binding site of UBR1 targets CUP9, a transcriptional repressor of the peptide transporter PTR2. Previous work demonstrated that dipeptides with destabilizing N-terminal residues allosterically accelerate the UBR1-dependent degradation of CUP9, an effect that mediates positive feedback in the import of peptides. Here we show that these dipeptides cause dissociation of the C-terminal autoinhibitory domain of UBR1 from its N-terminal region that contains all three substrate-binding sites. Moreover, this dissociation, which allows the interaction between UBR1 and CUP9, is strongly increased only if *both* type 1 and type 2 binding sites of UBR1 are occupied by dipeptides. The discovery of autoinhibition in UBR1 indicates that this regulatory mechanism may also control the activity of other Ub-dependent pathways.

Introduction, results and discussion

One Ub-dependent proteolytic system, termed the N-end rule pathway, targets proteins carrying a degradation signal called the N-degron (*9*), which comprises a substrate's destabilizing N-terminal residue and a lysine residue (*10-13*). In *S. cerevisiae*, the 225 kD, RING domain-containing E3 called UBR1 directly recognizes two classes of destabilizing N-terminal residues: basic, or type 1 (Arg,

Lys, and His) and bulky hydrophobic, or type 2 (Phe, Leu, Tyr, Trp, and Ile) (*12*). A complex of UBR1 and the Ub-conjugating (E2) enzyme RAD6 (UBC2) produces a multi-Ub chain linked to a substrate's Lys residue (*14*). Dipeptides bearing either basic (type 1) or bulky hydrophobic (type 2) N-terminal residues inhibit degradation of N-end rule substrates carrying the same type of destabilizing residue by competing with substrates for binding to the type 1 or type 2 sites of UBR1 (*15-17*). One physiological substrate of the N-end rule pathway that is targeted through an N-degron is the 33 kD fragment of SCC1, a subunit of cohesin complexes that hold together sister chromatids (*18, 19*). The SCC1 fragment is produced through a cleavage by the ESP1 separase, and must be degraded by the N-end rule pathway to maintain the normal fidelity of chromosome segregation (*18*).

Another physiological function of *S. cerevisiae* UBR1 is the control of peptide import, through regulated degradation of the 35 kD homeodomain protein CUP9, a transcriptional repressor of the di- and tripeptide transporter PTR2 (*17, 20, 21*). A third substrate-binding site of UBR1 targets CUP9 through an internal (non-N-terminal) degron. Previous work (*17*) has shown that the degradation of CUP9 is allosterically activated by dipeptides with destabilizing N-terminal residues. In the resulting positive feedback circuit, imported dipeptides bind to UBR1 and accelerate the UBR1-dependent degradation of CUP9, thereby derepressing the expression of the PTR2 transporter and increasing the cell's capacity to import peptides (*17*).

In the present work, we addressed the mechanism of dipeptide-mediated activation of UBR1 by asking, first, whether dipeptides enhance the binding of UBR1 to CUP9. These experiments employed the GST (glutathione S-transferase)

pulldown assay (*22, 23*). Hexahistidine-tagged CUP9 or its mutant derivatives were expressed in *E. coli* as fusions to the C-terminus of GST (GST-CUP9-his_6), and were purified by Ni-NTA affinity chromatography (*24*). The degradation of GST-CUP9-his_6 (denoted below as GST-CUP9) in *S cerevisiae* was UBR1-dependent and otherwise similar to that of wild-type CUP9 ($t_{1/2} \approx 5$ min in the absence of dipeptides) (ref. (*17*) and data not shown). Equal amounts of an extract from *S. cerevisiae* overexpressing the full-length flag-tagged UBR1 (^{f}UBR1) (*24*), which could fully rescue the N-end rule pathway in *ubr1Δ S. cerevisiae*, were incubated in either the absence or presence of various dipeptides (at 1 mM) with glutathione-Sepharose beads preloaded with GST-CUP9. The bound proteins were eluted, fractionated by SDS-PAGE, and the amount of bound ^{f}UBR1 was assessed by immunoblotting (*25*). The cognate Ub-conjugating (E2) enzyme RAD6 (GST-RAD6), a known UBR1 ligand (*14*), bound to ^{f}UBR1, whereas GST alone did not, in either the presence or absence of dipeptides (Fig. 1A, lanes 2 and 10; Fig. 1B, lane 10; Fig. 1C, lane 10). In the absence of added dipeptides or in the presence of dipeptides bearing stabilizing N-terminal residues, such as Ala-Arg or Ala-Leu, ^{f}UBR1 exhibited a detectable but weak binding to GST-CUP9 (Fig. 1A, lanes 3, 5, 7). The binding of ^{f}UBR1 to GST-CUP9 was slightly but reproducibly increased in the presence of Arg-Ala, a type 1 dipeptide, but not in the presence of Leu-Ala, a type 2 dipeptide (Fig. 1A, lanes 4 and 6 versus lanes 3, 5, 7). Remarkably, the binding of ^{f}UBR1 to GST-CUP9 was greatly increased in the presence of *both* type 1 and type 2 dipeptides, Arg-Ala and Leu-Ala (Fig. 1A, lane 8 versus lanes 3, 4, 6).

To determine whether the highly synergistic effect of type 1 and type 2 dipeptides was due to the presence of both *kinds* of dipeptides in a set, rather than to the identities of their N-terminal residues, we tested various combinations of other type 1 dipeptides (Lys-Ala and His-Ala) and type 2 dipeptides (Phe-Ala and Trp-Ala). Strong enhancement of the ^{f}UBR1-CUP9 interaction was observed only if both a type 1 and a type 2 dipeptide were present in the binding assay (Fig. 1A-C). The synergistic effect of a pair of type 1 and type 2 dipeptides on the interaction between ^{f}UBR1 and GST-CUP9 became detectable at 10 μM Arg-Ala plus 10 μM Leu-Ala, was strongly increased at 0.1 mM, and reached near-saturation at 1 mM Arg-Ala plus 1 mM Leu-Ala (Fig. 1D and data not shown). Note that 0.1 mM Arg-Ala plus 0.1 mM Leu-Ala increased the binding of ^{f}UBR1 to GST-CUP9 significantly more than did Arg-Ala alone at 1 mM (Fig. 1D, lane 6 versus lane 8). In contrast to type 1 and type 2 dipeptides, the corresponding free amino acids, in several tested combinations, did not increase the binding of UBR1 to CUP9 (data not shown).

To verify the specificity of the ^{f}UBR1-CUP9 interaction, we performed GST-pulldown assays using CUP9 mutants whose UBR1-dependent *in vivo* degradation was decreased or nearly abolished (Fig. 2B, E). These single-residue mutants of CUP9 were isolated through a genetic screen for metabolically stabilized CUP9 (*26*). All of the substitutions were located within the last 33 residues of the 306-residue CUP9, thereby defining an essential part of CUP9 degron (Fig. 2B, E). GST fusions to these CUP9 mutants, and also to a CUP9 mutant that contained the Asn-215 → Ser mutation in the homeodomain ($CUP9^{N215S}$), were produced in *E. coli*, purified as described above, and subjected

to GST-pulldown assays with ^{f}UBR1 (Fig. 3A-C). CUP9^{N215S} was degraded *in vivo* indistinguishably from wild-type CUP9 (*17*). In agreement with these results, GST-CUP9 and GST-CUP9^{N215S} exhibited similar binding to ^{f}UBR1 in GST-pulldown assays (Fig. 3A, lanes 3 and 4). By contrast, all of the metabolically stabilized CUP9 mutants exhibited weaker binding to ^{f}UBR1 (Fig. 3A, B). Moreover, the relative decrease in binding of ^{f}UBR1 to CUP9 mutants correlated with the extent of their metabolic stabilization in *S. cerevisiae*. For example, GST-CUP9^{K289E} and GST-CUP9^{L294P}, which had the longest *in vivo* half-lives (Fig. 2B, E), were found to bind most weakly to ^{f}UBR1 (Fig. 3B). The slightly-more unstable GST-CUP9^{L294V} and GST-CUP9^{R287H} bound more strongly to ^{f}UBR1, but still significantly below the binding of wild-type GST-CUP9 under the same conditions (Fig. 3B). These results (Figs. 2b, E and 3B) confirmed the specificity of the UBR1-CUP9 interaction *in vitro*.

The strikingly synergistic effect of a pair of type 1 and type 2 dipeptides on the UBR1-CUP9 interaction (Fig. 1) makes UBR1, operationally, a sensor for the simultaneous presence of both type 1 and type 2 dipeptides. To determine whether this *in vitro* effect could be detected at the level of dipeptide uptake *in vivo*, we used *S. cerevisiae* auxotrophic for Leu. Ala-Leu, bearing a stabilizing N-terminal residue, could not rescue the strain's growth if present at 230 μM in the growth medium (Fig. 2C, D). However, the growth was rescued if the Leu-lacking type 1 and type 2 dipeptides Arg-Ala and Phe-Ala were present together at 0.6 μM each, in addition to 230 μM Ala-Leu (Fig. 2C, D). Under the same conditions, Arg-Ala alone (type 1 dipeptide) was significantly less effective, whereas Phe-Ala alone (type 2 dipeptide) could not rescue cell growth (Fig. 2C,

D). Since food sources of *S. cerevisiae* in the wild are likely to contain diverse mixtures of short peptides, the sensitivity of UBR1-CUP9 interaction to the simultaneous presence of a type 1 and a type 2 dipeptide is likely to be physiologically relevant. At higher concentrations in the medium (> 2 μM), single dipeptides with destabilizing N-terminal residues were observed to stimulate the import of a nutritionally essential dipeptide bearing a stabilizing N-terminal residue (*17*). This may result, for example, from subthreshold *in vivo* levels of diverse dipeptides present in cells at all times, and/or from the highly cooperative dose-response aspect of peptide import, given the circuit's positive feedback (*17*). These issues remain to be explored in detail.

The type 1 and type 2 binding sites of UBR1 have been mapped, using a genetic screen, to a ~600-residue N-terminal region of the 1,950-residue UBR1 (*27, 28*). To map the approximate location of a third substrate-binding site of UBR1 that recognizes CUP9 (Fig. 2B), we examined the binding of GST-CUP9 to ^{f}UBR1$^{1-1175}$ and ^{f}UBR1$^{1-717}$, two N-terminal fragments of the 1,950-residue ^{f}UBR1 (Fig. 2A). (The superscript numbers refer to residue positions in the untagged wild-type UBR1.) These assays, carried out in the presence of 1 mM Arg-Ala plus 1 mM Leu-Ala, showed that ^{f}UBR1$^{1-1175}$ bound to GST-CUP9, whereas ^{f}UBR1$^{1-717}$ did not (Fig. 3D), indicating that UBR1 sequences between residues 717 and 1175 were required for the UBR1 binding to CUP9. Both the ^{f}UBR1$^{1-1175}$ fragment (which binds to CUP9) and the ^{f}UBR1$^{1-717}$ fragment (which does not bind to CUP9) retained the capacity to bind either dipeptides or test proteins bearing destabilizing N-terminal residues (data not shown), confirming that the type 1 and type 2 binding sites of UBR1 are distinct from its CUP9-binding site (*17*).

Remarkably, whereas strong binding of the full-length ^{f}UBR1 to GST-CUP9 required the presence of a pair of type 1 and type 2 dipeptides (Fig. 1), no such dependence was observed with the ^{f}UBR1$^{1-1175}$ fragment: it exhibited strong binding to GST-CUP9 in either the presence or absence of dipeptides (Fig. 3E). The same results were obtained with inputs of ^{f}UBR1$^{1-1175}$ that differed by six-fold (Fig. 3g), confirming this conclusion. Additional assays confirmed the specificity of observed interaction, in that the ^{f}UBR1$^{1-1175}$ fragment and the full-length ^{f}UBR1 exhibited similar patterns of affinity for the CUP9 mutants (Fig. 2B, D), with GST-CUP9^{K289E} and GST-CUP9^{L294P} binding most weakly, in either the absence or presence of dipeptides (Fig. 3F versus Fig. 3A, B). The ^{f}UBR1$^{1-1175}$ fragment lacked the RING-H2 domain of UBR1 but contained its BRR (basic residues-rich) region (Fig. 2A). Together, these two adjacent regions mediate the interaction between UBR1 and RAD6, the cognate E2 enzyme (*14*). The UBR1$^{1-1140f}$ fragment, which lacked the entire RAD6-binding site (Fig. 2A), also bound to GST-CUP9 constitutively (irrespective of the presence or absence of dipeptides) (Fig. 3H). Thus, the CUP9-binding site and the RAD6-binding site can be separated in UBR1 fragments.

^{f}UBR1$^{1-1367}$ and UBR1$^{1-1140f}$, which, respectively, contained and lacked the RAD6-binding site (Fig. 2A), bound to GST-CUP9 with similar apparent affinities (Fig. 4B, lanes 4-6 versus lanes 1-3). In contrast, both of the longer fragments, ^{f}UBR1$^{1-1540}$ and ^{f}UBR1$^{1-1700}$, exhibited weak binding to GST-CUP9 in either the presence or absence of Arg-Ala plus Leu-Ala (Fig. 4A, lanes 7-9, and Fig. 4B, lanes 7-9). The still longer ^{f}UBR1$^{1-1818}$ fragment, which encompassed the

108-residue, C-terminus-proximal region termed UBLC (UBR/Leu/Cys) (Fig. 2A) (*13, 28*), bound to GST-CUP9 strongly in the presence of Arg-Ala plus Leu-Ala; the omission of dipeptides decreased the binding, but to a lesser extent than with the full-length ^{f}UBR1 (Fig. 4A, lanes 4-6 versus lanes 1-3). The C-terminal UBLC domain of UBR1 (Fig. 2A) (*28*) contains two conserved cysteines, Cys-1703 and Cys-1706. ^{f}UBR1$^{C1703,1706S(A)}$, two variants of the full-length UBR1 in which both Cys residues were converted to either Ser or Ala, exhibited a pattern of CUP9 binding similar to that of ^{f}UBR1$^{1-1700}$, which lacks the UBLC domain (Fig. 4C versus Fig. 4a, lanes 7-9). Specifically, no strong increase in the binding of ^{f}UBR1$^{C1703,1706S(A)}$ to GST-CUP9 was observed in the presence of Arg-Ala plus Leu-Ala, while in the absence of dipeptides the binding of ^{f}UBR1$^{C1703,1706S(A)}$ to GST-CUP9 was slightly stronger than the binding of wild-type UBR1 (Fig. 4C, lanes 4-9 versus lanes 1-3).

The type 1 and type 2 binding sites, located within the first ~600 residues of the 1,950-residue UBR1, are followed by the CUP9-binding site, which, as shown above (Fig. 3), precedes the RAD6-binding site (the BRR region plus the RING-H2 domain) (Figs. 2A and 3D, H) (*14*). As shown in Figure 3E, G, H, the ~600-residue C-terminal region of UBR1, downstream from the RING-H2 domain (Fig. 2A), is required for the suppression of CUP9 binding to UBR1 in the absence of dipeptides, but is not required for the high-affinity interaction between UBR1 and CUP9. Thus, the binding of CUP9 by wild-type UBR1 is autoinhibited by its C-terminal region, which folds back and interacts, in part through the UBLC domain (see below), with a region that encompasses (or is adjacent to) the CUP9-binding site of UBR1. In this "closed" conformation, steric

hindrance by the C-terminal region of UBR1 precludes the binding of UBR1 to CUP9 (Fig. 5). A pair of type 1 and type 2 dipeptides synergistically relieves the autoinhibition by inducing an opening of the closed UBR1 conformation, thus abolishing the occlusion of the CUP9-binding site. (Previously known examples of such reversibly autoinhibited, jackknife-like proteins include N-WASP, a regulator of actin polymerization (*29, 30*).) In the absence of type 1 and type 2 dipeptides, the probability of closed → open transition in UBR1 is small. As a result, there are few UBR1 molecules in the open conformation that is capable of binding to CUP9, thus accounting for the observed weak binding UBR1 to CUP9 in the absence of dipeptides (Fig. 1). A pair of type 1 and type 2 dipeptides, upon their binding to the type 1 and type 2 sites of UBR1, greatly increases the probability of closed → open transition, causing a larger fraction of UBR1 molecules to acquire the binding-competent conformation, and yielding the observed strong binding to CUP9 (Fig. 5). According to this (parsimonious) model, the weak binding of UBR1 to CUP9 in the absence of a pair of type 1 and type 2 dipeptides (Fig. 1) reflects not a weak binding of CUP9 by most UBR1 molecules, but rather a low but non-zero probability of the closed → open transition in UBR1 under these conditions. In other words, the apparently weak binding of UBR1 for CUP9 in the absence of dipeptides (Fig. 1) is a manifestation of the strong binding of CUP9 by a minority of open-conformation UBR1 molecules in the largely closed-conformation UBR1 ensemble. The binding of a pair of type 1 and type 2 dipeptides to UBR1 shifts the equilibrium toward an ensemble where the open, CUP9 binding-competent conformation of UBR1 becomes a predominant species (Fig. 5).

We carried out yet another test of this model by asking whether GST-UBR1$^{1678\text{-}1950}$, a fusion between GST and the 273-residue C-terminal fragment of UBR1 that contained the UBLC domain, could bind, "in *trans*," to UBR1$^{1\text{-}1140f}$, the N-terminal fragment of UBR1 that contained all three of its substrate-binding sites but lacked the RAD6-binding site (Fig. 2A). The UBR1$^{1\text{-}1140f}$ fragment indeed bound to the UBLC-containing GST-UBR1$^{1678\text{-}1950}$ fragment (Fig. 4D). Remarkably, this interaction was strongly inhibited by a type 2 dipeptide (Leu-Ala), but neither by a type 1 dipeptide (Arg-Ala) nor by dipeptides with stabilizing N-terminal residues (Fig. 4D). Note, in conjunction with this result, that the binding of CUP9 to the full-length UBR1 was not increased by a type 2 dipeptide (Leu-Ala), was weakly increased by a type 1 dipeptide (Arg-Ala), and was greatly increased in the presence of both Arg-Ala and Leu-Ala (Fig. 1).

A version of the autoinhibition mechanism that accounts for this seemingly paradoxical set of results, and also explains specific earlier observations, is as follows. The model (Fig. 5) posits that the binding of Arg-Ala to the (always accessible) type 1 site of UBR1 induces a conformational change that increases the accessibility (and/or affinity) of the type 2 binding site to Leu-Ala. It is the latter binding event that induces the closed → open transition in UBR1 and the unmasking of its CUP9-binding site. In addition to explaining the fact that both type 1 and type 2 dipeptides must be present together for a strong increase in the binding of full-length UBR1 to CUP9 (Fig. 1), this sequential-binding mechanism is also consistent with the finding that a type 2 dipeptide alone, but not a type 1 dipeptide alone, is sufficient for abolishing the binding of the UBLC-containing C-terminal fragment UBR1$^{1678\text{-}1950}$ to the

N-terminal fragment UBR1$^{1-1140}$ (Fig. 4D). Specifically, whereas the binding of a type 1 dipeptide increases the accessibility of the type 2 binding site in the full-length UBR1, the type 2 site is likely to be always accessible in the UBR1$^{1-1140}$-UBR1$^{1678-1950}$ complex, because the UBR1$^{1678-1950}$ fragment contains the UBLC domain but not the entire C-terminal region of UBR1 (Fig. 5). Interestingly, the sequential-binding mechanism (Fig. 5) also explains two earlier findings: that type 1 dipeptides accelerate the degradation of type 2 protein substrates by the N-end rule pathway, whereas type 2 dipeptides do not accelerate the degradation of type 1 substrates (*13, 15, 16*).

Neither the full-length ^{f}UBR1$^{C1703,1706S(A)}$ mutants nor the ^{f}UBR1$^{1-1700}$ fragment of wild-type UBR1 that lacked the UBLC domain could rescue the N-end rule pathway in *ubr1Δ S. cerevisiae*, even upon overexpression (data not shown). Thus, the C-terminus-proximal UBLC domain (Fig. 2A) performs yet another, currently unknown, *in vivo* function, in addition to making possible the closed conformation of UBR1 (Fig. 5). One possibility is that the UBLC domain might be the site of a Ub~UBR1 thioester. Although UBR1 and other E3s of the RING-domain family are presumed not to form Ub~E3 thioesters, in contrast to the HECT-domain family of E3s (*1, 5*), the possibility of a functionally relevant Ub thioester in a RING-containing E3 remains to be precluded definitively, at least with UBR1.

Autoinhibition is an essential feature of many regulatory proteins, including protein kinases and phosphatases (*31*), nitric oxide synthases (*32*), transcription factors (*33*), and regulators of actin polymerization (*29, 30*). Autoinhibition has not been reported previously for Ub ligases. Mammalian

cells express at least three distinct homologs of *S. cerevisiae* UBR1, termed UBR1, UBR2 and UBR3, that contain the conserved (and similarly arranged) UBHC, RING-H2, and UBLC domains (*13, 28*). Moreover, similarly to *S. cerevisiae* UBR1, the mouse UBR1 and UBR2 E3s contain binding sites for the type 1 and type 2 destabilizing N-terminal residues, and the functional properties of these sites (type 1 dipeptides stimulate the degradation of type 2 N-end rule substrates, but not vice versa) are also similar to those of yeast UBR1 (*13, 15, 16*). Mouse UBR1 (E3α) is known to contain a third substrate-binding site as well (*34*). These extensive similarities to yeast UBR1 suggest that metazoan E3s of the UBR family are controlled through autoinhibition. Our findings also indicate that natural compounds (either small molecules or proteins) may regulate the activity of diverse Ub-dependent pathways through the abrogation or induction of autoinhibition in Ub ligases.

References and notes

1. A. M. Weissman, *Nature Rev. Mol. Cell. Biol.* **2**, 169 (2001).
2. W. P. Tansey, *Genes & Development* **15**, 1045 (2001).
3. A. Hershko, A. Ciechanover, A. Varshavsky, *Nature Med.* **10**, 1073 (2000).
4. J. D. Laney, M. Hochstrasser, *Cell* **97**, 427 (1999).
5. C. Pickart, *Annu. Rev. Biochem.* **70**, 503 (2001).
6. D. Voges, P. Zwickl, W. Baumeister, *Annu. Rev. Biochem.* **68**, 1015 (1999).
7. M. Rechsteiner, in *Ubiquitin and the Biology of the Cell* J. M. Peters, J. R. Harris, D. Finley, Eds. (Plenum Press, New York, 1998) p. 147.
8. B. A. Schulman *et al.*, *Nature* **408**, 381 (2000).
9. A. Bachmair, D. Finley, A. Varshavsky, *Science* **234**, 179 (1986).
10. A. Bachmair, A. Varshavsky, *Cell* **56**, 1019 (1989).
11. T. Suzuki, A. Varshavsky, *EMBO J.* **18**, 6017 (1999 Nov 1, 1999).
12. A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **93**, 12142 (1996).
13. Y. T. Kwon, Z.-X. Xia, I. V. Davydov, S. H. Lecker, A. Varshavsky, *Mol. Cell. Biol.* (in press) (2001).
14. Y. Xie, A. Varshavsky, *EMBO J.* **18**, 6832 (1999).
15. D. K. Gonda *et al.*, *J. Biol. Chem.* **264**, 16700 (1989).
16. R. T. Baker, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **87**, 2374 (1991).
17. G. C. Turner, F. Du, A. Varshavsky, *Nature* **405**, 579 (2000).
18. H. Rao, F. Uhlmann, K. Nasmyth, A. Varshavsky, *Nature* **410**, 955 (2001).
19. K. Nasmyth, J. M. Peters, F. Uhlmann, *Science* **288**, 1379 (2000).
20. C. Byrd, G. C. Turner, A. Varshavsky, *EMBO J.* **17**, 269 (1998).
21. K. Alagramam, F. Naider, J. M. Becker, *Mol. Microbiol.* **15**, 225 (1995).
22. Y. Xie, A. Varshavsky, *Proc. Natl. Acad. Sci. USA* **97**, 2497 (2000).

23. F. M. Ausubel *et al.*, Eds., *Current Protocols in Molecular Biology.* (Wiley-Interscience, New York, 2000).

24. DNA fragments encoding either wild-type CUP9, or its mutant derivatives, or RAD6, or the 273-residue C-terminal fragment of UBR1 ($UBR1^{1678-1950}$) were subcloned into pGEX-2TK (Pharmacia), downstream of, and in frame with, the ORF of GST. The final constructs (except the one containing RAD6) also bore the C-terminal His_6 tag (*23*). The resulting plasmids were cotransformed into *E. coli* BL21(DE3) together with pRI952 (a gift from Dr. S. W. Stevens, Division of Biology, Caltech), expressing tRNAs for the codons AGG, AGA and AUA. Transformants were grown in antibiotics-supplemented LB medium (*23*) at 37°C to A_{600} of 0.8-1.0. Isopropyl-β-D-thiogalactopyranoside (IPTG) was then added to a final concentration of 1 mM, followed by a 3-hr incubation. The cells were harvested by centrifugations, washed with PBS, and frozen in liquid N_2. The cell pellet was resuspended in buffer A (10% glycerol, 0.3 M NaCl, 10 mM imidazole, 10 mM β-mercaptoethanol, 25 mM K_2HPO_4-KH_2PO_4 (pH 7.8)) (6 ml of buffer per 1 g of pellet). Freshly dissolved chicken egg lysozyme in the same buffer was added to a final concentration of 0.5 mg/ml, followed by incubation at 4°C for 30 min. The cells were further disrupted by sonication, 3 times for 1 min each, at 1-min intervals, followed by the addition of NP40 to a final concentration of 0.5%. The suspension was centrifuged at 11,200*g* for 30 min. The supernatant was added to a tube containing 2 ml of Ni-NTA resin (Qiagen). After gentle rotation at 4°C for 10 min, the resin was transferred to a 5-ml column and washed with 10

ml of buffer A. A GST fusion protein was then eluted with buffer A containing 0.3 M imidazole. To overexpress UBR1 and its fragments in *S. cerevisiae*, we used SC295 (*MATa, GAL4 GAL80 ura3-52, leu2-3,112 reg1-501 gal1 pep4-3*), a gift from Dr. S. A. Johnston (Univ. of Texas, SW Medical Center, Dallas TX, USA). N-terminally flag-tagged (*23*), full-length *S. cerevisiae* UBR1 (^{f}UBR1) and its truncated derivatives were expressed in SC295 from the P_{ADH1} promoter in a high copy vector. Construction details are available upon request. A colony of yeast transformant was inoculated into 20 ml of SD medium (*23*), and cells were grown at 30°C until mid-exponential phase, then re-inoculated into 2 l of SD medium, and were grown to A_{600} of ~1.0. The cells were harvested by centrifugation, washed with cold PBS, and frozen in liquid N_2. The frozen pellet was ground to fine powder in liquid N_2, using mortar and pestle. The powder was thawed and resuspended (6 ml of buffer per 1 g of pellet) in the lysis buffer (10% glycerol, 0.5% Nonidet-P40 (NP40), 0.2 M KCl, 1 mM dithiothreitol (DTT), 50 mM HEPES (pH 7.5)) containing protease inhibitors (Roche). The suspension was centrifuged at 11,200*g* for 30 min at 4°C. Samples of the supernatant were stored at -80°C.

25. Either a GST fusion protein or GST alone (~2 μg) was diluted to 0.5 ml in the loading buffer (10% glycerol, 0.5 M NaCl, 1 mM EDTA, 1% NP-40, 50 mM Tris-HCl, pH 8.0), and incubated with 20 μl of glutathione-Sepharose beads (Pharmacia) for 20 min at 4°C. The beads were washed once with 0.5 ml of the loading buffer and once with 0.5 ml of the binding buffer (10% glycerol, 50 mM NaCl, 0.05% NP-40, 50 mM Na-HEPES, pH 7.8). The washed beads,

in 100 μl of the binding buffer, were incubated at 4°C for 1 hr with a yeast extract (160 μl) containing ^{f}UBR1 or its truncated derivative, in the absence or presence of dipeptides (Sigma) at stated concentrations. Ovalbumin (0.1 mg/ml in the binding buffer) was used to dilute yeast extract. The beads were washed three times with 0.2 ml of the binding buffer containing dipeptides at the same concentrations. The beads were then suspended in 20 μl of SDS-PAGE loading buffer, and heated at 100°C for 5 min, followed by SDS-8% PAGE, electrophoretic transfer of proteins onto PVDF membrane (Millipore), and immunoblotting with anti-FLAG M2 antibody (Sigma) (22).

26. F. Navarro-Garcia, G. C. Turner, and A. Varshavsky, unpublished data.

27. A. Webster, M. Ghislain, and A. Varshavsky, unpublished data.

28. Y. T. Kwon *et al.*, *Proc. Natl. Acad. Sci. USA* **95**, 7898 (1998).

29. R. Rohatgi, H. H. Ho, M. Kirschner, *J. Cell. Biol.* **150**, 1299 (2000).

30. K. E. Prehoda, J. A. Scott, R. D. Mullins, W. A. Lim, *Science* **290**, 801 (2000).

31. W. Xu, A. Doshi, M. Lei, M. J. Eck, S. C. Harrison, *Mol. Cell* **3**, 629 (1999).

32. C. R. Nishida, P. R. Ortis de Montellano, *J. Biol. Chem.* **274**, 14692 (1999).

33. W. Y. Kim *et al.*, *EMBO J.* **18**, 1609 (1999).

34. T. G. Lawson *et al.*, *J. Biol. Chem.* **274**, 9871 (1999).

35. We are grateful to S. A. Johnston, S. W. Stevens and A. Webster for strains and plasmids, and to G. C. Turner, A. Webster and M. Ghislain for permission to cite unpublished data. We thank former and current members of the Varshavsky laboratory, particularly G. C. Turner and H. Rao, for discussions and advice, and G. C. Turner for his comments on the manuscript. This work was supported by

grants to A. V. from the NIH. F. N.-G. was supported by postdoctoral fellowships from the Spanish Ministry of Education, the Fulbright Foundation, and the Del Amo-UCM Foundation.

Figures and legends

Figure 1 Type 1 and type 2 dipeptides, if present together, greatly increase the binding of UBR1 to CUP9. Equal amounts of an extract from *S. cerevisiae* overexpressing the N-terminally flag-tagged UBR1 (^{f}UBR1) were incubated with glutathione-Sepharose beads preloaded either with GST alone, or with GST-CUP9, or with GST-RAD6 , in either the presence or absence of dipeptides. The bound proteins were eluted, fractionated by SDS-PAGE, and immunoblotted with anti-flag antibody. Unless otherwise stated, each dipeptide was present at 1 mM. An asterisk denotes a fragment of ^{f}UBR1. The "5% input" lanes refer to a directly loaded sample of the yeast extract that corresponded to 5% of the extract's amount used in GST-pulldown assays.

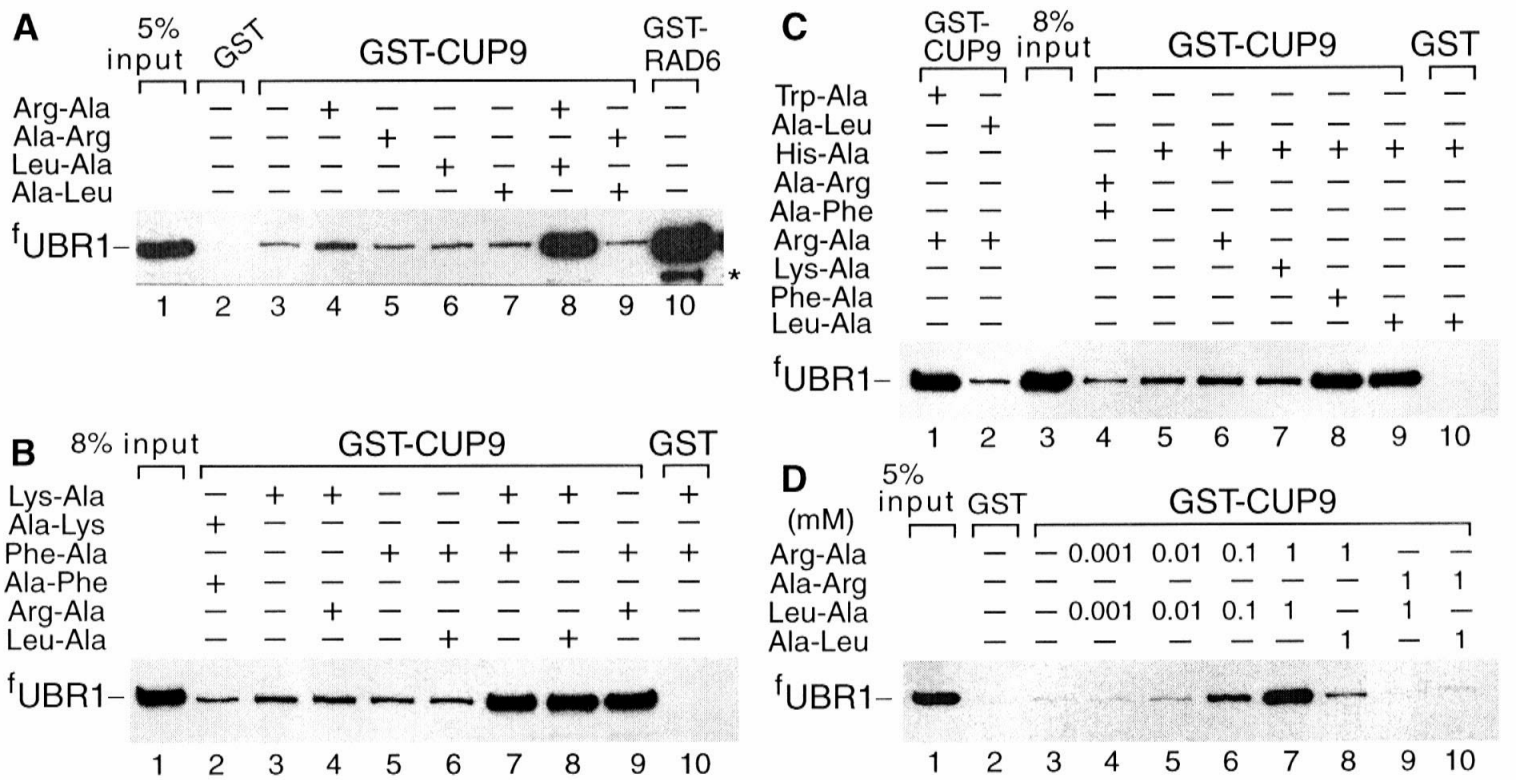

A
5% input
GST
GST-CUP9
GST-RAD6
Arg-Ala
Ala-Arg
Leu-Ala
Ala-Leu
fUBR1–
*
1 2 3 4 5 6 7 8 9 10
B
8% input
GST-CUP9
GST
Lys-Ala
Ala-Lys
Phe-Ala
Ala-Phe
Arg-Ala
Leu-Ala
fUBR1–
1 2 3 4 5 6 7 8 9 10
C
GST-CUP9
8% input
GST-CUP9
GST
Trp-Ala
Ala-Leu
His-Ala
Ala-Arg
Ala-Phe
Arg-Ala
Lys-Ala
Phe-Ala
Leu-Ala
fUBR1–
1 2 3 4 5 6 7 8 9 10
D
5% input
GST
GST-CUP9
(mM)
Arg-Ala
Ala-Arg
Leu-Ala
Ala-Leu
0.001
0.01
0.1
1
fUBR1–
1 2 3 4 5 6 7 8 9 10

Figure 2 *S. cerevisiae* UBR1, CUP9, and plating efficiency assays. (**A**) A diagram of the 1,950-residue UBR1, indicating the regions conserved in fungal and metazoan UBR proteins (*28*). His and Cys are some of the conserved residues in the UBHC region (UBR/His/Cys) (residues 123-193) that encompasses Gly-173 and Asp-176, which are essential for the integrity of the type 1 substrate-binding site of UBR1 (*27*). The BRR (basic/residues/rich) region (residues 1,165-1,175) mediates the interaction of UBR1 with the polyacidic tail of the RAD6 E2 enzyme (*14*). The RING-H2 domain (residues 1219-1333), which together with the BRR region mediates the UBR1-RAD6 interaction (*14*), is an essential domain present in a large class of E3s, including UBR-family E3s (*1, 5*). Leu and Cys are some of the conserved residues in the C-terminus-proximal, 108-residue region (1,702-1,809) termed UBLC (UBR/Leu/Cys) (*13, 28*). Also shown are the C-terminally truncated derivatives of UBR1 used in the GST-CUP9 binding assays, and a fragment containing the UBLC domain, used as a GST-UBR1$^{1678\text{-}1950}$ fusion in the UBR1$^{1\text{-}1140}$-UBR1$^{1678\text{-}1950}$ binding assay (Fig. 4D) (see also the main text). UBR1 and its fragments bore N-terminal flag (*24*), with the exception of UBR1$^{1\text{-}1140}$, which contained C-terminal flag. (**B**) A diagram of the 306-residue CUP9 highlighting its homeodomain, the site of toxicity-reducing Asn-215 → Ser mutation (*17*), and the C-terminal region, whose sequence is shown below, together with single-residue alterations that decreased the UBR1-dependent degradation of the corresponding CUP9 derivatives in *S. cerevisiae* (see panel **E**). (**C**) Plating efficiency assays, carried out in minimal (SD) medium with *S. cerevisiae* SC295 (Leu$^-$) in the presence of 230 μM Ala-Leu and the indicated pairs of non-nutritious dipeptides, at 0.6 μM each. (**D**) Representative images of plates from

the assays in **C**. (**E**) Pulse-chase analysis of CUP9 and some of its long-lived mutants. *S. cerevisiae* JD52 (*MAT***a** *lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*) (*17*) were used that carried low copy number plasmids expressing, from the P_{MET25} promoter, Ub fusion proteins of the UPR (Ub/protein/reference) technique (refs. (*11, 17*), and refs. therein). Upon cotranslational cleavage by deubiquitylating enzymes (DUBs), these ^{f}DHFR-Ub-CUP9^{f} fusions yielded the long-lived reference protein flag-dihydrofolate reductase-Ub (^{f}DHFR-Ub) and either the flag-tagged wild-type CUP9 (CUP9^{f}) or its mutant derivatives (see panel **B**). Cells were labeled with ^{35}S-methionine for 5 min at 30°C, followed by chase for 5, 15, 30, and 60 min, extraction of proteins, immunoprecipitation with anti-flag antibody, SDS-PAGE, and quantitation, essentially as described (*11, 17, 18*). Thus determined *in vivo* half lives of CUP9 and its single-substitution derivatives in panel **B**: wild-type CUP9 (5 min); CUP9^{L294P} (59 min); CUP9^{L294V} (43 min); CUP9^{K289E} (61 min); CUP9^{E274K} (45 min); CUP9^{R287H} (47 min), CUP9^{T285A} (44 min) (data in **E**, and data not shown).

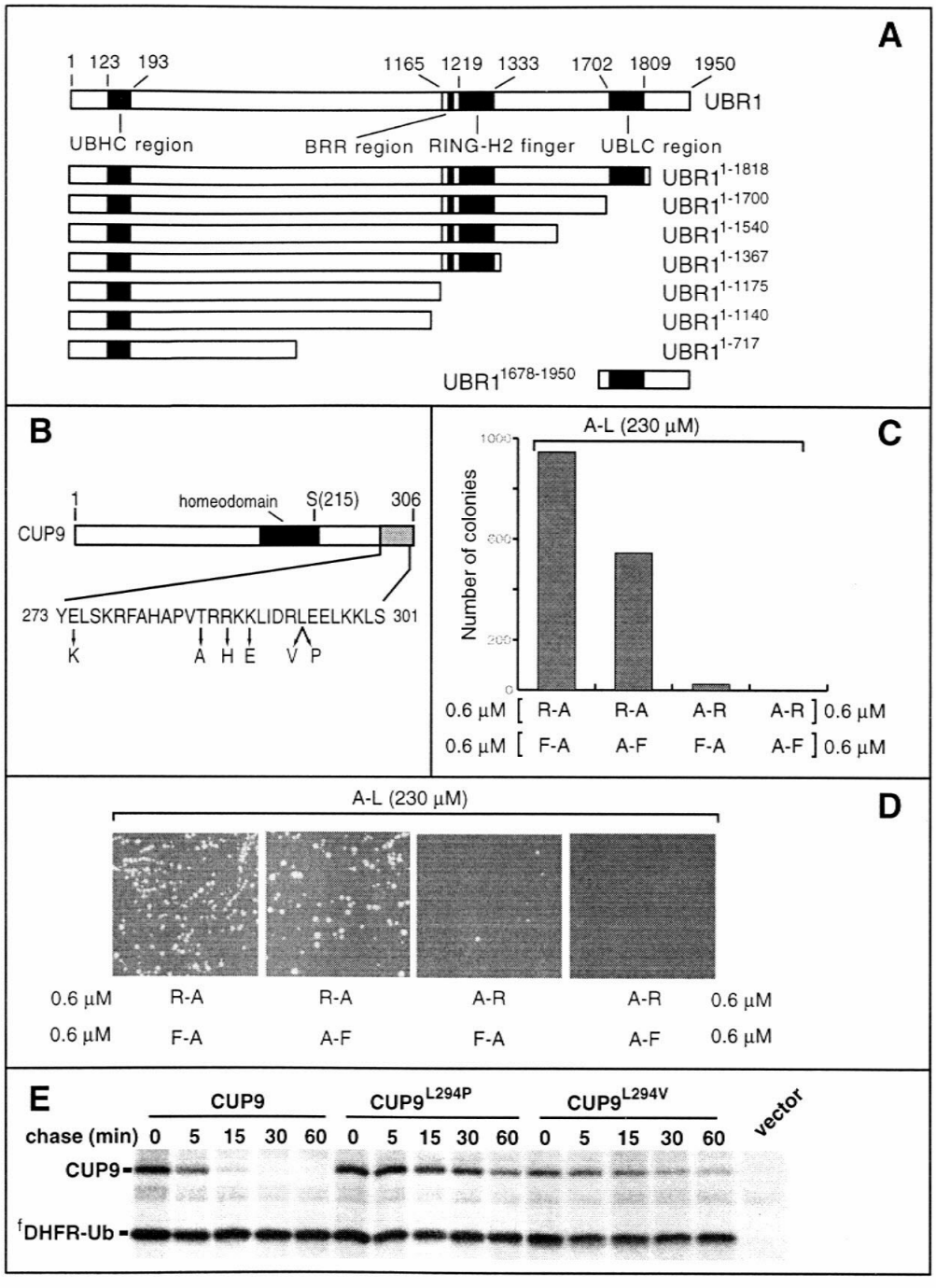

A
1 123 193
1165 1219 1333
1702 1809 1950
UBR1
UBHC region
BRR region
RING-H2 finger
UBLC region
UBR1^1-1818
UBR1^1-1700
UBR1^1-1540
UBR1^1-1367
UBR1^1-1175
UBR1^1-1140
UBR1^1-717
UBR1^1678-1950
B
1
homeodomain
S(215)
306
CUP9
273 YELSKRFAHAPVTRRKKLIDRLEELKKLS 301
K
A H E
V P
C
A-L (230 μM)
Number of colonies
1000
600
200
0
0.6 μM [R-A R-A A-R A-R] 0.6 μM
0.6 μM [F-A A-F F-A A-F] 0.6 μM
D
A-L (230 μM)
0.6 μM R-A R-A A-R A-R 0.6 μM
0.6 μM F-A A-F F-A A-F 0.6 μM
E
CUP9
CUP9^L294P
CUP9^L294V
vector
chase (min) 0 5 15 30 60 0 5 15 30 60 0 5 15 30 60
CUP9
fDHFR-Ub

Figure 3 Dipeptide-independent, high-affinity binding of CUP9 by N-terminal fragments of UBR1. (**A** and **B**) Comparisons of ^{f}UBR1 binding to GST-CUP9 fusions containing either wild-type CUP9 or its single-residue variants indicated in Fig. 2B. GST-pulldown assays were performed in the presence of 1 mM Arg-Ala and 1 mM Leu-Ala. (**C**) 30% of the GST-CUP9 and GST-RAD6 samples used in GST-pulldown assays of **B** were fractionated by SDS-10% PAGE and stained with Coomassie. (**D**) GST-pulldown assays, in the presence of 1 mM Arg-Ala and 1 mM Leu-Ala, with either GST alone or GST-CUP9, and either full-length ^{f}UBR1, or ^{f}UBR1$^{1-1175}$, or ^{f}UBR1$^{1-717}$. (**E**) Same as in **D** but with ^{f}UBR1$^{1-1175}$ in the presence of different dipeptides. Asterisks in **D** and **E** indicate cross-reacting proteins. **F**, Comparisons of the ^{f}UBR1$^{1-1175}$ binding to GST alone or to GST-CUP9 fusions containing either wild-type CUP9 or its mutant variants indicated in Fig. 2B. **G**, GST-pulldown assays with either full-length ^{f}UBR1 or ^{f}UBR1$^{1-1175}$ in the presence of either 1 mM Arg-Ala and 1 mM Leu-Ala (plus signs) or 1 mM Ala-Arg and 1 mM Ala-Leu (minus signs). Two input amounts of ^{f}UBR1$^{1-1175}$, differing by six-fold, were used (lanes 2 and 3). The corresponding ^{f}UBR1$^{1-1175}$ binding assays are in lanes 7 and 8 versus 9 and 10, respectively. **H**, UBR1$^{1-1140f}$ (which contained the C-terminal flag epitope), binds to CUP9 but not to RAD6. Comparisons, in the presence of either 1 mM Arg-Ala and 1 mM Leu-Ala (plus signs) or 1 mM Ala-Arg and 1 mM Ala-Leu (minus signs) of the ^{f}UBR1 and UBR1$^{1-1140f}$ binding to either GST-CUP9 or GST-RAD6.

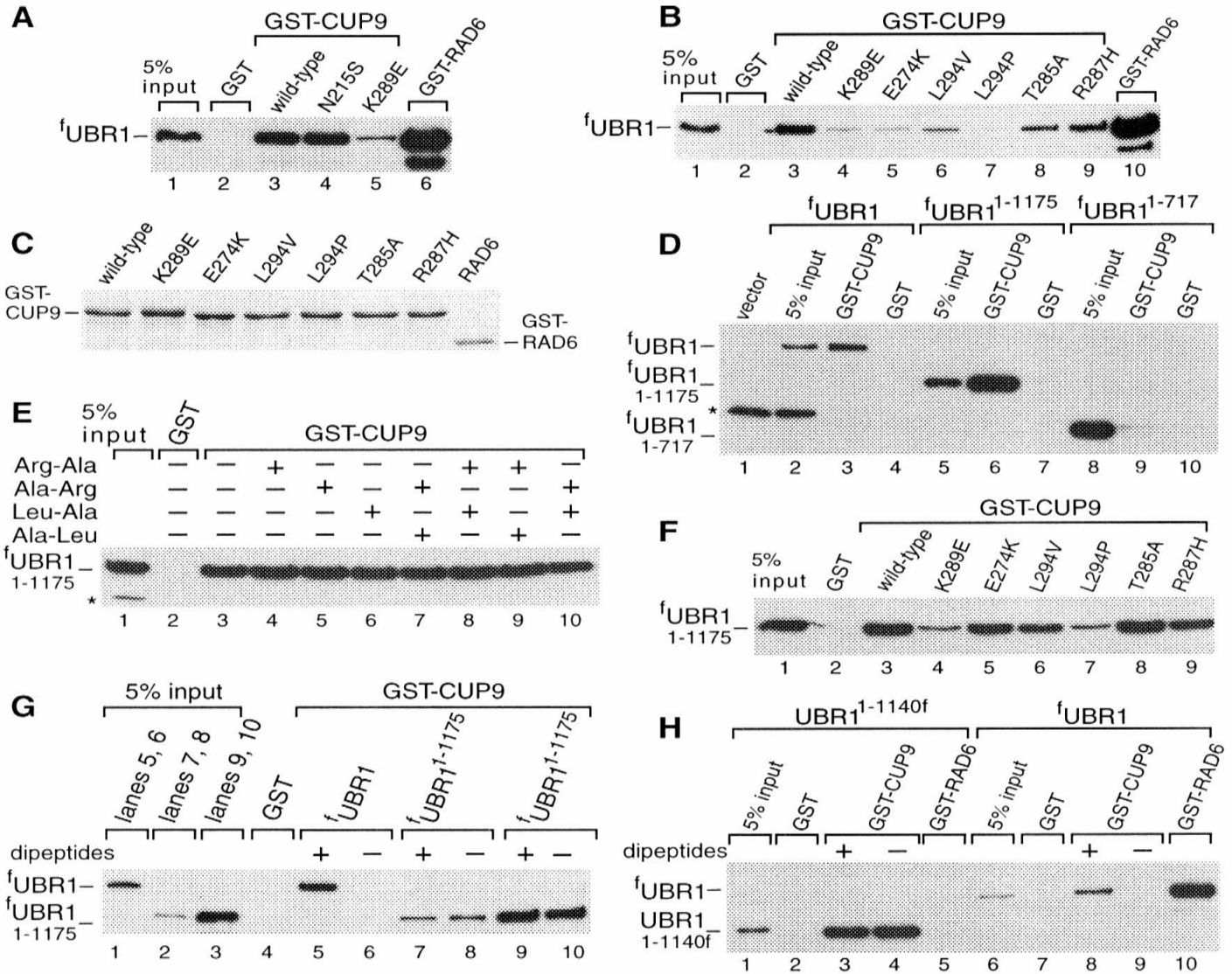

A
GST-CUP9
5% input
GST
wild-type
N215S
K289E
GST-RAD6
fUBR1–
1 2 3 4 5 6
B
GST-CUP9
5% input
GST
wild-type
K289E
E274K
L294V
L294P
T285A
R287H
GST-RAD6
fUBR1–
1 2 3 4 5 6 7 8 9 10
C
wild-type
K289E
E274K
L294V
L294P
T285A
R287H
RAD6
GST-CUP9–
GST-RAD6
D
fUBR1
fUBR1 1-1175
fUBR1 1-717
vector
5% input
GST-CUP9
GST
fUBR1–
fUBR1 1-1175–
fUBR1 1-717–
1 2 3 4 5 6 7 8 9 10
E
5% input
GST
GST-CUP9
Arg-Ala
Ala-Arg
Leu-Ala
Ala-Leu
fUBR1 1-1175–
1 2 3 4 5 6 7 8 9 10
F
GST-CUP9
5% input
GST
wild-type
K289E
E274K
L294V
L294P
T285A
R287H
fUBR1 1-1175–
1 2 3 4 5 6 7 8 9
G
5% input
GST-CUP9
lanes 5, 6
lanes 7, 8
lanes 9, 10
GST
fUBR1
fUBR1 1-1175
fUBR1 1-1175
dipeptides
fUBR1–
fUBR1 1-1175–
1 2 3 4 5 6 7 8 9 10
H
UBR1 1-1140f
fUBR1
5% input
GST
GST-CUP9
GST-RAD6
dipeptides
fUBR1–
UBR1 1-1140f–
1 2 3 4 5 6 7 8 9 10

Figure 4 The C-terminal UBLC domain of UBR1 binds to N-terminal region of UBR1. (**A**) GST-pulldown assays with GST-CUP9 and either the full-length ^{f}UBR1, or ^{f}UBR1$^{1-1818}$, or ^{f}UBR1$^{1-1700}$, in the presence of either 1 mM Arg-Ala and 1 mM Leu-Ala (plus signs) or 1 mM Ala-Arg and 1 mM Ala-Leu (minus signs). (**B**) Same as in **a**, but with the UBR1$^{1-1140f}$, ^{f}UBR1$^{1-1367}$, and ^{f}UBR1$^{1-1540}$ fragments. (**C**) Same as in **A**, but with the full-length wild-type ^{f}UBR1 and its full-length derivatives ^{f}UBR1^{C1703,1706A} and ^{f}UBR1^{C1703,1706S}. **D**, GST-pulldown assay for the interaction between UBR1$^{1-1140f}$ and GST-UBR1$^{1678-1950}$ (see Fig. 2A and the main text), in the presence of different dipeptides. Lanes 9 and 10 show the results of analogous assays with UBR1$^{1-1140f}$ and either GST-RAD6 or GST-CUP9. An asterisk indicates a nonspecific (cross-reacting) protein.

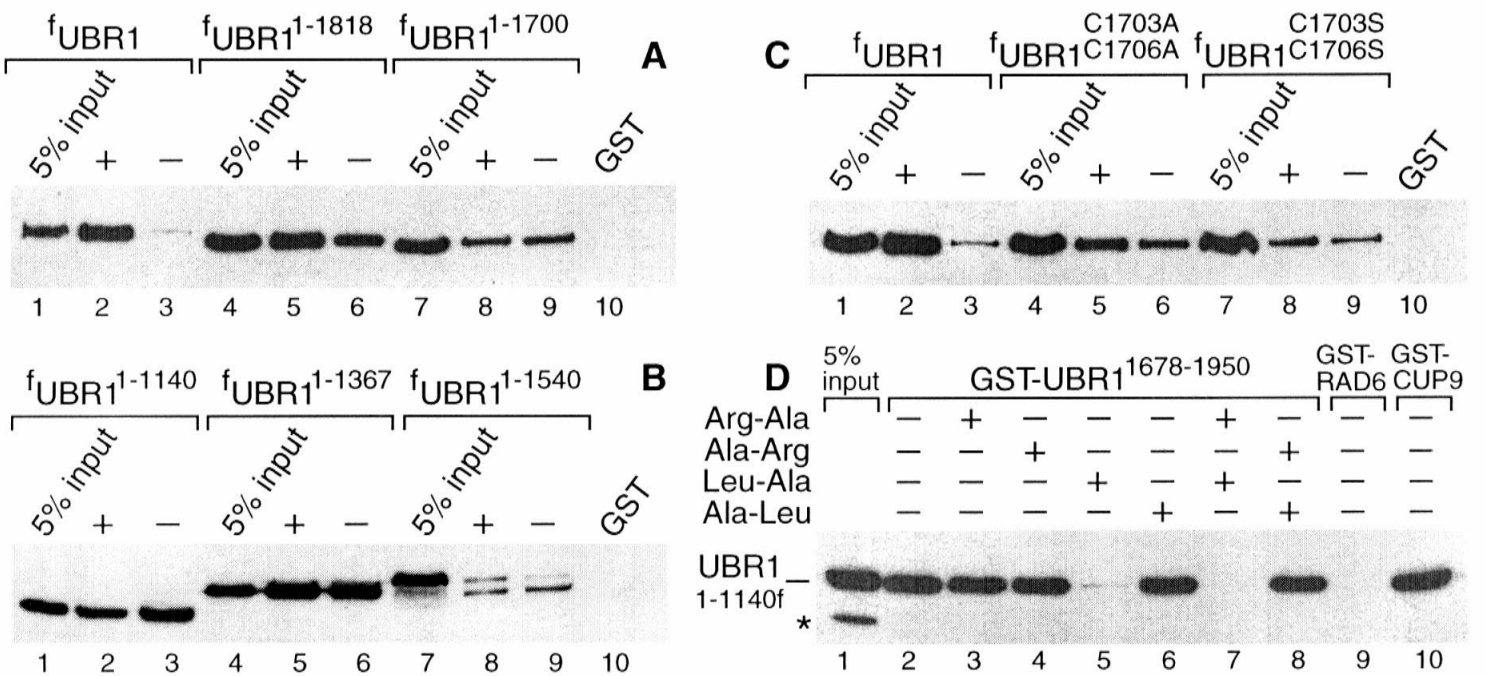
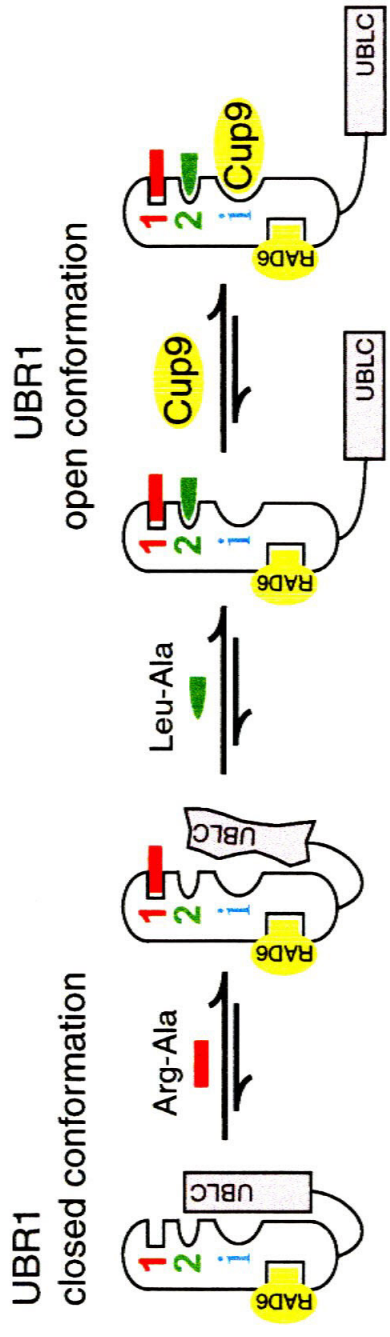

fUBR1
fUBR1 1-1818
fUBR1 1-1700
A
5% input
+
–
GST
1 2 3 4 5 6 7 8 9 10
fUBR1 1-1140
fUBR1 1-1367
fUBR1 1-1540
B
C
fUBR1
fUBR1 C1703A C1706A
fUBR1 C1703S C1706S
D
5% input
GST-UBR1 1678-1950
GST-RAD6
GST-CUP9
Arg-Ala
Ala-Arg
Leu-Ala
Ala-Leu
UBR1_ 1-1140f
*

Figure 5 Regulated autoinhibition of the ubiquitin ligase UBR1. The results of the present work (Figs. 1-4) indicate that the binding of the transcriptional repressor CUP9 by the Ub ligase UBR1 is autoinhibited by its C-terminal region, which folds back and interacts, in part through the UBLC domain, with a region that encompasses (or is adjacent to) the CUP9-binding site of UBR1, denoted by "i" (a site binding to a substrate's internal degron). In this "closed" conformation, steric hindrance by the C-terminal region of UBR1 precludes the binding of UBR1 to CUP9. A pair of type 1 and type 2 dipeptides (red rectangle and green bullet, respectively) synergistically relieves the autoinhibition, by causing an opening of the closed UBR1 conformation, thus abolishing the occlusion of the CUP9-binding site. In the absence of type 1 and type 2 dipeptides, the probability of the closed → open transition is low. As a result, the open (CUP9-binding) conformation of UBR1 is a minor species in the largely closed-conformation UBR1 ensemble. The binding of a pair of type 1 and type 2 dipeptides to the type 1 and type 2 binding sites of UBR1 shifts the equilibrium toward an ensemble where the CUP9-binding, open-conformation UBR1 becomes a predominant species. To account for the strong cooperativity between type 1 and type 2 dipeptides in the enhancement of CUP9 binding by UBR1 (Fig. 1), the model additionally posits that the binding of Arg-Ala (a type 1 dipeptide) to the always accessible type 1 site of UBR1 induces a conformational change that increases the accessibility (and/or affinity) of the type 2 binding site to Leu-Ala (a type 2 dipeptide), and that it is the latter binding which induces the closed → open transition in UBR1 and the unmasking of its CUP9-binding site. This version of the model explains yet another result of the present work, and two earlier findings as well (*13, 15, 16*) (see the main text). It is unknown whether the postulated conformational alteration of

UBR1, upon its binding to a type 1 dipeptide, involves the UBLC-containing C-terminal region of UBR1 (as depicted in the diagram) or its other domains. Experimental evidence for the arrangement of the three substrate-binding sites of UBR1 is described in the main text. The depicted orientation of the RAD6-binding site is solely to indicate its constitutive availability, in contrast to the CUP9-binding site.


UBR1
closed conformation
UBLC
1
2
i
RAD6
Arg-Ala
UBLC
1
2
i
RAD6
Leu-Ala
UBR1
open conformation
1
2
i
RAD6
Cup9
UBLC
1
2
i
RAD6
Cup9
UBLC

Chapter 5

Missions unfinished—

Experiments related to the N-end rule pathway

5.1 The internal degron of CUP9

In contrast to engineered N-end rule substrates which are targeted by UBR1 through their N-degron, CUP9 is targeted through an internal degron [1] (Chapter 2). In an early attempt to locate the internal degron of CUP9 [2], DNA sequences encoding the full-length CUP9 or its truncations that lacked, respectively, N-terminal 81, 153 or 221 residues (denoted as CUP9ΔN, where "N" referred to 81, 153 or 221) were fused in-frame to the DNA sequences encoding $DHFR^m$, a mouse DHFR variant with a C-terminal Myc epitope tag. The fusion proteins (denoted as CUP9-$DHFR^m$ or CUP9ΔN-$DHFR^m$) were expressed in the wild-type yeast cells and their half-lives measured using pulse-chase analysis. CUP9-$DHFR^m$ and CUP9Δ81-$DHFR^m$ were found to be degraded through the N-end rule pathway similarly to CUP9; in contrast, both CUP9Δ153-$DHFR^m$ and CUP9Δ221-$DHFR^m$ were stabilized dramatically [2]. These results suggested that the N-terminal 81-residue fragment of CUP9 was not required for the targeting of CUP9 by UBR1, and that the 72-residue region between residues 81 and 153 of CUP9 was essential for the integrity of its internal degron. It was unknown, however, whether CUP9Δ153-$DHFR^m$ and CUP9Δ221-$DHFR^m$ bound specifically to UBR1, since metabolic stabilization of CUP9Δ153-$DHFR^m$ and CUP9Δ221-$DHFR^m$ *in vivo* could result from the absence of "spatially appropriate" Lys residues.

A genetic screen (F. Navarro-Garcia, G. C. Turner and A. Varshavsky, personal communication) (Chapter 1) was carried out later to search for single-residue substitutions that stabilized URA3-CUP9 in the wild-type yeast cells. The *in vivo* half-life of the URA3-CUP9 fusion was similar to that of CUP9.

All of the single-residue substitutions that stabilized both the URA3-CUP9 fusion and CUP9 were located within the C-terminal 33 residues of the 306-residue CUP9. These single-residue substitutions were further demonstrated to weaken the affinity between UBR1 and CUP9, and the extent of affinity decrease correlated with the degree of the metabolic stabilization of the corresponding CUP9 mutants in *S. cerevisiae* (Chapter 4), suggesting that these single-residue substitutions directly perturbed the region through which CUP9 bound to UBR1.

In this report, GST-pulldown assays (Chapter 4) were carried out to identify the minimal region of CUP9 that was sufficient to interact specifically with UBR1 (Figure 5.1.1a). Yeast extracts containing ^{f}UBR1, the full-length UBR1 with an N-terminal FLAG epitope tag (Figure 5.1.1c) and UBR1$^{1\text{-}1140f}$, the N-terminal 1140-residue fragment of UBR1 with a C-terminal FLAG tag (Figure 5.1. 1b) were used. Consistent with the data in Chapter 4, ^{f}UBR1 bound to GST-CUP9 with high affinity in the presence of both type 1 and type 2 dipeptides, but with much lower affinity in their absence (Figure 5.1.1c, lanes 2-3), while UBR1$^{1\text{-}1140f}$ bound to CUP9 constitutively (Figure 5.1.1b, lanes 2).

GST-CUP9$^{1\text{-}221}$, the GST fusion of the homeodomain-containing N-terminal fragment of CUP9, did not bind to UBR1$^{1\text{-}1140f}$ (Figure 5.1.1b, lane 3), or to ^{f}UBR1 in the presence or absence of type 1 and type 2 dipeptides (Figure 5.1.1c, lanes 4-5). In contrast, GST-CUP9$^{222\text{-}306}$ and GST-CUP9$^{256\text{-}306}$ bound tightly to UBR1$^{1\text{-}1140f}$ (Figure 5.1.1b, lanes 4-5), and to ^{f}UBR1 in the presence of both type 1 and type 2 dipeptides (Figure 5.1.1c, lanes 6-7), but not in their absence. These results indicated that the homeodomain-containing N-terminal 221-residue fragment of CUP9 was not required for its interaction with UBR1,

and that the C-terminal 51-residue fragment of CUP9, from residue 256 to the C-terminus, contained most, if not all, of the binding surface of CUP9 that was recognized specifically by UBR1. GST-CUP9$^{274-306}$ and GST-CUP9$^{287-306}$, the fusions of GST and the most C-terminal 33-residue and 20-residue fragments of CUP9, bound to UBR1$^{1-1140f}$ with reduced affinity (Figure 5.1.1b, lanes 6-7); their interaction with full-length ^{f}UBR1 were very weak even in the presence of type 1 and type 2 dipeptides (Figure 5.1.1c, lanes 8-9). These results, combined with the earlier data that all of the single-residue substitutions that stabilized CUP9 *in vivo* were located within its C-terminal 33 residues, strongly suggested that the C-terminal region of CUP9 encompassing its most C-terminal 33 residues specifically interacted with UBR1.

What structural features might be possessed by the C-terminal fragment of CUP9 encompassing its most C-terminal 33 residues? The C-terminal region of CUP9, starting from residues 272-274 to residues 300-303, was predicted, through several secondary-structure-prediction programs, to form two α helices separated by about four residues including Pro283, a known α helix breaker (Figure 5.1.2a). Helical wheel representations of the two α helices (Figure 5.1.2b and Figure 5.1.2c) revealed that the 16-residue α helix, from residues 286 to 300, formed an amphipathic helix— hydrophobic residues concentrate on one side and charged residues on another side (Figure 5.1.2c). Notably, three single-residue substitutions that stabilized CUP9 *in vivo* (L294P, L294V and R287H) were located on the hydrophobic surface of the amphipathic helix. In

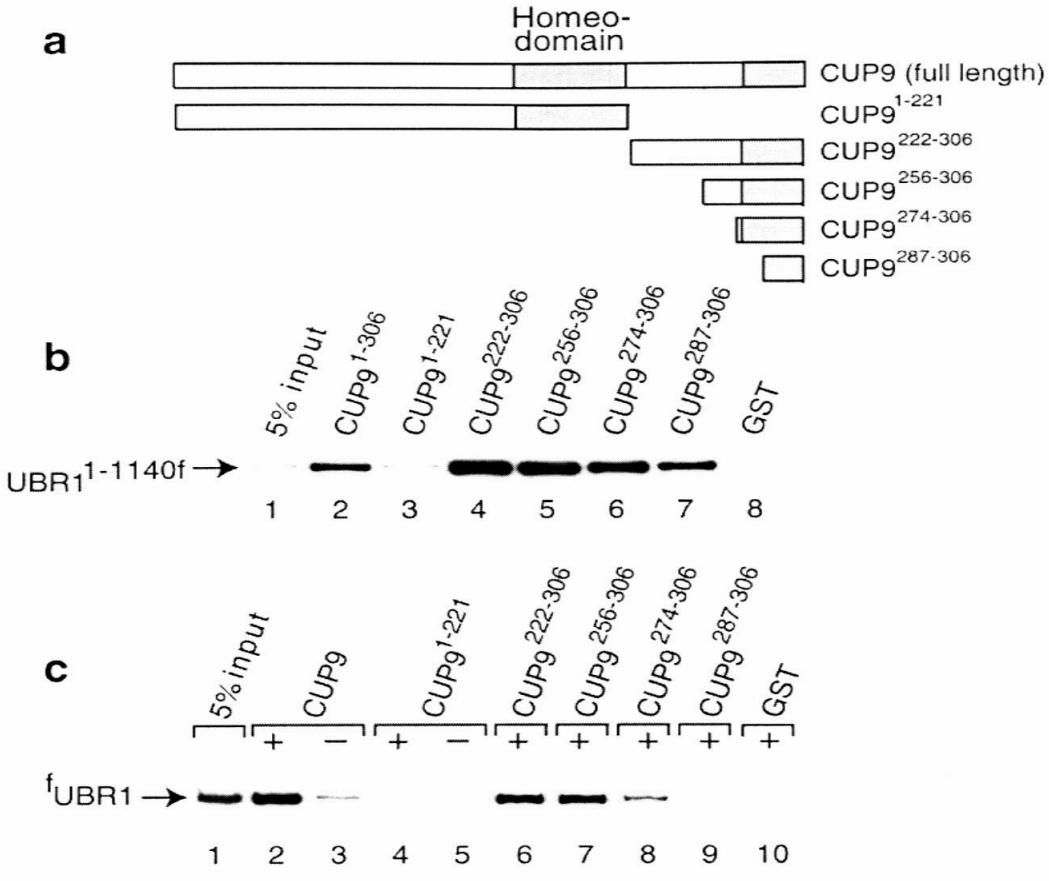


Figure 5.1.1 The C-terminal fragment of CUP9 encompassing its most C-terminal 33 residues mediates the interaction of CUP9 to UBR1. **a**, A diagram of CUP9, indicating its homeodomain and the most C-terminal 33-residue fragment. The CUP9 fragments shown (each of the fragments was also tagged with C-terminal hexahistidine) were fused in-frame to the C-terminus of GST. The fusion proteins were expressed in bacteria, purified through Ni-NTA affinity chromatography, and used in GST-pulldown assays. **b** and **c,** GST-pulldown assays with UBR1$^{1\text{-}1140f}$ **(b)** or ^{f}UBR1 **(c).** No dipeptides were present in **b.** The assay in **c** contained 1 mM Arg-Ala and 1 mM Leu-Ala (denoted as "+"above the lanes), or 1 mM Ala-Arg and 1 mM Ala-Leu (denoted as "-"above the lanes).

particular, Leu294 was central to the amphipathic helix, and its substitution with Pro, a known α helix breaker, was predicted to destroy the amphipathic helix. Indeed, CUP9(L294P), a CUP9 variant with its Leu294 mutated to Pro, was found to be the most stabilized CUP9 mutant *in vivo* (F. Navarro-Garcia, G. C. Turner and A. Varshavsky, personal communication). One mutation (E274K) was located in the 9-residue α helix that precedes the amphipathic helix. These two close α helices may stabilize each other through specific interactions.

CUP9 may be recognized by UBR1 mainly through the hydrophobic surface of its C-terminal 16-residue amphipathic helix, similar to MATα2, another homeodomain protein that was found to be targeted through its amphipathic helix [3]. In the future, more point mutations in the predicted amphipathic helix will be required to verify the above model. For example, mutations of Leu290 or Leu297 (both located on the hydrophobic surface) to Pro or charged residues are predicted to destroy the interaction of CUP9 to UBR1. Consequently, CUP9 variants harboring these mutations are expected to be stabilized significantly in the wild-type yeast cells. Obviously, cocrystal structures of UBR1 and CUP9, or their interacting domains, will be required to understand, at atomic resolution, the mechanism of CUP9 targeting by UBR1.

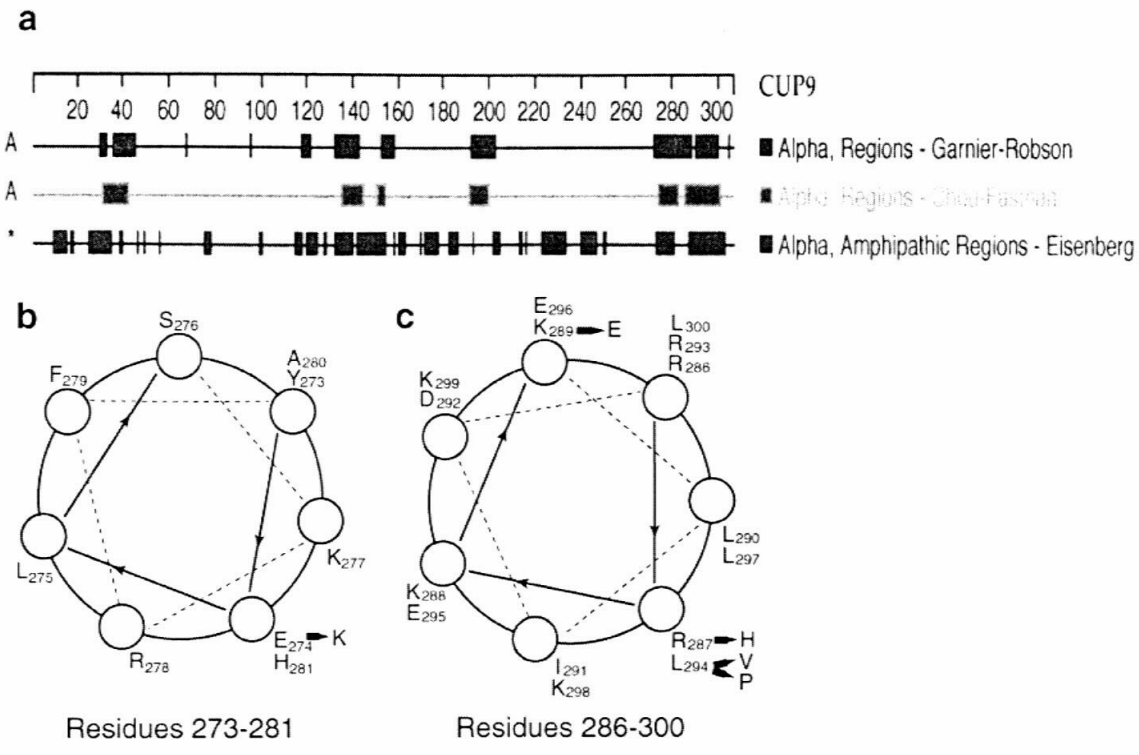


Figure 5.1.2 The α helices in the C-terminal region of CUP9. **a**, The C-terminal region of CUP9, from residues 272-274 to residues 300-303, was predicted through three secondary-structure-prediction programs to form two α helices. This figure was generated using DNASTAR package (DNASTAR, Madison, WI). **b**, Helical wheel representation of the predicted α helix between residues 273-281 of CUP9. A point mutation (E274K) found earlier to stabilize CUP9 in the wild-type yeast cells is denoted. **c**, The amphipathic helix between residues 286-300 of CUP9 revealed through helical wheel representation. Four point mutations that stabilized CUP9 in the wild-type yeast cells are indicated. Note that the right side of the helical wheel forms hydrophobic surface while the left side is filled with charged residues.

References

1. Byrd, C., G.C. Turner, and A. Varshavsky, *The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor.* EMBO J., 1998. 17: p. 269-277.

2. Byrd, C., *Regulation of Peptide Transport*, in *Department of Biology*. 1998, Caltech: Pasadena, CA. p. 155.

3. Johnson, P.R., et al., *Degradation signal masking by heterodimerization of MATalpha2 and MATa1 blocks their mutual destruction by the ubiquitin-proteasome pathway.* Cell, 1998. 94(2): p. 217-27.

5.2 Interaction of the type 1 and type 2 sites of UBR1 with dipeptides or other derivatives of amino acids

Type 1 and type 2 dipeptides and other derivatives of amino acids, such as methyl esters of Arg, Lys and Leu, were found to competitively inhibit the ubiquitylation and subsequent degradation of N-end rule substrates bearing the same type of destabilizing residues, both in cellular extracts [1] [2] [3] and in living cells [4]. More recently, binding of the type 1 or type 2 site of UBR1 by type 1 or type 2 dipeptides was found to activate UBR1 by increasing its affinity for CUP9, an effect mediated, apparently, through an increase in proportion of UBR1 molecules with "open" conformation (Chapters 3 and 4). In the experiments described below, several dipeptides and derivatives of amino acids were paired with either Arg-Ala or Leu-Ala, and assayed for their relative potency of enhancing the interaction of ^{f}UBR1 and GST-CUP9, using GST-pulldown assays. Dipeptides with N-terminal Arg and other derivatives of Arg were also tested for their capacity to activate the UBR1-dependent positive feedback circuit governing peptide import in yeast, using colony formation assays.

In GST-pulldown assays, the binding of ^{f}UBR1 to GST-CUP9 was very weak in the presence of dipeptides with stabilizing N-terminal residues (Figure 5.2a, lane 3; Figure 5.2b, lane 2); it was increased slightly in the presence of type 1, but not type 2 dipeptide alone (Figure 5.2a, lane 10; Figure 5.2b, lane 10), or the free amino acid constituents of dipeptides (Figure 5.2a, lane 5 versus lane 10; Figure 5.2b, lane 4 versus lane 10). Remarkably, the binding of ^{f}UBR1 to GST-CUP9 was increased dramatically in the presence of both type 1 and type 2

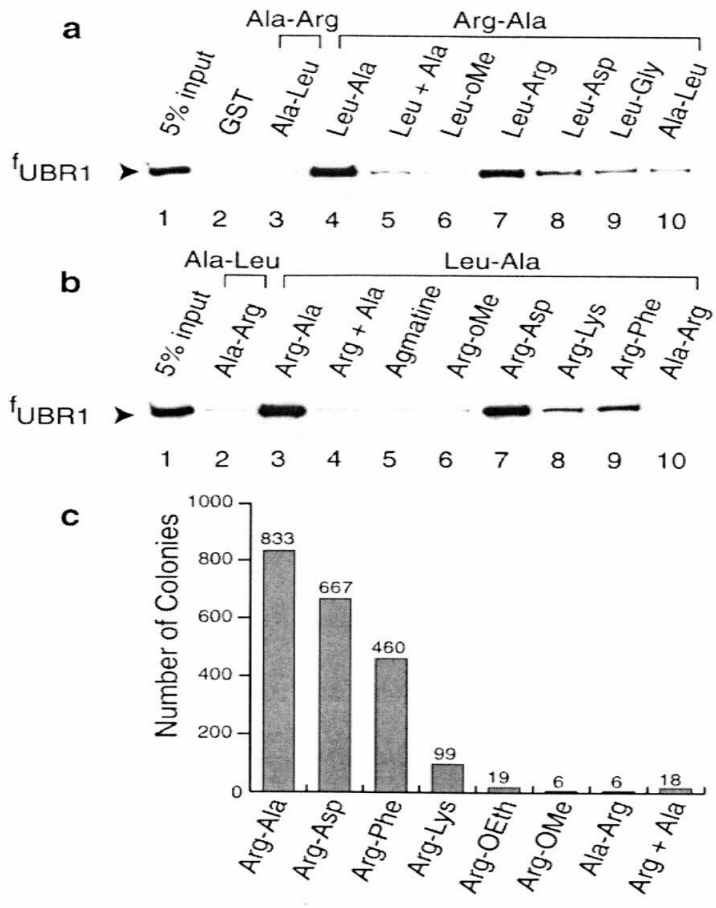


Figure 5.2 The relative potency of dipeptides or derivatives of amino acids in activating UBR1. **a** and **b**, GST-pulldown assays (Chapter 4) with GST-CUP9 and ^{f}UBR1 in the presence of pairs of dipeptides, or pairs of dipeptides and derivatives of amino acids. Each of the dipeptides, other derivatives of amino acids, or free amino acids were present at 1 mM. **c**, Colony formation assays (Chapter 3). The same number of RJD350 (*Matα leu2-3, 112*) yeast cells requiring Leu for growth was spread onto plates containing Ala-Leu (0.23 mM), and the indicated compound (10 μM). The number of yeast colonies, shown above each bar, was an average from two plates. Yeast colonies were counted after the plates were incubated at 30°C for ~80 hr.

dipeptides (Figure 5.2a lane 4; Figure 5.2b lane 3). These results were consistent with the earlier data (Chapter 4).

Dipeptides with the same N-terminal but different C-terminal residues varied widely in their potency in enhancing the binding of ^{f}UBR1 to GST-CUP9. In one set of assays (Figure 5.2a), each of the four dipeptides with an N-terminal Leu residue, as well as Leu methyl ester, was paired with the type 1 dipeptide Arg-Ala, and assayed for their potency in enhancing the interaction of ^{f}UBR1 to GST-CUP9. The dipeptide Leu-Ala was the most potent (Figure 5.2a, lane 4), and Leu-Arg was slightly less potent (Figure 5.2a, lane 7). In contrast, Leu-Asp, Leu-Gly and Leu methyl ester were nearly ineffective (Figure 5.2a, lanes 6, 8, 9). In another set of assays, the dipeptide Leu-Ala was kept constant in the pairs (Figure 5.2b, lanes 3-10). Arg-Ala was found to be the most potent (Figure 5.2b, lane 3), and Arg-Asp was slightly less potent (Figure 5.2b, lane 7), while both Arg-Phe and Arg-Lys were far less potent in enhancing the interaction of ^{f}UBR1 to GST-CUP9 (Figure 5.2b, lanes 8, 9). In contrast, both Arg methyl ester and agmatine (decarboxylated Arg) did not have any effect (Figure 5.2b, lanes 5, 6). Among all the pairs of type 1 and type 2 dipeptides tested, the combination of Arg-Ala and Leu-Ala was the most potent. Neither the pair of Arg methyl ester and Leu-Ala, nor the pair of Leu methyl ester and Arg-Ala, enhanced the interaction of ^{f}UBR1 to GST-CUP9. These observations, coupled with earlier findings that both Arg methyl ester and Leu methyl ester competitively inhibited the degradation of type 1 or type 2 N-end rule substrates respectively, indicated that binding of derivatives of amino acids to the type 1 or type 2 sites of UBR1

was necessary, but not sufficient to enhance the interaction of ^{f}UBR1 with GST-CUP9. It is possible that the methyl group in methyl esters of amino acids is not bulky enough to "open" the UBR1 protein. Yet another, mutually non-exclusive possibility is that esters of amino acids lack the peptide bond present in a dipeptide, which may be important in causing the "opening" of UBR1 protein, perhaps because it forms hydrogen bonds with residues surrounding the type 1 and type 2 site of UBR1.

The relative potency of dipeptides to enhance the interaction of ^{f}UBR1 to GST-CUP9 may correlate with their affinity for the type 1 or type 2 sites of UBR1, and thus may reveal the molecular environment of the type 1 and type 2 sites of UBR1. Specifically, the relative potency of dipeptides with an N-terminal Leu residue (Figure 5.2a) suggested that the type 2 site of UBR1 bound most tightly to dipeptides with C-terminal hydrophobic residues, and weakly to dipeptides with C-terminal acidic residues. The relatively strong binding of Leu-Arg to the type 2 site might result from the partly hydrophobic side chain of the C-terminal Arg residue, rather than from its positive charge. One prediction, based on the above analysis, is that dipeptides with an N-terminal Leu residue and a C-terminal bulky hydrophobic residue bind to the type 2 site of UBR1 even more strongly than Leu-Ala. The type 1 site of UBR1, however, displayed different preferences for the C-terminal residues of dipeptides (Figure 5.2b). The relative potency of dipeptides with an N-terminal Arg residue suggested that the type 1 site of UBR1 bound tightly to type 1 dipeptides with a small hydrophobic or an acidic C-terminal residue, significantly less tightly to type 1 dipeptides with a bulky hydrophobic C-terminal residue. Interestingly, type 1 dipeptides with a strongly

basic C-terminal residue bound weakly to the type 1 site of UBR1, which may underlie the earlier finding that Lys residue was not tolerated as the penultimate residue of type 1 N-end rule substrates [5].

Additionally, colony formation assays [6] were employed to measure the relative potency of dipeptides with an N-terminal Arg residue, or derivatives of Arg, in activating the UBR1-dependent positive feedback circuit regulating the dipeptide import in yeast (Figure 5.2c). Dipeptides of different compositions were most likely imported into yeast cells with similar efficiency due to the promiscuity of the dipeptide transporter PTR2 [7]. The plates supplemented with Arg-Ala had the highest number of colonies, followed by the plates supplemented with Arg-Asp or Arg-Phe. However, the number of yeast colonies on the plates supplemented with Arg-Lys was only about 1/8 of that on the plates supplemented with Arg-Ala. In contrast, the plates supplemented with Arg methyl ester, Arg ethyl ester, Ala-Arg, or free amino acids Arg and Ala, contained very few yeast colonies. All plates also contained 230 μM Ala-Leu as the source of Leu. Notably, the order of relative potency for dipeptides with an N-terminal Arg residue observed in the *in vitro* GST-pulldown assays is the same as that in colony formation assays, suggesting that the potency of dipeptides to enhance the binding of ^{f}UBR1 to GST-CUP9 may underlie their capacity to activate the UBR1-dependent positive feedback circuit regulating the dipeptide import in yeast.

In summary, dipeptides with the same N-terminal residue but different C-terminal residues were found to vary widely in their potency in activating UBR1 through the *in vitro* GST-pulldown assays and the *in vivo* colony formation

assays, possibly because of their different affinity for the type 1 or type 2 sites of UBR1. These results may be employed to adjust the strength of N-degron by varying the identity of the second residue of N-end rule substrates. It was also demonstrated here that small compounds, such as Arg methyl ester and Leu methyl ester, could inhibit the degradation of N-end rule substrates bearing N-degron without, in the meantime, accelerating the degradation of CUP9. It was discovered previously that low concentration of type 1 or type 2 dipeptides accelerated the degradation of CUP9 without inhibiting the degradation of N-degron-bearing N-end rule substrates. Mechanistically similar but more metabolically stable small compounds are likely to be of value as reagents that activate the third (internal degron-binding) site of UBR1 without significant inhibition of UBR1 binding to proteins bearing N-degrons.

References

1. Reiss, Y., D. Kaim, and A. Hershko, *Specificity of binding of N-terminal residues of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites.* J. Biol. Chem., 1988. **263**: p. 2693-2698.
2. Gonda, D.K., et al., *Universality and structure of the N-end rule.* J. Biol. Chem., 1989. **264**: p. 16700-16712.
3. Davydov, I.V., D. Patra, and A. Varshavsky, *The N-end rule pathway in Xenopus egg extracts.* Arch. Biochem. Biophys., 1998. **357**: p. 317-325.
4. Baker, R.T. and A. Varshavsky, *Inhibition of the N-end rule pathway in living cells.* Proc. Natl. Acad. Sci. USA, 1991. **87**: p. 2374-2378.
5. Suzuki, T. and A. Varshavsky, *Degradation signals in the lysine-asparagine sequence space.* EMBO J., 1999. **18**(21): p. 6017-6026.
6. Turner, G.C., F. Du, and A. Varshavsky, *Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway.* Nature, 2000. **405**(6786): p. 579-583.
7. Perry, J.R., et al., *Isolation and characterization of a Saccharomyces cerevisiae peptide transport gene.* Mol. Cell. Biol, 1994. **14**(1): p. 104-15.

5.3 Cofractionation of Leu-ILV1$^{34-576}$ and TDH2/TDH3 with ^{f}UBR1^{h}

To identify the proteins that cofractionate with UBR1, ^{f}UBR1^{h}, a functionally active derivative of UBR1 bearing N-terminally FLAG tag and C-terminal hexahistidine tag, was overexpressed in *S. cerevisiae* from the P_{ADH1} promoter in a high-copy vector, and purified, successively, through Ni-NTA and anti-FLAG affinity columns. Three cofractionated polypeptides, named P80, P54 and P33 for their apparent molecular weights in SDS-PAGE patterns (Figure 5.3a), were subjected to N-terminal sequencing by Edman degradation. P80 did not yield conclusive sequence for unknown reason — its N-terminal residue could be Pro, Lys or Leu. N-terminal sequencing of P33 yielded VRVAINGFGRIGR, which is identical to the N-terminal fragments of two almost identical (96% identity) yeast isozymes of glyceraldehyde-3-phosphate dehydrogenase encoded by *TDH2* and *TDH3* (Figure 5.3b). The initiator Met residues of TDH2 and TDH3 are predicted to be cleaved off, since Met is followed here by Val, a residue with a small radius of gyration [1]. The deduced molecular weight of the mature TDH2 or TDH3 is ~35.7 kDa, close to the 33 kDa that was observed (Figure 5.3a). TDH2 and TDH3 catalyze the reversible oxidation and phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphosphoglycerate in glycolysis. These two isozymes were found to cofractionate with quite a few yeast proteins, at least partly due to their abundance in yeast cells. It is unknown whether ^{f}UBR1^{h} directly binds to TDH2 or TDH3, and whether this interaction is physiologically relevant.

N-terminal sequencing of P54 yielded LHRQHLSP, which matches the 8-residue amino acid sequence, from residues 34 to 41, of the abundant serine

and threonine dehydratase encoded by *ILV1* (Figure 5.3b). In addition, the deduced molecular weight of ILV1 lacking its N-terminal 33 residues is ~60 kDa, which is not far from the 54 kDa observed through SDS-PAGE (Figure 5.3a). Thus, P54 is most likely the C-terminal, 543 residues long (from Leu34 to the C-terminus) fragment of ILV1, denoted hereafter as Leu-ILV1$^{34\text{-}576}$. ILV1 catalyzes the conversion of threonine to α-ketobutyrate, the first step in the isoleucine biosynthesis pathway. It was reported to be a tetrameric enzyme whose activity was inhibited by free Ile through allosteric feedback control [2]. Yeast strain lacking *ILV1* is auxotrophic for Ile, a phenotype that can be suppressed with either Thr or Ser as the sole nitrogen source [3]. ILV1 is believed to be localized in the matrix of mitochondria, based primarily on the biochemical fractionation experiments [4]. The N-terminal sequence of unprocessed ILV1 is characteristic of a mitochondrial translocation signal [5]. Specifically, among the N-terminal 33 residues of ILV1, there are six strongly basic residues (Arg and Lys), five hydroxylated residues (Ser and Thr), and no negatively charged residues at all. However, the mitochondrial translocation signal of ILV1 has not been experimentally characterized. There have been conflicting reports about whether ILV1 is cleaved after it is imported into mitochondria [6]. So far, it has not been possible, in most cases, to predict reliably whether the mitochondrial translocation signal of a mitochondrial protein is cleaved in the matrix of mitochondria. Leu-ILV1$^{34\text{-}576}$, which lacks the N-terminal 33 residues of ILV1 and exposes Leu34 as its new N-terminal residue, is most probably the mature form of ILV1 in mitochondria. It is unlikely that Leu-ILV1$^{34\text{-}576}$ binds to ^{f}UBR1^{h} in living cells since these two proteins are localized in different cellular

compartments: Leu-ILV1$^{34-576}$ in mitochondria and ^{f}UBR1^{h} in the cytosol (possibly in the nucleus too). In yeast extracts, where proteins from different compartments become intermixed, Leu-ILV1$^{34-576}$, most likely, bound tightly to the type 2 site of ^{f}UBR1^{h} through its type 2 destabilizing N-terminal Leu residue, resulting in their cofractionation.

To determine whether Leu-ILV1$^{34-576}$ is targeted through the N-end rule pathway if it is present in the cytosol, we constructed Ub-Leu-ILV1$^{34-576}$, Ub-Val-ILV1$^{34-576}$ and Ub-Met-ILV1$^{34-576}$ using the Ub fusion technique to vary the N-terminal residue of ILV1 *in vivo*. These proteins were expressed, in both *UBR1* and *ubr1Δ* yeast cells from the P_{CUP1} promoter in a high-copy vector. Val-ILV1$^{34-576}$ and Met-ILV1$^{34-576}$ are identical to Leu-ILV1$^{34-576}$ except that their N-terminal residues (Val and Met) are stabilizing in the N-end rule. The expression of Val-ILV1$^{34-576}$ or Met-ILV1$^{34-576}$ in *UBR1* or *ubr1Δ* yeast cells did not inhibit their growth significantly (Figure 5.3c). Surprisingly, however, the expression of Leu-ILV1$^{34-576}$ was lethal to *UBR1* cells, but not to *ubr1Δ* cells (Figure 5.3c). Moreover, the lethality was partially suppressed by the presence of a mixture of amino acids containing Ile, Arg, Lys, Met, Phe and Thr (denoted as aa-4 in Figure 5.3c) in the medium. For reasons that remain unclear, the finding of lethality of Leu-ILV1$^{34-576}$ expression in UBR1 cells could not be reproduced in later experiments.

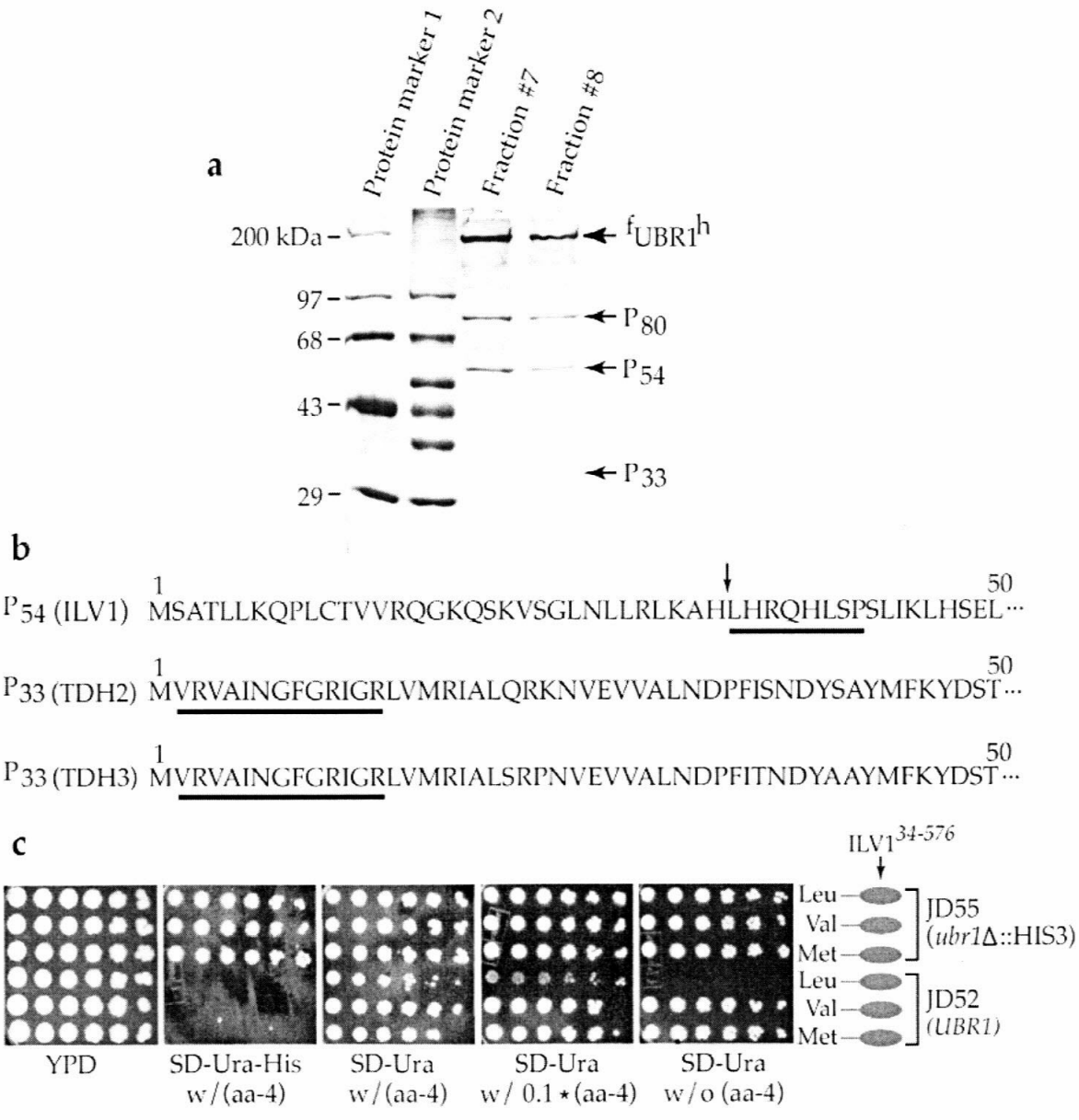

a
Protein marker 1
Protein marker 2
Fraction #7
Fraction #8
200 kDa
97
68
43
29
fUBR1h
P80
P54
P33
b
P54 (ILV1) 1 MSATLLKQPLCTVVRQGKQSKVSGLNLLRLKAHLHRQHLSPSLIKLHSEL 50 ···
P33 (TDH2) 1 MVRVAINGFGRIGRLVMRIALQRKNVEVVALNDPFISNDYSAYMFKYDST 50 ···
P33 (TDH3) 1 MVRVAINGFGRIGRLVMRIALSRPNVEVVALNDPFITNDYAAYMFKYDST 50 ···
c
ILV1 34-576
Leu
Val
Met
JD55
(ubr1Δ::HIS3)
Leu
Val
Met
JD52
(UBR1)
YPD
SD-Ura-His w/(aa-4)
SD-Ura w/(aa-4)
SD-Ura w/ 0.1 ⋆(aa-4)
SD-Ura w/o (aa-4)

Figure 5.3 **a**, Cofractionation of P80, P54 and P33 with $^{f}UBR1^{h}$. $^{f}UBR1^{h}$ was fractionated successively though Ni-NTA and anti-FLAG M2 affinity columns. Proteins in two fractions from the anti-FLAG M2 affinity column were further separated on SDS-8% PAGE. The gel was stained with Coomassie. The positions of the three cofractionated polypeptides are indicated with arrows. **b**, The identities of P54 and P33. The N-terminal 50-residue fragments of ILV1, TDH2 and TDH3 are displayed, and the sequences obtained through Edman degradation are underlined. The arrow between the residues 33 and 34 of ILV1 indicates the possible cleavage site of ILV1 in the matrix of mitochondria. **c**, Expression of Leu-ILV1$^{34-576}$ in the cytosol was lethal to *UBR1* cells, but not to *ubr1Δ* yeast cells. Serial dilutions of *UBR1* and *ubr1Δ S. cerevisiae* cells expressing Leu-ILV1$^{34-576}$, Val-ILV1$^{34-576}$ or Met-ILV1$^{34-576}$ were deposited onto SD plates supplemented with different nutrients. The SD-Ura plates contain 0.67% yeast nitrogen base without amino acids (Difco), 2% glucose and auxotrophic nutrients, including Trp, Leu, His and Lys at concentrations specified by Sherman *et al.* [7]. The SD-Ura-His plate is the same as SD-Ura plate except for the absence of His. The aa-4 refers to a mixture of amino acids, including Ile, Arg, Lys, Met, Phe and Thr [7] while 0.1*(aa-4) indicates 1/10 concentration of amino acids present in aa-4. The plates were incubated at 30°C for ~80 hr.

References

1. Bradshaw, R.A., W.W. Brickey, and K.W. Walker, *N-terminal processing: the methionine aminopeptidase and N alpha-acetyl transferase families.* Trends Biochem. Sci., 1998. **23**(7): p. 263-7.
2. Robichon-Szulmajster, H. and P.T. Magee, *The regulation of isoleucine-valine biosynthesis in Saccharomyces cerevisiae. I. Threonine deaminase.* Eur. J. Biochem., 1968. **3**(4): p. 492-501.
3. Ramos, F. and J.M. Wiame, *Occurrence of a catabolic L-serine (L-threonine) deaminase in Saccharomyces cerevisiae.* Eur. J. Biochem., 1982. **123**(3): p. 571-6.
4. Ryan, E.D. and G.B. Kohlhaw, *Subcellular localization of isoleucine-valine biosynthetic enzymes in yeast.* J. Bacteriol., 1974. **120**(2): p. 631-7.
5. Neupert, W., *Protein import into mitochondria.* Annu. Rev. Biochem., 1997. **66**: p. 863-917.
6. Petersen, J.G., et al., *Molecular genetics of serine and threonine catabolism in Saccharomyces cerevisiae.* Genetics, 1988. **119**(3): p. 527-34.
7. Sherman, F., *Getting started with yeast.* Meth. Enzymol., 1991. **194**: p. 3-21.

5.4 Purification of the recombinant *S. cerevisiae* NTA1 and its enzymatic activity assay

Introduction

The *S. cerevisiae* N-terminal amidase (Nt-amidase) encoded by *NTA1* is a component of the yeast N-end rule pathway. Nt-amidase converts the tertiary destabilizing N-terminal Asn and Gln residues to Asp and Glu, respectively [1]. The N-terminal amidases from the yeast *S. cerevisiae* and *S. pombe* constitute one family of the nitrilase superfamily [2] (Chapter 1). Previously, Nt^N-amidase (NTAN1), the N-terminal amidase that is specific for N-terminal Asn, but not Gln, was purified to homogeneity from porcine liver [3]. However, the amino acid sequences of porcine or murine NTAN1, which are 88% identical to each other, are not similar to *S. cerevisiae* NTA1 [4]. Moreover, mammalian Nt^N-amidase does not appear to be a member of the nitrilase superfamily (Chapter 1), suggesting that *S. cerevisiae* NTA1 and mammalian NTAN1 may employ different mechanisms to catalyze deamidation reaction.

Materials and methods

Construction of pET-11c-NTA1

The plasmid pET-11c-NTA1 was constructed by inserting the *NTA1* ORF between the unique *Nde I* and *BamH I* restriction sites of pET-11c (Novagen, Madison, WI). The *NTA1* ORF, starting from the ATG(+1) and thus containing the putative mitochondrial translocation signal (Chapter 1), was amplified by PCR using the primers NTA1PC3 (5'- CCAGTCGACATATGCTAATAGACGCA

ATTC-3′) and NTA1PC4 (5′-CCAGGATCCTCACCTAAACACTTCAAA-3′). The *NTA1* ORF was verified by DNA sequencing. NTA1 was untagged since extensions at its N-terminus or C-terminus were shown to reduce its *in vivo* activity [1] (F. Du and A. Varshavsky, data not shown).

Overexpression and purification of NTA1

A fresh colony of BL21(DE3) strain carrying pET-11c-NTA1 was dispersed into 1 ml of LB medium, then diluted 100 fold with fresh LB medium. 150 μl of diluted culture was spread onto an LB plate containing ampicillin (60 μg/ml). After incubation at 37°C for ~18 hr, bacterial colonies on the plate were scraped and inoculated into 1 l of ampicillin-containing LB medium. The culture was grown at 37°C to A_{600} ~1.0, cooled down to 23°C and IPTG (Calbiochem, San Diego, CA) was added to a final concentration of 1 mM, followed by an additional 5-hr incubation at 23°C. The cells were harvested by centrifugation, washed with PBS, and frozen in liquid N_2.

The cell pellet was resuspended (6 ml of buffer per 1g of pellet) in buffer A (10% glycerol, 50 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 mM Tris-HCl (pH 7.6) containing protease inhibitors (50 μM AEBSF, 1 mM pepstatin). Freshly dissolved chicken egg white lysozyme in the same buffer was added to a final concentration of 0.5 mg/ml, followed by incubation at 4°C for 30 min, and the cells were further disrupted by sonication. The suspension was clarified by centrifugation at 11,200*g* for 30 min, and the supernatant was filtered through 0.8μm filter. The protein concentration in the filtrate was determined through the Bradford assay with bovine serum albumin (BSA) as standard, and adjusted

to ~1 mg/ml with buffer A. The filtrate was then loaded onto a DEAE column equilibrated in buffer A. The DEAE column was developed with 20-column volumes linear gradient of NaCl (50-300 mM) in 10% glycerol, 1 mM DTT, 1 mM EDTA, 50 mM Tris-HCl (pH 7.6). NTA1 eluted at ~ 110 mM (Figure 5.4.1b, lane 4). The NTA1- containing fractions were pooled and dialyzed into buffer C (10% glycerol, 1 mM DTT, 0.2 mM EDTA, 10 mM K_2HPO_4-KH_2PO_4 (pH 6.0)), followed by passage through S-Sepharose column equilibrated in buffer C. NTA1 was present in the flow-through (Figure 5.4.1b, lane 5). The NTA1-containing flow-through was adjusted to pH 6.8 with 10 mM K_2HPO_4 and loaded onto hydroxyapatite column equilibrated in buffer E (10% glycerol, 1 mM DTT, 0.2 mM EDTA, 10 mM K_2HPO_4-KH_2PO_4 (pH 6.8)). The hydroxyapatite column was developed with 6-column volumes linear gradient of phosphate buffer (10 –200 mM, pH 6.8) in 10% glycerol, 1 mM DTT, 0.2 mM EDTA. NTA1 eluted at ~80 mM (Figure 5.4.1b, lane 6). The NTA1-containing fractions were pooled and dialyzed into buffer G (10% glycerol, 50 mM NaCl, 1 mM DTT, 0.2 mM EDTA, 20 mM bis-Tris. (pH 6.0)), and loaded onto Mono-Q column that was equilibrated in buffer G. The Mono-Q column was developed with 5-column volumes linear gradient of NaCl (50-500 mM) in 10% glycerol, 1 mM DTT, 0.2 mM EDTA, 20 mM bis-Tris (pH 6.0). NTA1 eluted at ~270 mM (Figure 5.4.1b, lane 7). The NTA1-containing fractions were pooled, concentrated and dialyzed into buffer K (20% glycerol, 100 mM KCl, 1 mM DTT, 0.1 mM EDTA, 20 mM HEPES (pH 7.6)), and fractionated through Superdex 75 gel filtration column equilibrated in buffer K. The purified NTA1 protein (Figure 5.4.1c) was concentrated to 11.5 mg/ml in

buffer K, frozen in liquid N_2 and stored at -80°C. The typical yield was ~1 mg pure NTA1 protein per 1 l of bacterial culture.

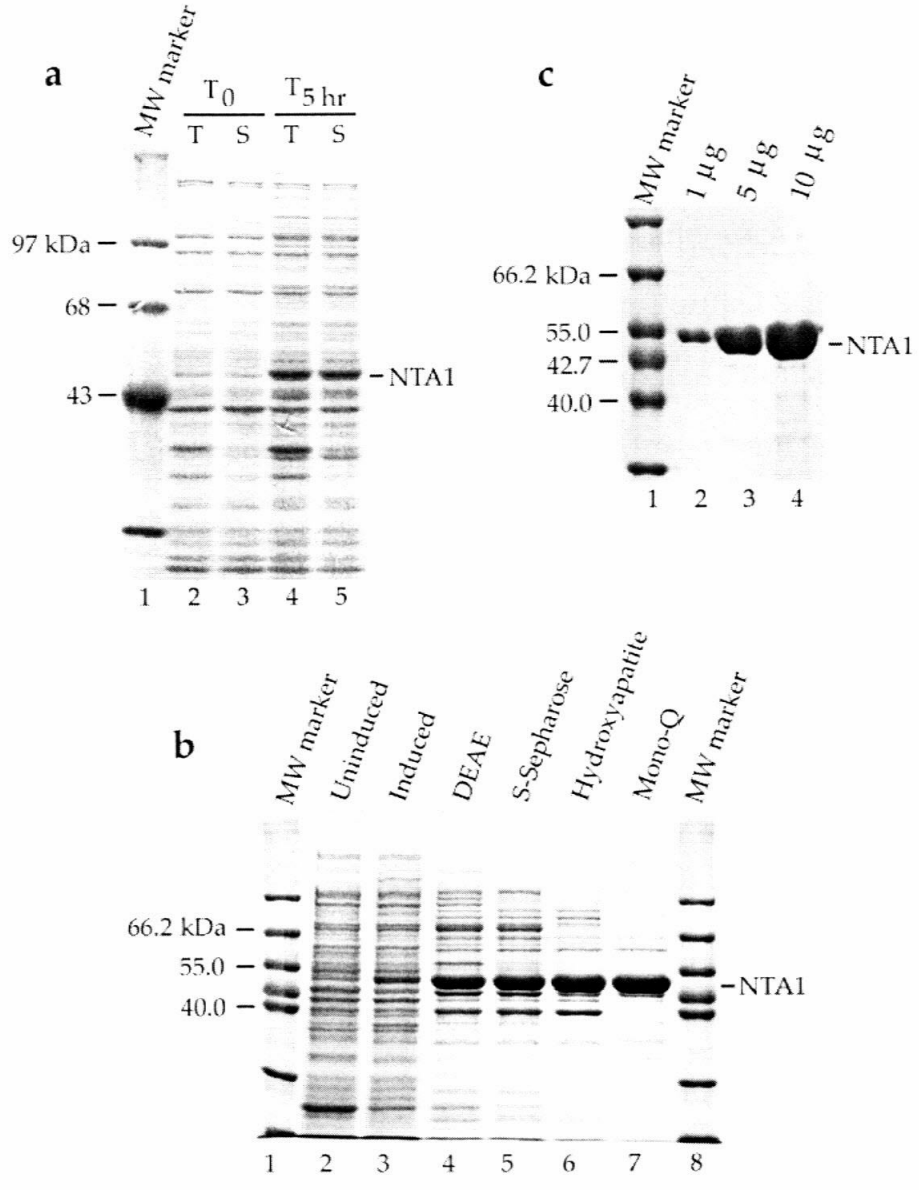


Figure 5.4.1 Purification of *S. cerevisiae* NTA1. **a**, Overexpression of NTA1 in bacteria. Bacterial cells in 2 ml of culture (A_{600} ~1.0) before the addition of IPTG (denoted as "T_0" above the lanes 2-3) and 0.8 ml of culture (A_{600} ~2.5) 5 hr after

the addition of IPTG to 1 mM (denoted as "T_{5hr}" above the lanes 4-5) were collected. Cellular proteins were extracted, either by boiling the cells in 300 µl of SDS-PAGE loading buffer for 5 min (denoted as "T"above lanes 2,4), or by solubilizing proteins in 300 µl of buffer S (150 mM NaCl, 2 mM EDTA, 25 mM HEPES (pH 7.6)) (denoted as "S" above lanes 3, 5). 20 µl of samples were fractionated by SDS-10% PAGE. **b**, Purification of *S. cerevisiae* NTA1. 10 µl of samples from pooled NTA1-containing fractions after each chromatographic column were fractionated by SDS-10% PAGE. **c**, Purified *S. cerevisiae* NTA1. 1 µg, 5 µg and 10 µg of NTA1 fractionated through five chromatographic columns, namely, DEAE, S-Sepharose, Hydroxyapatite, Mono-Q and Superdex 75 gel filtration columns, were displayed on SDS-10% PAGE. The gels were stained with Coomassie, and the positions of NTA1 were indicated.

Activity assay of NTA1

The Asn-peptide (NH_2-Asn-His-Gly-Ser-Gly-Ala-Trp-Leu-COOH) and Asp-peptide (the same as the Asn-peptide except for its N-terminal Asp residue) were synthesized and verified by mass spectrometry (Caltech). The peptide sequences were derived from the N-terminal e^k region of engineered N-end rule substrates. Both peptides were dissolved in H_2O, and stored at -80°C as 10 mM stock solution.

Capillary electrophoresis (CE) was carried out at 20°C in phosphate running buffer (20 mM H_3PO_4-NaH_2PO_4 (pH 2.5)), using the 270 A Capillary Electrophoresis System (Perkin Elmer, Boston, MA). The voltage applied was 30

kV, and one run of CE took 10 min. For each run, less than 0.1 μl of sample was consumed. Peptides were detected through their absorption at 220 nm (A_{220}).

Enzymatic activity of NTA1 was assayed by generation of Asp-peptide through deamidation of the N-terminal Asn residue of Asn-peptide. The premix of NTA1 deamidation reactions contained 0.6 mM of the Asn-peptide in buffer R (10% glycerol, 100 mM NaCl, 1 mM DTT, 0.1 mg/ml acetylated BSA, 20 mM HEPES (pH 7.6)). Acetylated BSA was added as a carrier protein. 5 μl of NTA1-containing bacterial extract or chromatographic fractions were added into 45 μl of the premix, and the reactions were shifted to 37°C. For each reaction, 10 μl of the reaction mix was withdrawn at 0, 10, 20 and 40 min, respectively, heated at 95°C for 4 min to inactivate NTA1, and diluted with 10 μl of H_2O before it was subjected to CE (Figure 5.4.2). The peak areas corresponding to Asn-peptide and Asp-peptide were quantitated.

RESULTS AND DISCUSSIONS

The untagged *S. cerevisiae* NTA1 was overexpressed in *E. coli* and purified to more than 95% homogeneity through five chromatographic steps, specifically, DEAE, S-Sepharose, Hydroxyapatite, Mono-Q, and the Superdex 75 gel filtration columns (Figures 5.4.1b and 5.4.1c). The identity of NTA1 was verified by N-terminal sequencing. As expected, the initiator Met residue of NTA1 was retained.

NTA1 activity was assayed with a peptide bearing N-terminal Asn as its substrate, the same peptide that was used as a substrate for porcine NTAN1 [3]. The N-terminal Asn residue of the Asn-peptide was converted to Asp through

NTA1-catalyzed deamidation, and the resulting Asp-peptide was well-separated from the substrate Asn-peptide using CE: the Asn-peptide was detected ~3.8 min while the Asp-peptide was detected ~5.2 min through A_{220} after the start of electrophoresis (Figure 5.4.2). Detection of peptides through UV absorption limited the sensitivity of this assay (the lower limit is ~10 ng of NTA1), which can be increased at least 1,000 fold if a peptide substrate is fluorescently labeled, and the reactions are monitored by CE with laser-induced fluorescence.

One unit of NTA1 activity was operationally defined as the amount of enzyme that deamidated 1 nmol of N-terminal Asn residue of Asn-peptide at 37°C in 20 min. This definition is the same as that adopted for porcine NTAN1 [3]. The specific activity of purified yeast NTA1 was ~120,000 units/mg, which was significantly higher than that of purified porcine NTAN1 (23,000 units/mg). The purified NTA1 remained stable at 4°C for several weeks, and was able to withstand at least three rounds of freezing and thawing without appreciable loss of enzymatic activity; in contrast, the enzymatic activity of porcine NTAN1 was significantly reduced after freezing and thawing [3].

A total of 14 mg of purified *S. cerevisiae* NTA1 (11.5 mg/ml) was supplied to Pavel Strop (Caltech) for crystallization. NTA1 does not form well-ordered, large crystals suitable for structural analysis by X-ray crystallography, although it forms small crystals under a variety of conditions (P. Strop and D. C. Rees, personal communication).

In summary, we described the overexpression of *S. cerevisiae* NTA1 in *E. coli* and a five-step procedure for its purification. The enzymatic activity of NTA1 was monitored through deamidation of the N-terminal Asn residue of a

peptide substrate (Asn-peptide), and the product Asp-peptide was separated from the substrate Asn-peptide by capillary electrophoresis (CE). This facile and sensitive assay for amidase activity has not been described previously. Purified *S. cerevisiae* NTA1 and the procedure developed for its purification should facilitate both enzymological and structural studies of this enzyme.

References

1. Baker, R.T. and A. Varshavsky, *Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway.* J. Biol. Chem., 1995. **270**(20): p. 12065-12074.
2. Pace, H.C. and C. Brenner, *The nitrilase superfamily: classification, structure and function.* Genome Biology, 2001. **2**(1): p. 1-9.
3. Stewart, A.E., S.M. Arfin, and R.A. Bradshaw, *Protein NH2-terminal asparagine deamidase. Isolation and characterization of a new enzyme.* J. Biol. Chem., 1994. **269**(38): p. 23509-17.
4. Grigoryev, S., et al., *A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway.* J. Biol. Chem., 1996. **271**: p. 28521-28532.

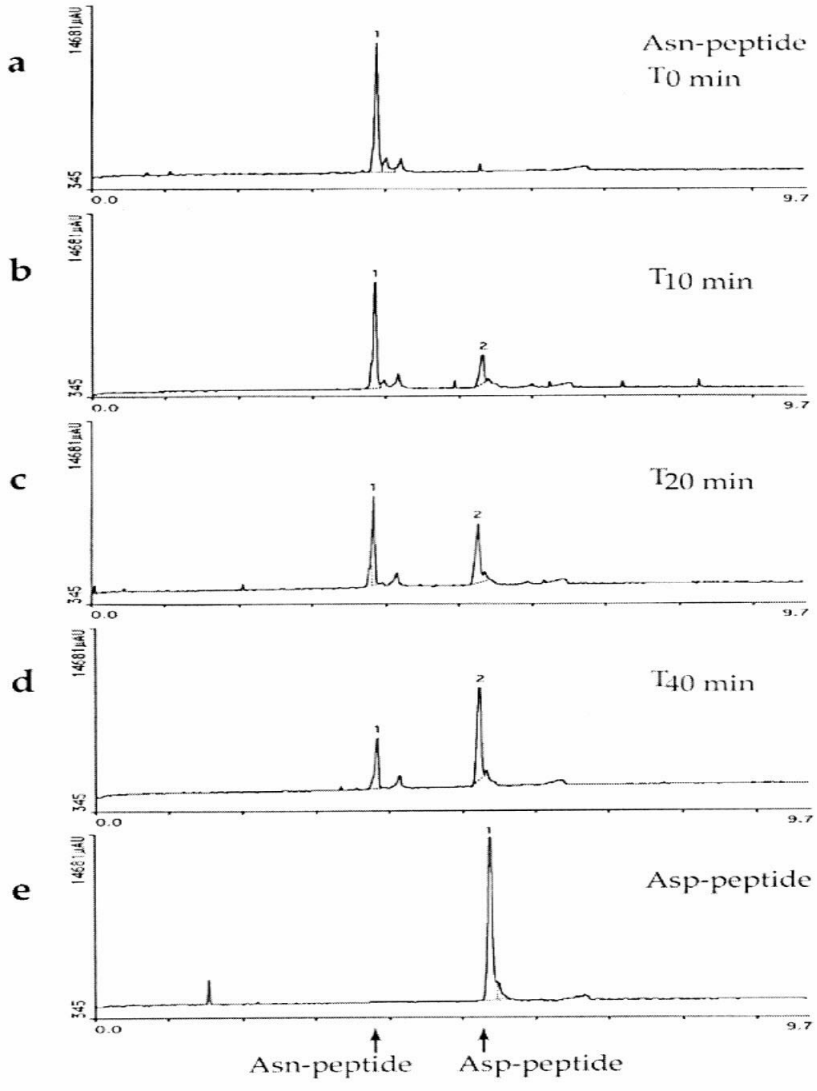


Figure 5.4.2 Separation of the Asn-peptide and Asp-peptide by CE in enzymatic activity assays of NTA1. **a-d**, Electropherogram of the reaction mix at 0, 10, 20 and 40 min. Added into 45 µl of premix containing the Asn-peptide (0.6 mM) in buffer R (see Materials and methods) was 5 µl of bacterial extract (containing ~3 µg of total proteins), and the reaction was shifted to 37°C. Withdrawn at each time point and subjected to CE was 10 µl of reaction mix. **e**, The electropherogram of the synthesized Asp-peptides. The peaks corresponding to the Asn-peptide (~3.8 min) and the Asp-peptide (5.2 min) were denoted.

5.5 Multicopy suppressors of ATE1 overexpression-mediated growth arrest of *S. cerevisiae* cells

Fangyong Du, Hong-Rui Wang and Alexander Varshavsky

Introduction

Arginyl-tRNA-protein transferase (R-transferase) transfers the activated Arg of Arg-tRNA^{Arg} to N-terminal Asp or Glu of polypeptides, and also to the oxidized N-terminal Cys residue (sulfinic acid residue and/or cysteic acid residue) (Chapter 1) [1] (Y. T. Kwon and A. Varshavsky, personal communication). R-transferase is a component of the N-end rule pathway, since arginylated N-end rule substrates are targeted by E3 of the N-end rule pathway for Ub-dependent degradation [1]. The sole R-transferase of the yeast *S. cerevisiae* is encoded by *ATE1*. Yeast mutants lacking *ATE1* appear normal, except for their inability to degrade engineered N-end rule substrates bearing tertiary and secondary destabilizing residues [1]. By contrast, overexpression of ATE1 in yeast cells was found to cause growth arrest [2]. Here we report the analysis of the molecular cause(s) of toxicity of overexpressed ATE1.

Materials and methods

Strains, media and β–galactosidase assay

The *S. cerevisiae* strains used in this work were JD52 (*MAT***a** *lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*); JD53 (*MATα lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-3,112*); JD55 (a *ubr1Δ::HIS3 derivative of JD52*); GMY1, a diploid strain derived from the cross between JD52 and JD53; JD54 (*a GAL::UBR1 derivative of JD52*);

SGY4 (*MATα lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 ade2-101 nta1Δ::LEU2*), and SGY3 (*MATα lys2-801 ura3-52 trp1-Δ63 his3-Δ200 leu2-Δ1 ade2-101 ate1Δ::LEU2*). Yeast strains were grown in synthetic media containing 0.67% yeast nitrogen base without amino acids (Difco), auxotrophic nutrients at the concentrations specified by Sherman et al. [3], and either 2% glucose (SD media) or 2% galactose (SGal media) as carbon source. The enzymatic activity of β–galactosidase (βgal) in yeast extract from culture of A_{600} ~1 was determined as described previously [4], using *o*-nitrophenyl-D-galactopyranoside (ONPG) as its substrate.

Construction of plasmids

The plasmids pYES2-ATE1, pYES2-ATE1C20S and pYES2-ATE1C23A were constructed to express, respectively, *S. cerevisiae* ATE1, or its derivatives ATE1C20S (Cys20→Ser) and ATE1C23A (Cys23→Ala) from the P_{GAL1} promoter in the episomal vector pYES2 (Invitrogen). The DNA fragments encoding ATE1, ATE1C20S or ATE1C23A were amplified by PCR using the primers ATE1PC9 (5′-GGCCAAGCTTATGTCCGATAGATTCGTTATT-3′) and ATE1PC4 (5′-CCA CTCGAGTCACATTTGCTCACTATA-3′), and were inserted between the unique *Hind III* and *Xho I* sites of pYES2. The plasmids containing the *ATE1C20S ORF* and *ATE1C23A ORF* were gifts from Dr. C. M. Pickart (Johns Hopkins Univ. Baltimore, MD). The plasmid pYES2-ATE1^f was constructed to express ATE1^f, a functionally active ATE1 derivative bearing C-terminal FLAG epitope tag. from the P_{GAL1} promoter in pYES2. The DNA fragment encoding ATE1^f was amplified by PCR from the plasmid pRS314$_{GAL1/10}$ATE1^f using the primers

ATE1PC9 and ATE1FLAG1 (5′-ACCACTCGAGTCACTTGTCATCGTCGTC-3′), and was inserted between the unique *Hind III* and *Xho I* sites of pYES2. The plasmids pRS314$_{GAL1/10}$ ATE1 and pRS314$_{GAL1/10}$ ATE1^{f} were constructed to express ATE1 and ATE1^{f} from the P_{GAL1} promoter in the centromeric vector pRS314 [5] (M. Ghislain and A. Varshavsky, unpublished data). The plasmids p413MET25-ATE1, p413MET25-ATE1C20S and p413MET25-ATE1C23A were constructed to express ATE1, ATE1C20S and ATE1C23A from the P_{MET25} promoter in the centromeric vector p413MET25 [6]. The DNA fragments encoding ATE1, ATE1C20S or ATE1C23A were amplified by PCR using the primers ATE1PC7 (5′-GGCTCTAGAATGTCCGATAGATTCGTT-3′) and ATE1PC4, and were inserted between the unique *Xba I* and *Xho I* restriction sites of p413MET25. The plasmid p413MET25-ATE1$^{1-412}$ was constructed similarly; it expressed the N-terminal 412-residue fragment of ATE1 (ATE1$^{1-412}$) from the P_{MET25} promoter in p413MET25. The plasmid pINP123-306 was derived from the integrating vector pRS306 [5]; it was constructed to facilitate the integration of YFG (your favorite gene) to the *LYS2* or *Lys2-801* locus of S, *cerevisiae*. Three DNA fragments of *LYS2 ORF* (Figure 5.5.4), termed INP1 (888 bp), INP2 (321 bp) and INP3 (358 bp), were amplified by PCR using the primer pairs IP1L (5′-CA GCCGATATCATTGTGCCCAATGGTAAATCAAC-3′) and IP1R (5′-CAGCC GAGCTCTGGTATTGGCGAAATAGGTGAGA-3′); IP2L(5′-CAGCCGGTACC TGATCCAATTGTCCAGCAGCTC -3′) and IP2R (5′-CAGCCCTCGAGGGCG CGCCAGAGAGAACCTGTGTTGT-3′); IP3L (5′-CAGCCCTCGAGGGCGCGCC TTCAGCATTGTCCTGGAAAATGTC-3′) and IP3R (5′-CAGCCGATATCGAAA GGACCCCTCAGTTGTTCC-3′). The three fragments were cloned into the

vector pBluescript II KS to generate pINP1-B2, pINP2-B2 and pINP3-B2. All three inserts were transferred to pRS306 to generate pINP123-306 through three additional intermediate plasmids pINP1-306, pINP23-B2 and pINP23-306. The plasmid pINP123-ATE1 was constructed by inserting the DNA fragment containing, from 5′ to 3′, the P_{GAL1} promoter, *ATE1 ORF*, and the *CYC1* terminator, between the unique *Asp718* and *Spe I* sites of pINP123-306. The construction details are available upon request.

Plating efficiency assay

Yeast transformants were grown in selective SD medium to A_{600} of ~1. Cells in 1 ml of culture were spun down, washed once with H_2O, and resuspended into 1 ml of H_2O. The samples were then serially diluted 3,000 fold with H_2O, and 120 µl of each sample was plated out on duplicate SD and SGal plates. After the plates were incubated at 30°C for ~3 days (or ~6 days as specified), the colonies were counted. Plating efficiency was expressed as the percentage of average number of colonies on SGal plates versus average number of colonies on SD plates.

Isolation of multicopy suppressors of ATE1 overexpression-mediated toxicity

The *ubr1Δ* cells with the integrated P_{GAL1}–ATE1-T_{CYC1} were grown in YPD medium, and transformed with *S. cerevisiae* genomic DNA library in the episomal vector YEp13. The transformants were grown on selective SD plates, and replicate-plated onto SGal plates. The colonies on SGal plates were purified by restreaking on fresh SGal plates. The plasmids in yeast cells were recovered,

and transformed back to *ubr1Δ* cells with the integrated P_{GAL1}–ATE1-T_{CYC1} to ensure that they were directly responsible for suppression of ATE1 overexpression-mediated growth arrest. Both ends of genomic DNA inserts in the plasmids were sequenced, and locations of the inserts in the fully-sequenced *S. cerevisiae* genome were determined.

Results and discussion

Overexpression of ATE1 is more toxic to haploid cells than to diploid cells.

It was observed previously that overexpression of ATE1 from the P_{GAL1} promoter on a high-copy vector caused growth arrest of α haploid yeast cells [2]. To verify and extend this observation, we transformed the empty high copy vector pYES2, or its derivatives pYES2-ATE1 and pYES2-ATE1^f expressing ATE1 or ATE1^f (a functionally active ATE1 derivative bearing C-terminal FLAG tag) from the P_{GAL1} promoter, into congenic haploid (**a** and α) and diploid (**a** /α) cells. The transformants were grown on plasmid-retaining SD medium (where the P_{GAL1} promoter was repressed), and thereafter assayed for plating efficiency on SGal plates, where the P_{GAL1} promoter was fully induced. The haploid and diploid cells harboring empty vector grew robustly on SGal plates. In contrast, the **a** and α haploid cells harboring pYES2-ATE1 or pYES2-ATE1^f did not form colonies after three-day incubation at 30°C (Figure 5.5.1). The results were different with diploid cells: ~60% of the **a** /α cells harboring pYES2-ATE1 or pYES2-ATE1^f formed colonies on SGal plates (Figure 5.5.1), although some of these colonies were significantly smaller than their counterparts on SD plates (Figure 5.5.1b). Furthermore, ~20% of haploid cells harboring pYES2-ATE1 were

observed to form colonies on SGal plates after six-day incubation (Figure 5.5.2a, c). These results demonstrated that overexpression of ATE1 from the P_{GAL1} promoter on a high-copy vector caused either cell death or severe growth retardation of both types of haploid cells, but was significantly less toxic to diploid cells.

We also examined the plating efficiency of haploid cells carrying a low-copy vector expressing ATE1 from the P_{GAL1} promoter (Figure 5.5.2a). Some small colonies were formed after three-day incubation, and ~70% of haploid cells formed colonies on SGal plates after six-day incubation (Figure 5.5.2a), indicating that overexpression of ATE1 from the P_{GAL1} promoter on a low-copy vector was significantly less toxic to haploid cells.

Overexpression of ATE1 is more toxic to *ubr1Δ* cells than to *UBR1* cells.

What might be the molecular mechanism of ATE1 overexpression-mediated toxicity to haploid cells? One model is that a substrate (or substrates) of ATE1, which is required for cell growth, may be arginylated very efficiently in the presence of overexpressed ATE1, and therefore is degraded by the N-end rule pathway at an unphysiologically high rate, so that its steady-state level becomes too low to support normal growth of yeast cells. This model implies that arginylation of this essential N-end rule substrate is a rate-limiting step, since overexpression of UBR1 alone, which accelerates the degradation of engineered N-end rule substrates, is not significantly toxic to haploid cells [7]. To address this possibility, we compared the plating efficiency of congenic *UBR1* and *ubr1Δ* haploid cells carrying a high copy or low copy (*CEN*) plasmid

expressing ATE1 from the P_{GAL1} promoter. Surprisingly, overexpression of ATE1 was even more toxic to *ubr1Δ* cells than to *UBR1* cells: whereas ~11% of *ubr1Δ* haploid cells harboring the episomal plasmid formed small colonies after six-day incubation, ~20% of *UBR1* haploid cells harboring the high copy plasmid formed much bigger colonies under the same conditions (Figure 5.5.2). However, overexpression of UBR1 did not increase further the plating efficiency (data not shown). Although similar percentage (~70%) of *UBR1* and *ubr1Δ* haploid cells carrying the low copy plasmid overexpressing ATE1 formed colonies after six-day incubation (Figure 5.5.2), the colonies of *ubr1Δ* cells were smaller than those of *UBR1* cells (data not shown). These results indicated that UBR1 was not required for ATE1 overexpression-mediated toxicity, and that the accelerated degradation of a substrate(s) of ATE1 by the N-end rule pathway was not the cause of toxicity of overexpressed ATE1. Neither the absence of NTA1, the sole yeast N-terminal amidase that converts tertiary destabilizing N-terminal residues Asn and Gln to secondary destabilizing residues Asp and Glu [8], nor an overexpression of NTA1 affected the plating efficiency of yeast cells carrying the high copy plasmid expressing ATE1 from the P_{GAL1} promoter (data not shown), suggesting that neither NTA1, nor its physiological substrate(s), were involved in ATE1 overexpression-mediated toxicity.

Overexpression of ATE1 mutant with reduced enzymatic activity is less toxic to haploid cells.

To determine whether the enzymatic activity of ATE1 is required for ATE1 overexpression-mediated toxicity, we took advantage of two ATE1

mutants, ATE1C20S (Cys20→Ser) and ATE1C23A (Cys23→Ala), that were demonstrated previously to have very low (<2% of the wild-type ATE1) specific activity in an *in vitro* arginylation assay [9]. These two ATE1 mutants (the gifts from Dr. Cecile Pickart) were compared with wild-type ATE1 in mediating the *in vivo* degradation of N-end rule substrates bearing secondary destabilizing residues. The yeast cells lacking *ATE1* were transformed with two plasmids: a low copy plasmid expressing either ATE1C20S, or ATE1C23A, or wild-type ATE1 from the P_{MET25} promoter, and a high copy plasmid expressing either Ub-Asp-βgal or Ub-Glu-βgal. The latter test proteins were cotranslationally deubiquitylated *in vivo* to produce Asp-βgal or Glu-βgal, the N-end rule substrates bearing secondary destabilizing N-terminal residues. The enzymatic activity of βgal in yeast extract was shown to be a sensitive measure of its metabolic stability *in vivo* [4].

As expected, both Asp-βgal and Glu-βgal were long-lived in *ate1Δ* yeast cells (Figure 5.5.3a), and were short-lived in *ate1Δ* yeast cells expressing the wild-type ATE1. Surprisingly, Asp-βgal and Glu-βgal were also short-lived in cells expressing the nearly inactive ATE1C20S mutant of ATE1 (Figure 5.5.3a). In contrast, Asp-βgal was slightly stabilized, and Glu-βgal was significantly stabilized in *ate1Δ* yeast cells that expressed ATE1C23A (Figure 5.5.3a). These results suggested that the reduced activity of ATE1C20S R-transferase was sufficient to mediate wild-type levels of the N-end rule pathway's activity. In contrast, the residual activity of ATE1C23A mutant yielded only a partial rescue of the N-end rule pathway in *ate1Δ* cells. One explanation of this paradox is that the low enzymatic activities of ATE1C20S or ATE1C23A mutants of R-transferase

in the *in vitro* arginylation assays were either almost completely (for ATE1C20S), or partially (for ATE1C23A) "compensated" by their overexpresion in yeast cells (from the P_{MET25} promoter on a low-copy plasmid) (Figure 5.5.3c, lanes 4, 7, 8). The wild-type ATE1 R-transferase is apparently very efficient in arginylating its substrates, since very low levels of ATE1 in wild-type haploid or diploid yeast cells (undetectable with anti-ATE1 antibody) (Figure 5.5.3c, lanes 1-3 versus lane 6) are sufficient for arginylation of highly expressed Asp-βgal and Glu-βgal (from the P_{GAL1} promoter on a high copy vector) [1].

We compared the plating efficiency of *UBR1* and *ubr1Δ* haploid cells carrying the high copy plasmids expressing wild-type ATE1, ATE1C20S or ATE1C23A from the P_{GAL1} promoter. No colonies of *ubr1Δ* cells carrying the plasmids expressing wild-type ATE1, ATE1C20S were formed after three-day incubation at 30°C, while ~54% of *ubr1Δ* cells carrying the plasmids expressing ATE1C23A, the ATE1 mutant that was strongly weakened in mediating the *in vivo* degradation of Asp-βgal and Glu-βgal, formed colonies (Figure 5.5.3b). Similar results were obtained using *UBR1* cells (data not shown). Thus, the enzymatic activity of ATE1 mutant appeared to correlate with the severity of its overexpression-mediated toxicity. Since the toxicity of ATE1 overexpression was caused, at least in part, by the titration of (essential) Arg-$tRNA^{Arg}$ (see below), one explanation of the correlation is that mutant ATE1 proteins had lower affinity for Arg-$tRNA^{Arg}$ than did wild-type ATE1.

Multicopy suppressors of ATE1 overexpression-mediated toxicity

To search for multicopy suppressors of ATE1 overexpression-mediated toxicity, we integrated a DNA fragment consisting of (from 5′ to 3′) the P_{GAL1} promoter, *ATE1 ORF* and the *CYC1* terminator (hereafter denoted as P_{GAL1}–ATE1-T_{CYC1}) into the *lys2-801* locus of *UBR1* and *ubr1Δ* strains through a two-step procedure using the integrating plasmid pINP123-ATE1 (Figure 5.5.4). The pINP123-ATE1 was derived from pINP123-306, a plasmid designed to facilitate the integration of YFG (your favorite gene) into the *LYS2* or *lys2-801* locus of *S. cerevisiae*. The integrating plasmid pINP123-306 has several very useful features: 1) It retains the unique restriction sites in the polylinker of the popular vector pRS306 [5]; 2) It is compatible with the widely used series of plasmids p4XXprom, all of which contain the *CYC1* terminator and one of the four promoters: P_{GAL1}, P_{GALL}, P_{GALS} and P_{MET25}. Thus, once YFG is cloned into any of the p4XXprom plasmids, it is easily subcloned, together with promoter and the *CYC1* terminator, into pINP123-306; 3) Through a two-step procedure, only the DNA sequence containing a promoter, *YFG* and the *CYC1* terminator flanked by the two fragments of *LYS2 ORF*, but no other sequences, including the *URA3* marker from the plasmid pINP123-306, were integrated into the *LYS2* or *lys2-801* locus, thus allowing repeated use of URA3 selection (Figure 5.5.4).

Overexpression of ATE1 from the integrated P_{GAL1}–ATE1-T_{CYC1} at the *lys2-801* locus appeared to be more toxic to both *UBR1* cells and *ubr1Δ* cells than its overexpression from the P_{GAL1} promoter on a low copy plasmid, and the toxicity was more severe for *ubr1Δ* cells than for *UBR1* cells. Specifically, *ubr1Δ* cells only formed some tiny colonies after ~6-day incubation at 30°C (data not

shown). The multicopy suppressors were selected as described in MATERIALS AND METHODS, and both ends of the inserts in the recovered plasmids were sequenced. All of the inserts were found to contain at least one of the 19 genes encoding $tRNA^{Arg}$, and all four yeast $tRNA^{Arg}$ isoacceptors were recovered at least once, strongly suggesting that genes encoding $tRNA^{Arg}$ were responsible for suppression of ATE1 overexpression-mediated toxicity.

Thus, the ATE1 overexpression-mediated toxicity is likely to result, at least in part, from sequestration of the endogenous pool of $tRNA^{Arg}$ and/or arginylated $tRNA^{Arg}$ (Arg-$tRNA^{Arg}$) due to the binding of ATE1 to (most likely) Arg-$tRNA^{Arg}$. The resulting decrease in the levels of Arg-$tRNA^{Arg}$ would inhibit protein synthesis, this accounting for cell death or severe growth retardation observed upon overexpression of ATE1. The $tRNA^{Arg}$ expressed from high copy plasmids and the native promoter would be expected to suppress this toxicity, as observed. This interpretation suggests that the reduced enzymatic activity of ATE1C23A and the reduced toxicity caused by its overexpression may result from its lower affinity for Arg-$tRNA^{Arg}$. This interpretation still does not explain the fact that overexpression of ATE1 is more toxic to *ubr1Δ* cells than to *UBR1* cells. Thus, while we partially understand the cause of ATE1 overexpression-mediated toxicity, a detailed mechanism, which may prove helpful in understanding the physiological functions of ATE1, remains to be determined.

References

1. Balzi, E., et al., *Cloning and functional analysis of the arginyl-tRNA-protein transferase gene ATE1 of Saccharomyces cerevisiae.* J. Biol. Chem., 1990. **265**: p. 7464-7471.
2. Espinet, C., et al., *An efficient method to isolate yeast genes causing overexpression- mediated growth arrest.* Yeast, 1995. **11**(1): p. 25-32.
3. Sherman, F., *Getting started with yeast.* Meth. Enzymol., 1991. **194**: p. 3-21.
4. Xie, Y. and A. Varshavsky, *The E2-E3 interaction in the N-end rule pathway: the RING-H2 finger of E3 is required for the synthesis of multiubiquitin chain.* EMBO J., 1999. **18**: p. 6832-6844.
5. Sikorski, R.S. and P. Hieter, *A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in S. cerevisiae.* Genetics, 1989. **122**: p. 19-27.
6. Mumberg, D., R. Muller, and M. Funk, *Regulatable promoters of Saccharomyces cerevisiae - comparison of transcriptional activity and their use for heterologous expression.* Nucl. Acids Res., 1994. **22**(25): p. 5767-5768.
7. Xie, Y. and A. Varshavsky, *Physical association of ubiquitin ligases and the 26S proteasome.* Proc. Natl. Acad. Sci. USA, 2000. **97**(6): p. 2497-2502.
8. Baker, R.T. and A. Varshavsky, *Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway.* J. Biol. Chem., 1995. **270**(20): p. 12065-12074.
9. Li, J. and C.M. Pickart, *Binding of phenylarsenoxide to Arg-tRNA-protein transferase is independent of vicinal thiols.* Biochemistry, 1995. **34**: p. 15829-15837.

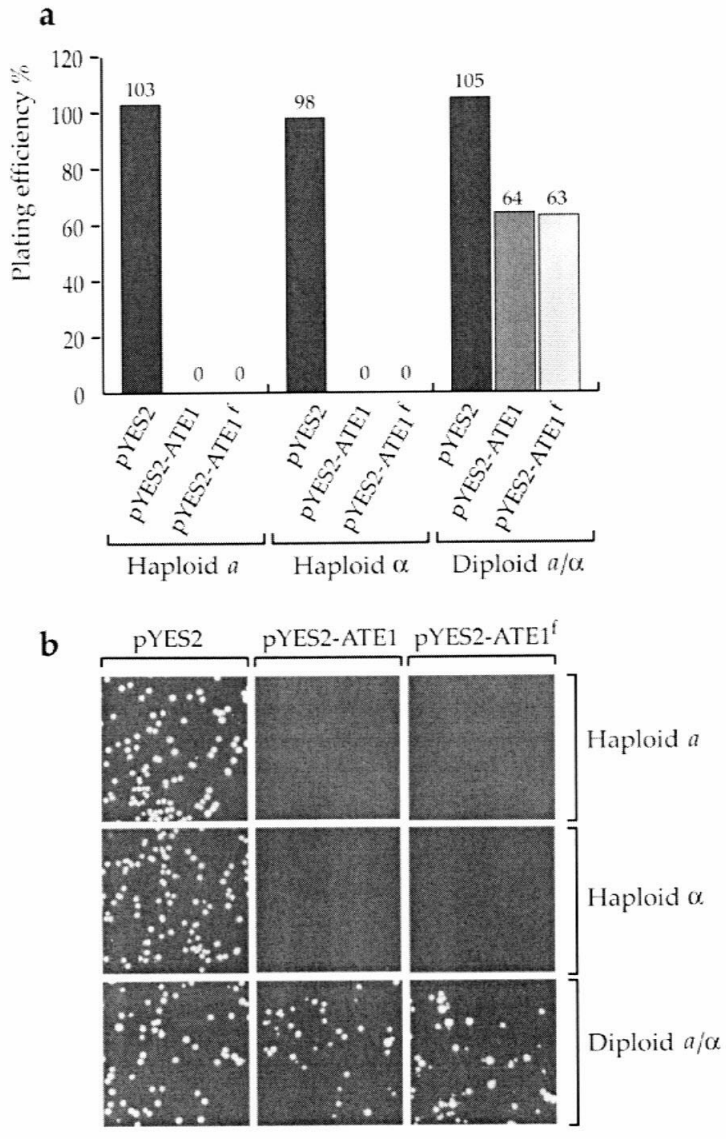


Figure 5.5.1 Overexpression of *S.cerevisiae* ATE1 or ATE1^{f} from the P_{GAL1} promoter in a high copy vector is more toxic to haploid cells than to diploid cells. **a**, Plating efficiency assays of *JD52* (**a** haploid), *JD53* (α haploid) and *GMY1* (**a** /α diploid) transformed with the high copy plasmids pYES2, pYES2-ATE1, or pYES2-ATE1^{f}. The yeast transformants grew robustly in SD media where the P_{GAL1} promoter was repressed. **b**, Representative images of SGal plates from the assays in **a**.

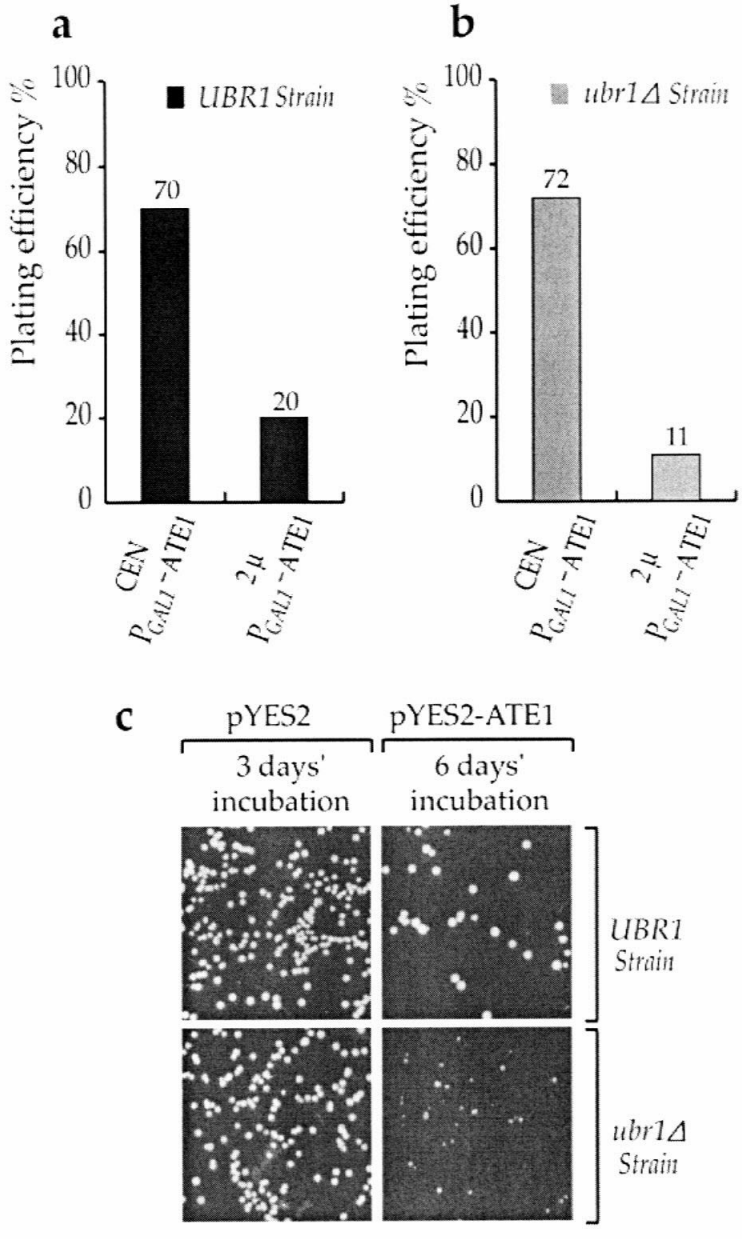


Figure 5.5.2 Overexpression of ATE1 is more toxic to *ubr1Δ* cells than to *UBR1* cells. **a-b**, Plating efficiency assays of *JD52*(**a**) or *JD55* (**b**) transformed with either the low copy plasmid pRS314$_{GAL1/10}$ ATE1 and the high copy plasmid pYES2-ATE1. The SGal plates were incubated at 30°C for ~6 days. **c**, Representative images of SGal plates from the assays in **a-b**.

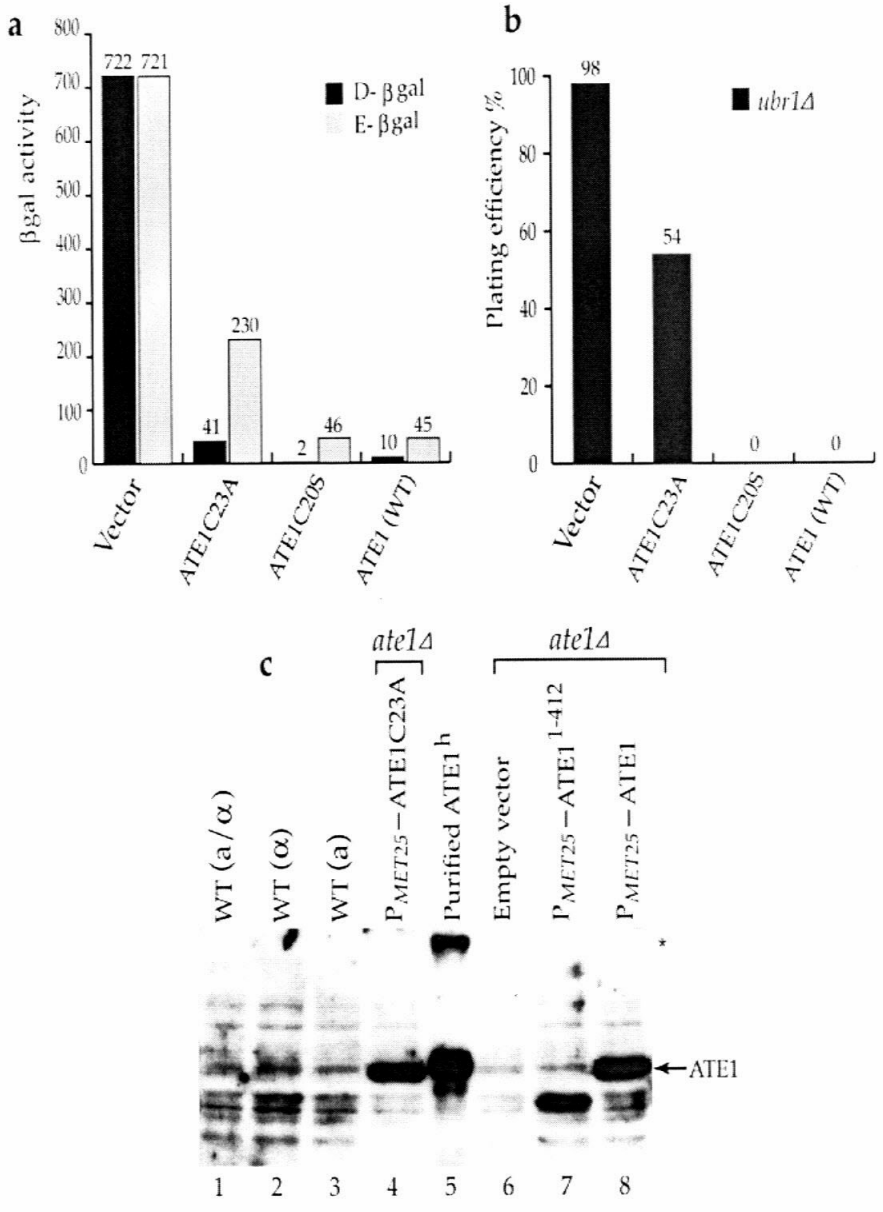

a
βgal activity
800
700
600
500
400
300
200
100
0
722 721
41
230
2
46
10
45
D- βgal
E- βgal
Vector
ATE1C23A
ATE1C20S
ATE1 (WT)
b
Plating efficiency %
100
80
60
40
20
0
98
54
0
0
ubr1Δ
Vector
ATE1C23A
ATE1C20S
ATE1 (WT)
c
ate1Δ
ate1Δ
WT (a/α)
WT (α)
WT (a)
P MET25 – ATE1C23A
Purified ATE1 h
Empty vector
P MET25 – ATE1 1-412
P MET25 – ATE1
ATE1
1 2 3 4 5 6 7 8

Figure 5.5.3 Overexpression of ATE1 mutant with reduced enzymatic activity was less toxic to *ubr1Δ* cells. **a**, Comparison of the wild-type ATE1 and two ATE1 mutants, ATE1C20S and ATE1C23A, in mediating the *in vivo* degradation of Asp-βgal and Glu-βgal. The wild-type ATE1, ATE1C20S and ATE1C23A were expressed in *ate1Δ* yeast cells (SGY3) from the P_{MET25} promoter in the low copy vectors p413MET25-ATE1, p413MET25-ATE1C20S and p413MET25-ATE1C23A, the *ate1Δ* cells also expressed either Asp-βgal or Glu-βgal. The βgal activity in yeast extract is used to measure the *in vivo* metabolic stability of βgal. The values shown are the averages of duplicate measurements. **b**, Plating efficiency assays of *ubr1Δ* cells harboring the high copy plasmids expressing ATE1, ATE1C20S or ATE1C23A from the P_{GAL1} promoter. **c**, Immunodetection of ATE1 and its derivatives in yeast extracts. Yeast extracts containing 10 μg of total proteins from *JD52* (WT a haploid), *JD53* (WT α haploid), *GMY1* (WT a /α diploid) and several transformants of *ate1Δ* strain were fractionated by SDS-10%PAGE, and immunoblotted with anti-ATE1 polyclonal antibody (a gift from Dr. C. M. Pickart). The centromeric plasmid p413MET25-$ATE1^{1-412}$ expresses the N-terminal 412-residue fragment of ATE1 from the P_{MET25} promoter; it is partially active in mediating the *in vivo* degradation of Asp-βgal and Glu-βgal. The purified $ATE1^{h}$ (a variant of ATE1 with C-terminal hexahistidine tag) in lane 5 (20 ng) was overexpressed in *E. coli*, and purified through Ni-NTA affinity chromatograph. An asterisk denotes an $ATE1^{h}$ dimer.

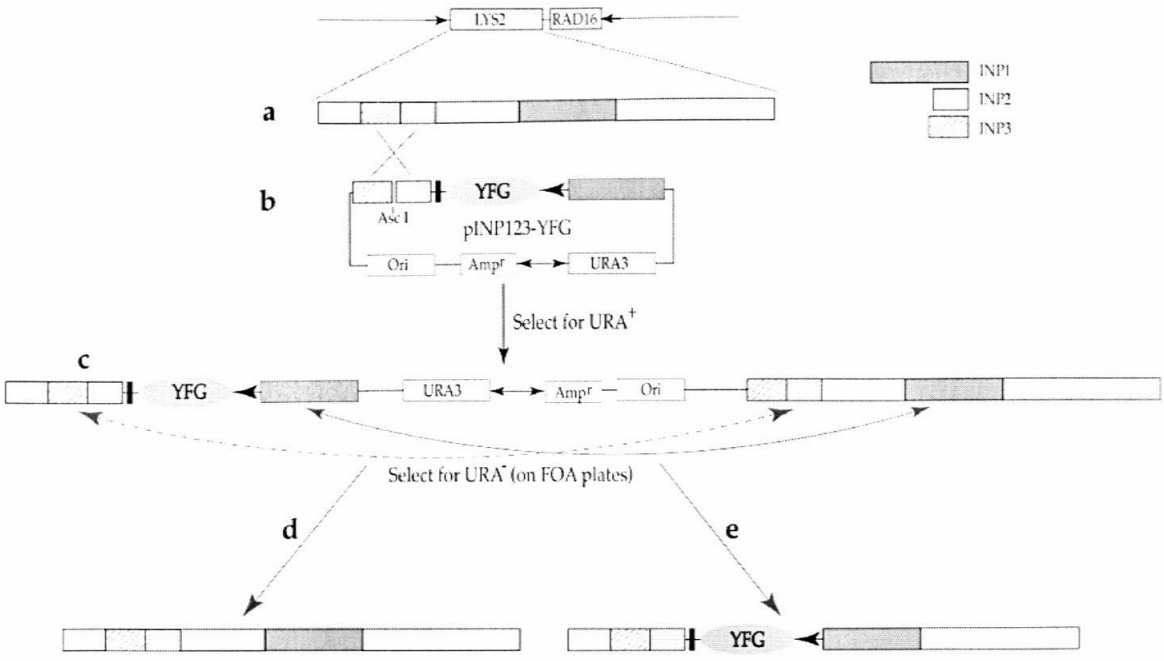


Figure 5.5.4 Integration of YFG (your favorite gene) to the *LYS2* or *lys2-801* locus of *S. cerevisiae* using the pINP123-YFG plasmid. **a**, The *LYS2 ORF* and the locations of its three fragments INP1, INP2 and INP3. **b**, The integrating plasmid pINP123-YFG. The arrows in the plasmid denote the directions of gene transcription, and the black block adjacent to YFG denotes the *CYC1* terminator. The pINP123-YFG plasmid is linearized at the unique *Asc I* site and transformed into yeast strains, and URA$^+$ colonies are selected on SD plates. The structure of *LYS2* or *lys2-801* locus with the integrated plasmid pINP123-YFG is shown in **c**. **d-e**, Two recombination events between direct repeat sequences of *LYS2 ORF*, both of which result in the "pop-out" of *URA3* marker, and thus are selected on YPD plates containing FOA. Only event **e**, usually identified by PCR or Southern blotting, has the desired integration.