The EMBO Journal Vol.18 No.21 pp.6017–6026, 1999

Degradation signals in the lysine–asparagine sequence space

Tetsuro Suzuki[1] and Alexander Varshavsky[2]

Division of Biology, California Institute of Technology, 1200 East California Boulevard, Pasadena, CA 91125, USA

[1]Present address: Department of Virology II, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan

[2]Corresponding author
e-mail: avarsh@its.caltech.edu

The N-degrons, a set of degradation signals recognized by the N-end rule pathway, comprise a protein's destabilizing N-terminal residue and an internal lysine residue. We show that the strength of an N-degron can be markedly increased, without loss of specificity, through the addition of lysine residues. A nearly exhaustive screen was carried out for N-degrons in the lysine (K)–asparagine (N) sequence space of the 14-residue peptides containing either K or N (16 384 different sequences). Of these sequences, 68 were found to function as N-degrons, and three of them were at least as active and specific as any of the previously known N-degrons. All 68 K/N-based N-degrons lacked the lysine at position 2, and all three of the strongest N-degrons contained lysines at positions 3 and 15. The results support a model of the targeting mechanism in which the binding of the E3–E2 complex to the substrate's destabilizing N-terminal residue is followed by a stochastic search for a sterically suitable lysine residue. Our strategy of screening a small library that encompasses the entire sequence space of two amino acids should be of use in many settings, including studies of protein targeting and folding.

Keywords: N-degron/N-end rule/proteolysis/simple sequences/ubiquitin

Introduction

Regulatory proteins are often short-lived *in vivo*, providing a way to generate their spatial gradients and to rapidly adjust their concentration or subunit composition through changes in the rate of their synthesis or degradation. Most of the damaged or otherwise abnormal proteins are metabolically unstable as well. Many other proteins, while long-lived as components of larger structures such as ribosomes and oligomeric proteins, are short-lived as free subunits (reviewed by Hochstrasser, 1996; Varshavsky, 1997; Hershko and Ciechanover, 1998; Scheffner *et al.*, 1998; Koepp *et al.*, 1999; Tyers and Willems, 1999).

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Laney and Hochstrasser, 1999). One class of degradation signals, called the N-degrons, comprises a protein's destabilizing N-terminal residue and an internal Lys residue (Bachmair *et al.*, 1986; Varshavsky, 1996). A set of N-degrons containing different N-terminal residues that are destabilizing in a given cell defines a rule, termed the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. The lysine determinant of an N-degron is the site of formation of a substrate-linked multi-ubiquitin chain (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989). The N-end rule pathway is thus one pathway of the ubiquitin (Ub) system. Ub is a 76-residue protein whose covalent conjugation to other proteins plays a role in a multitude of processes, including cell growth, division, differentiation, and responses to stress (Pickart, 1997; Varshavsky, 1997; Peters, 1998; Scheffner *et al.*, 1998). In many of these settings, Ub acts through routes that involve the degradation of Ub–protein conjugates by the 26S proteasome, an ATP-dependent multisubunit protease (Coux *et al.*, 1996; Hilt and Wolf, 1996; Baumeister *et al.*, 1998; Rechsteiner, 1998).

The N-end rule is organized hierarchically. In the yeast *Saccharomyces cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues in that they function through their conversion, by the *NTA1*-encoded N-terminal amidase, into the secondary destabilizing N-terminal residues Asp and Glu (Baker and Varshavsky, 1995). The destabilizing activity of N-terminal Asp and Glu requires their conjugation, by the *ATE1*-encoded Arg-tRNA-protein transferase, to Arg, one of the primary destabilizing residues. The primary N-terminal residues are bound directly by the *UBR1*-encoded N-recognin, the E3 (recognition) component of the N-end rule pathway. In *S.cerevisiae*, N-recognin is a 225 kDa protein that binds to potential N-end rule substrates through their primary destabilizing N-terminal residues: Phe, Leu, Trp, Tyr, Ile, Arg, Lys and His (Varshavsky, 1996). Analogous components of the mammalian N-end rule pathway have been identified as well (Stewart *et al.*, 1995; Grigoryev *et al.*, 1996; Kwon *et al.*, 1998, 1999).

Studies with engineered N-end rule substrates indicated the bipartite organization of N-degrons and suggested a stochastic model of their targeting, in which specific lysines of an N-end rule substrate could be assigned different probabilities of being used as a ubiquitylation site (Bachmair and Varshavsky, 1989; Chau *et al.*, 1989; Johnson *et al.*, 1990; Hill *et al.*, 1993; Varshavsky, 1996; Lévy *et al.*, 1999). Most of the evidence for this model was produced with a set of N-degrons in which a destabilizing N-terminal residue X was linked to the ~40-residue *Escherichia coli* Lac repressor-derived sequence termed e^K [extension (e) bearing lysines (K)] (Figure 1A) (Bachmair and Varshavsky, 1989). The resulting X-e^K sequence comprised a portable N-degron, which could confer short half-lives on test proteins such as *E.coli* β-galactosidase

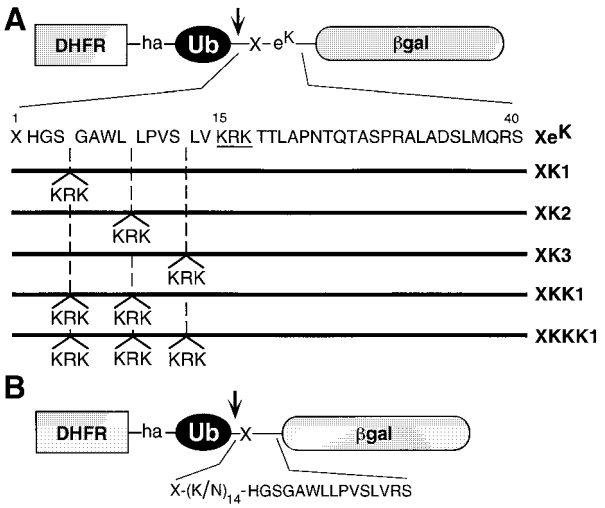


Fig. 1. Test proteins. (**A**) Fusions used in this work contained some of the following elements (see Materials and methods): DHFRha, a mouse dihydrofolate reductase moiety extended at the C-terminus by a sequence containing the hemagglutinin-derived ha epitope; the Ub^{R48} moiety bearing the Lys→Arg alteration at position 48; a 40-residue *E.coli* Lac repressor-derived sequence, termed e^K and shown in single-letter abbreviations for amino acids; a variable residue X (either Tyr, His or Met) between Ub^{R48} and e^K; the *E.coli* βgal moiety lacking the first 24 residues of wild-type βgal. The positions of single or multiple KRK insertions into e^K are indicated. The endogenous KRK sequence of e^K is underlined. The arrow indicates the site of *in vivo* cleavage by DUBs. (**B**) The fusion construct used for screening in the K/N sequence space. The 14 residues of e^K immediately following the residue X were replaced by a set of 14-residue sequences that comprised a random permutation of Lys and Asn residues, followed by the sequence HGSGAWLLPVSLVRS, derived from residues 2–14 of the e^K extension, followed by Arg-Ser.

(βgal) or mouse dihydrofolate reductase (DHFR) (Varshavsky, 1996). At least one of two lysines (K) in e^K, either K-15 or K-17, must be present for the N-degron to be active (Figure 1A) (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1990). Even though several other classes of N-degron, including the naturally occurring ones, have been described over the last decade (Townsend *et al.*, 1988; deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Sadis and Finley, 1995; Ghislain *et al.*, 1996; Sijts *et al.*, 1997; Tobery and Siliciano, 1999), the mechanistic understanding of these degradation signals remains confined largely to the e^K-based N-degrons (Varshavsky, 1996; Lévy *et al.*, 1999).

In the present work, we show that spiking an e^K-based N-degron with additional Lys residues can markedly increase its activity. We also show, using a new approach of searching in the sequence space of lysine and asparagine, that simple-sequence N-degrons can be as strong and specific as any of the previously known N-degrons. These findings provide independent evidence for the model of a bipartite N-degron and stochastic targeting mechanism (Bachmair and Varshavsky, 1989). The strategy of exhaustive searching in the sequence space of two amino acids should be of use in many settings, including studies of protein folding and degradation.

Results and discussion

The Ub/protein/reference technique

The assays below utilized the previously developed Ub/protein/reference (UPR) technique, which increases the accuracy of pulse–chase analysis by providing a 'built-in' reference protein (Lévy *et al.*, 1996). This method employs a linear fusion in which Ub is located between a protein of interest and a reference protein moiety (Figure 1A). The fusion is co-translationally cleaved by Ub-specific de-ubiquitylating enzymes (DUBs) (Wilkinson and Hochstrasser, 1998) after the last residue of Ub, producing equimolar amounts of the protein of interest and the reference protein bearing the C-terminal Ub moiety. If both the reference protein and the protein of interest are immunoprecipitated in a pulse–chase assay, the relative amounts of the protein of interest can be normalized against the reference protein in the same sample. The UPR technique can thus compensate for the scatter of immunoprecipitation yields, sample volumes and other sources of sample-to-sample variation (Lévy *et al.*, 1996, 1999).

Two previously introduced terms, ID^x, initial decay, i.e. the decay of a protein during the pulse of x min, and $t_{0.5}^{y-z}$, the protein's half-life averaged over the interval of y to z min of chase (Lévy *et al.*, 1996), are used below to describe the decay curves of test proteins. The ID^x term and the interval-specific term $t_{0.5}^{y-z}$ would be superfluous in the case of a strictly first-order decay, which is defined by a single half-life. However, the *in vivo* degradation of most proteins deviates from first-order kinetics. For example, the rate of degradation of short-lived proteins can be much higher during the pulse, in part because a newly labeled (either nascent or just-completed) polypeptide is conformationally immature and may, consequently, be targeted for degradation more efficiently than its mature counterpart. This enhanced early degradation, previously termed the 'zero-point' effect (Baker and Varshavsky, 1991), is described by the parameter ID^x (Lévy *et al.*, 1996). It was found that a large fraction of the zero-point effect results from the co-translational degradation of nascent (being synthesized) polypeptide chains, which never reach their mature size before their destruction by processive proteolysis (G.Turner and A.Varshavsky, unpublished data). The detection of a zero-point effect requires the comparison of a test protein's degradation between cells containing and lacking the relevant proteolytic pathway. Alternatively, the zero-point effect can be detected by comparing, through the UPR technique, the degradation of otherwise identical degron-containing and degron-lacking versions of a test protein (Lévy *et al.*, 1996, 1999). Although the degradation of a protein during the pulse can be strikingly high (Lévy *et al.*, 1996) (see also below), it is not detectable by a conventional, reference-lacking pulse–chase assay.

Increasing the strength of N-degrons by spiking them with additional lysines

The UPR constructs of the present work were DHFR-ha-Ub^{R48}-X-e^K-βgal fusions. They contained the metabolically stable, ha-epitope-bearing DHFR-ha-Ub^{R48} moiety as a reference protein, termed dha-Ub below. The dha-Ub-X-e^K-βgal proteins were co-translationally cleaved *in vivo*, yielding the test protein X-e^K-βgal and the reference dha-Ub (Figure 1A). To reduce the possibility that the C-terminal Ub moiety of dha-Ub could function as a ubiquitylation/degradation signal, the K-48 residue of Ub (a major site of isopeptide bonds in multi-Ub

chains) was converted to Arg, which cannot be ubiquitylated, yielding Ub^{R48} (Lévy *et al.*, 1996). These and related fusions (Figure 1A) were expressed in *S.cerevisiae* from low copy plasmids and the copper-inducible P_{CUP1} promoter.

X-e^K-βgal is an extensively analyzed class of N-end rule substrates, which contain a variable N-terminal residue X (produced through the DUB-mediated cleavage of dha-Ub-X-e^K-βgal at the Ub–X junction), a 40-residue N-terminal extension called e^K (see Introduction), and a βgal moiety lacking the first 24 residues of wild-type *E.coli* βgal (Figure 1A). If K-15 and K-17, the only lysines of the e^K extension (Figure 1A), are replaced by Arg residues, which cannot be ubiquitylated, the resulting X-$e^{\Delta K}$-βgal is long-lived even if its N-terminal residue is destabilizing in the N-end rule (Bachmair *et al.*, 1986; Johnson *et al.*, 1990). The inactivity of N-degron in X-$e^{\Delta K}$-βgal is caused by the absence of targetable lysines (Varshavsky, 1996). Specifically, the multiple lysines of the βgal moiety in X-$e^{\Delta K}$-βgal (Chau *et al.*, 1989) cannot serve as N-degron determinants, apparently because the most N-terminal Lys residue in X-$e^{\Delta K}$-βgal, at position 239, is too far from the protein's N-terminus.

One of our aims was to produce stronger N-degrons. We chose Tyr, a moderately destabilizing type 2 residue (Bachmair and Varshavsky, 1989; Varshavsky, 1996), as the N-terminal residue of an initial test protein (Figure 1A). More strongly destabilizing N-terminal residues, e.g. Leu or Arg, in the context of (expected) stronger N-degrons would have made the test proteins too short-lived for detection in a pulse–chase assay. Met was employed as a stabilizing N-terminal residue. The term ID^5 below (see Materials and methods) conveys the extent of degradation of a protein during the 5 min pulse, in comparison with the degradation, during the same pulse, of a control (degron-lacking, i.e. Met-bearing) version of the same protein.

To determine whether the degradation of Tyr-e^K-βgal in *S.cerevisiae* could be enhanced through the addition of Lys residues while remaining dependent on the Ubr1p N-recognin, the sequence Lys-Arg-Lys (KRK), identical to the sequence at positions 15–17 of e^K, was inserted at the indicated locations within e^K (Figure 1A). The unmodified Tyr-e^K-βgal had an ID^5 of ~48%, i.e. ~48% of the labeled Tyr-e^K-βgal was destroyed during the 5 min pulse, before time 0. The $t_{0.5}^{0–10}$ (half-life between 0 and 10 min of chase) of Tyr-e^K-βgal was ~26 min (Figures 2A and 3A). The KRK sequence inserted at any of the indicated three positions within e^K (Figure 1A) strongly destabilized the already short-lived Tyr-e^K-βgal: for example, Tyr-K1e^K-βgal (Figure 1A) had an ID^5 of ~75% and $t_{0.5}^{0–10}$ of ~5 min (Figures 2A and 3A). The increased degradation of Tyr-e^K-βgal derivatives containing extra KRK remained completely Ubr1p-dependent: Tyr-e^K-βgal, Tyr-K1e^K-βgal, Tyr-K2e^K-βgal and Tyr-K3e^K-βgal were all long-lived proteins ($t_{0.5}$ >10 h) in *ubr1*Δ cells (Figures 2A and 3A). In addition, Met-e^K-βgal, Met-K1e^K-βgal, Met-K2e^K-βgal and Met-K3e^K-βgal, the Met-bearing counterparts of the Tyr-based N-end rule substrates, were long-lived in either *UBR1 or ubr1*Δ cells (data not shown).

These results led us to examine the effects of adding more than one KRK sequence to e^K (Figure 1A). The resulting Tyr-e^K-βgal derivatives, bearing either two (Tyr-KK1e^K-βgal) or three (Tyr-KKK1e^K-βgal) KRK sequences, in addition to the original KRK of e^K, were extremely short-lived proteins, even though N-terminal Tyr is a weakly destabilizing residue (Varshavsky, 1996). For example, Tyr-KK1e^K-βgal (Figure 1A) had $t_{0.5}^{0–10}$ of ~4 min (in comparison with ~26 min in the case of Tyr-e^K-βgal) and an ID^5 of ~94%. In other words, ~94% of the labeled Tyr-KK1e^K-βgal was destroyed during the 5 min pulse, before time 0 (Figures 2B and 3B). At the same time, all of these proteins were long-lived in *ubr1*Δ cells (Figures 2B and 3B).

The N-terminal Tyr is bound by the type 2 site of N-recognin (Ubr1p) that recognizes substrates bearing bulky hydrophobic N-terminal residues (Varshavsky,

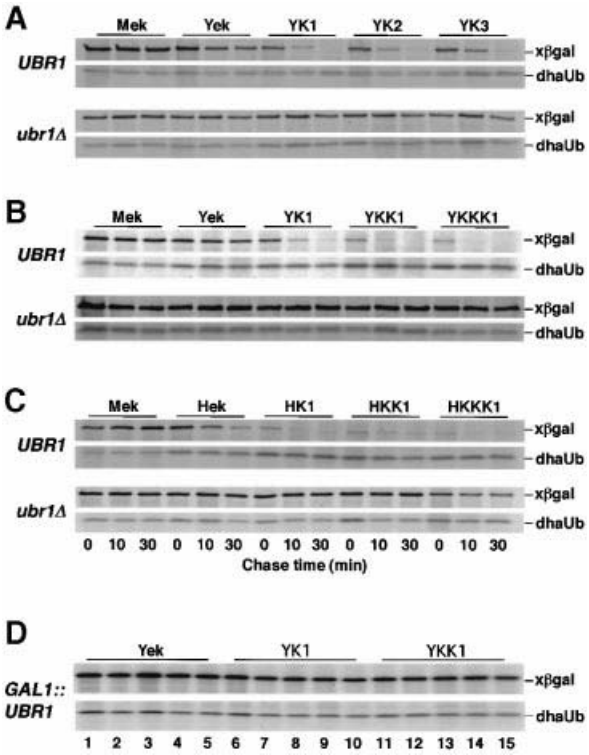


Fig. 2. Active N-degrons can be strongly enhanced by additional lysines. (**A**) Congenic *UBR1 (*wt) and *ubr1*Δ *S.cerevisiae* that expressed the UPR-based fusions Met-e^K-βgal (Mek) (DHFR-ha-Ub^{R48}-Met-e^K-βgal), Tyr-e^K-βgal (Yek), Tyr-K1e^K-βgal (YK1), Tyr-K2e^K-βgal (YK2) and Tyr-K3e^K-βgal (YK3) (see Figure 1A) were labeled with [^{35}S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS–PAGE, and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub^{R48} (reference protein) are indicated on the right. (**B**) As in (A), but with Met-e^K-βgal (Mek), Tyr-e^K-βgal (Yek), Tyr-K1e^K-βgal (YK1), Tyr-KK1e^K-βgal (YKK1) and Tyr-KKK1e^K-βgal (YKKK1) (see Figure 1A). (**C**) As in (A), but with Met-e^K-βgal (Mek), His-e^K-βgal (Hek), His-K1e^K-βgal (HK1), His-KK1e^K-βgal (HKK1) and His-KKK1e^K-βgal (HKKK1). (**D**) Metabolic stability of conformationally mature Tyr-e^K-βgal and its KRK-spiked derivatives. JD54 (P_{GAL1}-*UBR1*) cells expressing Tyr-e^K-βgal (Yek), Tyr-K1e^K-βgal (YK1) or Tyr-KK1e^K-βgal (YKK1) were grown in SM-raffinose medium (no expression of Ubr1p), then labeled with [^{35}S]methionine/cysteine for 10 min at 30°C. After a 20 min chase in SM-raffinose, galactose was added to 3% to induce Ubr1p expression, followed by a chase for 1, 3 and 6 h, and the analysis of immunoprecipitated test proteins. Lanes 1, 6 and 11, the end of ^{35}S labeling (time 0). Lanes 2, 7 and 12, the end of 20 min chase in SM-raffinose. Lanes 3, 8 and 13, 1 h chase with galactose. Lanes 4, 9 and 14, 3 h chase. Lanes 5, 10 and 15, 6 h chase.

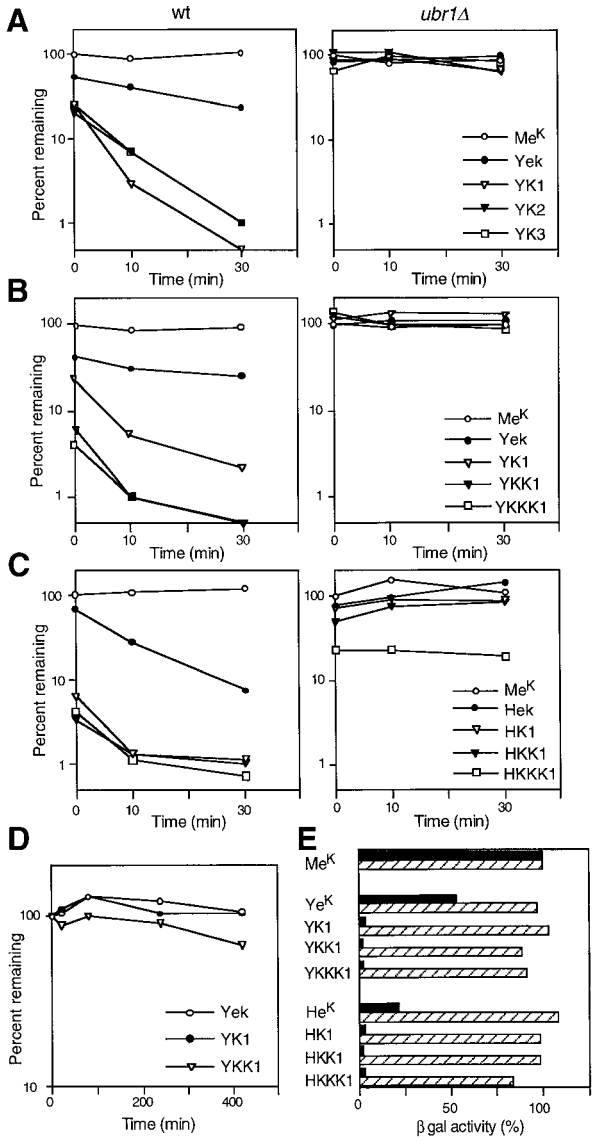


Fig. 3. Quantitation of degradation of the test proteins. (**A**) Pulse–chase patterns of Met-e^K-βgal (Me^K), Tyr-e^K-βgal (Ye^K), Tyr-K1e^K-βgal (YK1), Tyr-K2e^K-βgal (YK2) and Tyr-K3e^K-βgal (YK3) (see Figures 1A and 2A) were quantitated using the UPR technique and PhosphorImager (see Materials and methods). Time 0 refers to the end of the 5 min pulse; 100% refers to the relative amount of Met-e^K-βgal, normalized against the reference protein dha-Ub^{R48}. (**B**) As in (A) but with Met-e^K-βgal (Me^K), Tyr-e^K-βgal (Ye^K), Tyr-K1e^K-βgal (YK1), Tyr-KK1e^K-βgal (YKK1) and Tyr-KKK1e^K-βgal (YKKK1) (see Figures 1A and 2B). (**C**) As in (A) but with Met-e^K-βgal (Me^K), His-e^K-βgal (He^K), His-K1e^K-βgal (HK1), His-KK1e^K-βgal (HKK1) and His-KKK1e^K-βgal (HKKK1) (see Figures 1A and 2C). (**D**) As in (A), but quantitation of the post-translational degradation of Tyr-e^K-βgal (Ye^K), Tyr-K1e^K-βgal (YK1) and Tyr-KK1e^K-βgal (YKK1) in the P$_{GAL1}$-*UBR1* strain JD54 (see Figures 1A and 2A), following the induction of Ubr1p by galactose. Time 0 refers to the end of the 10 min labeling in raffinose (no Ubr1p). The cells were incubated for another 20 min in raffinose, followed by the addition of galactose to induce Ubr1p (see Materials and methods). (**E**) Relative enzymatic activities of βgal in *UBR1* cells (filled bars) and *ubr1*Δ cells (striped bars) expressing one of the following test proteins: Met-e^K-βgal (Me^K), Tyr-e^K-βgal (Ye^K), Tyr-K1e^K-βgal (YK1), Tyr-KK1e^K-βgal (YKK1), Tyr-KKK1ue^K-βgal (YKKK1), His-e^K-βgal (He^K), His-K1e^K-βgal (HK1), His-KK1e^K-βgal (HKK1) and His-KKK3e^K-βgal (HKKK1) (see Materials and methods). The activities of βgal were normalized to the activity of Met-e^K-βgal in each cell. Values shown are the means from duplicate measurements, which yielded results within 10% of the mean values.

1996). We asked whether the above findings were also relevant to the type 1 (basic) destabilizing N-terminal residues Arg, Lys and His, which are bound by the type 1 site of Ubr1p. Counterparts of the Tyr-βgal fusions that bore N-terminal His (a weak type 1 destabilizing residue) were constructed (Figure 1A) and tested in pulse–chase assays. The His residue was chosen as a type 1 destabilizing residue in these tests for the same reason as Tyr in the preceding tests: a stronger destabilizing residue would have made the measurements impractical with extremely short-lived substrates. The results (Figures 2C and 3C) confirmed the generality and specificity of degradation enhancement by the additional KRK sequences. For example, His-K1e^K-βgal had an ID[5] of ~95% (i.e. ~95% of the labeled His-K1e^K-βgal was destroyed during the 5 min pulse, before time 0), in comparison with the ID[5] of ~41% for His-e^K-βgal; the corresponding $t_{0.5}^{0-10}$ values were ~5 min and ~8 min for His-K1e^K-βgal and His-e^K-βgal, respectively (Figure 3C).

Similar to the results with Tyr-bearing substrates, their His-bearing, multiple KRK-containing counterparts were long-lived in *ubr1*Δ cells (Figures 2C and 3C). The only exception was His-KKK1e^K-βgal (Figure 1A), which contained three KRK sequences, in addition to the KRK of the original e^K: in contrast to Tyr-KKK1e^K-βgal, His-KKK1e^K-βgal was stabilized strongly but incompletely in the *ubr1*Δ genetic background (Figure 3C). Thus, the His-KKK1e^K extension, in contrast to the Tyr-KKK1e^K extension (Figure 1A), appears to contain a Ubr1p-independent degron.

Previous work (Madura *et al.*, 1993; Kwon *et al.*, 1999) has shown that the steady-state level of an X-βgal protein (determined by measuring the enzymatic activity of βgal in yeast extracts) is a sensitive measure of its metabolic stability. The results of this steady-state assay were in agreement with those derived from pulse–chase measurements: the level of Tyr-e^K-βgal in *UBR1* cells was 53% of the level of the long-lived Met-e^K-βgal, whereas Tyr-K1e^K-βgal was present at 4% of the Met-e^K-βgal level, and the concentration of Tyr-KKK1e^K-βgal was virtually indistinguishable from the assay's background (cells transformed with vector alone) (Figure 3E). Crucially, the levels of these extra KRK-bearing Tyr-e^K-βgal fusions in *ubr1*Δ cells became similar to that of Met-e^K-βgal (Figure 3E), in agreement with the pulse–chase data (Figures 2A, B and 3A, B).

Although the addition of extra KRK sequences to e^K yielded considerable decreases in the $t_{0.5}^{0-10}$ of the corresponding N-end rule substrates, by far the major effect of multiple KRK sequences was on the decay curves' ID[5] term, which conveys the extent of degradation of a protein during or shortly after its synthesis (Figures 2A–C and 3A–C). To examine this issue in a different way, Tyr-e^K-βgal, Tyr-K1e^K-βgal and Tyr-KK1e^K-βgal (Figure 1A) were produced in the JD54 *S.cerevisiae* strain, which expressed Ubr1p from the galactose-inducible, dextrose-repressible P$_{GAL1}$ promoter. JD54 cells expressing one of the test proteins were labeled in raffinose-containing SR medium (no Ubr1p), incubated for 20 min in the same medium and thereafter shifted to galactose, where Ubr1p was induced. Even though the N-end rule pathway became hyperactive in the presence of galactose (Madura and Varshavsky, 1994; Ghislain

et al., 1996; data not shown), the pre-labeled substrates Tyr-e^K-βgal and Tyr-K1e^K-βgal were barely degraded after the induction of Ubr1p; Tyr-KK1e^K-βgal was degraded only slightly (Figures 2D and 3D).

These findings were consistent with the earlier evidence for a strong retardation of the post-translational degradation of Arg-e^K-βgal under the same conditions (R.J.Dohmen and A.Varshavsky, unpublished data). Thus, in contrast to a newly formed, conformationally immature βgal-based test protein, a conformationally mature βgal tetramer is a poor substrate of the N-end rule pathway even in the presence of N-degron enhancements such as the additional KRK sequences. It is the βgal moiety of these test proteins (Figure 1A) that was responsible for the time-dependent decline in the rate of degradation, because the kinetics of *in vivo* degradation of e^K-DHFR-based N-end rule substrates was much closer to first-order decay (Lévy *et al.*, 1999; data not shown).

Locating N-degrons in the lysine–asparagine sequence space

The earlier work, which led to the bipartite model of N-degron (Bachmair and Varshavsky, 1989; Hill *et al.*, 1993), and particularly the present findings about the effects of adding KRK sequences to an e^K-based N-degron (Figures 2 and 3) suggested that a substrate's destabilizing N-terminal residue and a sterically suitable internal Lys residue (or residues) are the two necessary and sufficient components of an N-degron. However, since both the e^K-based and other previously analyzed N-degrons are embedded in complex sequence contexts (deGroot *et al.*, 1991; Dohmen *et al.*, 1994; Varshavsky, 1996), we wished to address the bipartite-degron model by constructing an N-degron from much simpler sequence motifs. Should this prove feasible, we also wanted to explore constraints on the structure of N-degrons through a screen in a simpler sequence setting. If the sequence space could be reduced strongly enough, one advantage of such a screen would be its exhaustiveness. The AAA codon for lysine differs by just one third-letter substitution from the codon for asparagine (AAU), a polar uncharged residue. Thus, one could attempt a screen for N-degrons in the sequence space of two amino acids: Lys (K) and Asn (N).

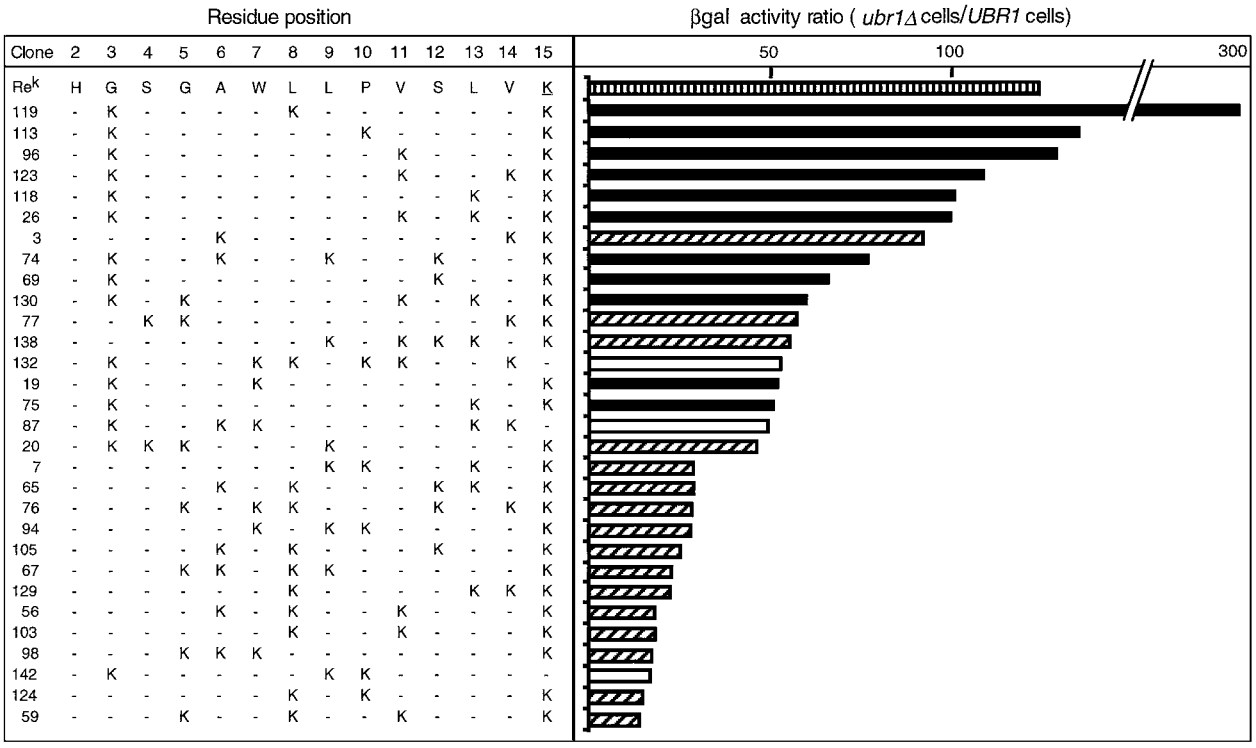
A double-stranded oligonucleotide that encoded random 14-residue K/N sequences (see Materials and methods) was used to replace the sequence encoding 14 residues of e^K immediately following the residue X (Figure 1B). In the resulting test proteins, this latter sequence, HGSG-AWLLPVSLVRS (plus the sequence RS), followed the quasi-random 14-residue K/N sequence (Figure 1A). The resulting K/N-based extensions either lacked the lysines or contained a variable number of them between residues 2 and 16. The K-17 of e^K (the only other lysine in e^K) was replaced by Arg. In these test fusions, dha-Ub^{R48}-Arg-$(K/N)_{14}$-e^Δ-βgals, Arg was used as a destabilizing N-terminal residue (Figure 1B). The number of different 14-residue sequences containing exclusively K or N is $2^{14} = 16\,384$. The bulk of a library of this complexity could be encompassed with conventional screening methods. Testing of the pRKN14-based library by amplifying it in *E.coli* indicated that >90% of the plasmids contained an oligonucleotide insert. The pRKN14 library was introduced into *S.cerevisiae* JD54 (Ghislain *et al.*, 1996), which expressed Ubr1p from the P_{GAL1} promoter, and screened for colonies that stained blue with XGal [high levels of Arg-$(K/N)_{14}$-e^Δ-βgal] on dextrose (SD) plates but stained white [low levels of Arg-$(K/N)_{14}$-e^Δ-βgal] on replica-plated galactose (SG) plates. Approximately 20 000 colonies were screened this way. A total of 68 isolates were identified in which the activity of βgal was significantly higher in the absence than in the presence of the N-end rule pathway.

The corresponding Arg-$(K/N)_{14}$-e^Δ-βgal test proteins were expressed in congenic *ubr1*Δ and *UBR1* strains, and the ratio of βgal activities was determined for each of the test proteins. The results are summarized in Figure 4, which shows the K/N sequences of the 30 most active N-degrons, and the ratios of the corresponding βgal activity in the *ubr1*Δ strain to that in the *UBR1* strain (higher ratios indicate stronger N-degrons). Remarkably, the strongest K/N-based N-degron was found to be more active than the strongest e^K-based N-degron (Figure 4). Black bars in Figure 4 denote βgal activity derived from constructs carrying K/N-based N-degrons with lysines present at positions 3 and 15; the strongest N-degrons were largely of this class (Figure 4). K-15 was present in the 15 strongest K/N-based N-degrons except one (clone 132), which had K at position 14, and was also, presumably in compensation for the absence of K-15, one of the most lysine-rich N-degrons in this set (Figure 4). Similarly, K-3 was present in the 15 strongest K/N-based N-degrons except three (clones 3, 77 and 138) (Figure 4). All of these exceptional clones bore K-15; in addition, one of them (clone 77) bore K-4 and K-5, as well as K-14 and K-15 (Figure 4).

A completely uniform feature of all 68 K/N-based N-degrons was the absence of K from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. To address this issue directly, we used site-directed mutagenesis, replacing asparagine at position 2 of clone 119 (the strongest K/N-based N-degron; Figure 4) with lysine. The resulting test protein was long-lived in *UBR1* cells (data not shown), confirming that lysine is not tolerated at position 2 of an N-degron.

Even though all of the strongest K/N-based N-degrons contained at least three lysines, the total number of lysines in a K/N-based N-degron was not a strong predictor of its activity, and the arrangement of additional lysines between positions 3 and 15 did not correlate, in an obvious way, with the activity of N-degrons (Figure 4). None of the 15 strongest K/N-based N-degrons had more than six lysines (most had from three to five) (Figure 4), indicating that the density of lysines *per se* is not the main feature of a strong K/N-based N-degron. Note that our screen rejected K/N-based degrons that exhibited a significant Ubr1p-independent activity. This may in part account for the upper limit on the number of lysines in the strongest N-degrons: more lysine-rich K/N sequences could contain motifs recognized by non-N-end rule pathways, similarly to the above His-KKK1-βgal test protein (Figures 1A, 2C and 3C).

The *in vivo* degradation of Arg-$(K/N)_{14}$-e^Δ-βgal proteins bearing K/N-based N-degrons [produced from the dha-Ub^{R48}-Arg-$(K/N)_{14}$-e^Δ-βgal fusions (Figure 1B)] was com-



Clone	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Re^K	H	G	S	G	A	W	L	L	P	V	S	L	V	K
119	-	K	-	-	-	-	K	-	-	-	-	-	-	K
113	-	K	-	-	-	-	-	-	K	-	-	-	-	K
96	-	K	-	-	-	-	-	-	-	K	-	-	-	K
123	-	K	-	-	-	-	-	-	-	K	-	-	K	K
118	-	K	-	-	-	-	-	-	-	-	-	K	-	K
26	-	K	-	-	-	-	-	-	-	K	-	K	-	K
3	-	-	-	-	K	-	-	-	-	-	-	-	K	K
74	-	K	-	-	K	-	-	K	-	-	K	-	-	K
69	-	K	-	-	-	-	-	-	-	-	K	-	-	K
130	-	K	-	K	-	-	-	-	-	K	-	K	-	K
77	-	-	K	K	-	-	-	-	-	-	-	-	K	K
138	-	-	-	-	-	-	-	K	-	K	K	K	-	K
132	-	K	-	-	-	K	K	-	K	K	-	-	K	-
19	-	K	-	-	-	K	-	-	-	-	-	-	-	K
75	-	K	-	-	-	-	-	-	-	-	-	K	-	K
87	-	K	-	-	K	K	-	-	-	-	-	K	K	-
20	-	K	K	K	-	-	-	K	-	-	-	-	-	K
7	-	-	-	-	-	-	-	K	K	-	-	K	-	K
65	-	-	-	-	K	-	K	-	-	-	K	K	-	K
76	-	-	-	K	-	K	K	-	-	-	K	-	K	K
94	-	-	-	-	-	K	-	K	K	-	-	-	-	K
105	-	-	-	-	K	-	K	-	-	-	K	-	-	K
67	-	-	-	K	K	-	K	K	-	-	-	-	-	K
129	-	-	-	-	-	-	K	-	-	-	-	K	K	K
56	-	-	-	-	K	-	K	-	-	K	-	-	-	K
103	-	-	-	-	-	-	K	-	-	K	-	-	-	K
98	-	-	-	K	K	K	-	-	-	-	-	-	-	K
142	-	K	-	-	-	-	-	K	K	-	-	-	-	-
124	-	-	-	-	-	-	K	-	K	-	-	-	-	K
59	-	-	-	K	-	-	K	-	-	K	-	-	-	K

Fig. 4. N-degrons in the K/N sequence space. The deduced sequences of the identified K/N N-degrons are shown in conjunction with the bar diagram of their relative activity, defined as the ratio of βgal activities in the *ubr1*Δ versus *UBR1* cells expressing a given Arg-$(K/N)_{14}$-e^{Δ}-βgal test protein. The top bar (vertical stripes) indicates the relative activity of N-degron in the original Arg-e^K-βgal (Re^K). Thirty 14-residue K/N extensions with the highest Ubr1p-dependent destabilizing activity are listed, out of the total of 68 isolates that were metabolically unstable in the presence but not in the absence of Ubr1p. The total number of possible 14-residue K/N sequences is 16 384 (see the main text). The Lys and Asn residues are denoted as the letter K and a hyphen, respectively. Position 1 in each clone was occupied by Arg. The extensions in which lysines were present either at positions 3 and 15, or only at 15, or only at 3, are marked, respectively, by the filled, striped and open bars.

pared with the degradation of Arg-e^K-βgal in pulse–chase assays (Figure 5A and B). The chosen K/N degrons (clones 119, 113 and 4) were the two most active ones in the βgal assay (clones 119 and 113) (Figure 4), and a relatively weak one (clone 4) (Figure 6A). As expected from the results of steady-state βgal assays (Figures 4), Arg-$(K/N)_{14}$-e^{Δ}-βgal proteins 119 and 113 were extremely short-lived in *UBR1* cells (Figure 5A), in contrast to their stability in *ubr1*Δ cells (Figure 5B). As with Arg-e^K-βgal, the bulk of degradation of Arg-$(K/N)_{14}$-e^{Δ}-βgals took place either during or shortly after their synthesis, so that even at time 0 (at the end of the 5 min pulse) the test proteins could be detected only by overexposing the autoradiograms (Figure 5A; data not shown; compare with Figure 5B). In contrast, Arg-$(K/N)_{14}$-e^{Δ}-βgal bearing the weaker N-degron of clone 4 was readily detectable at time 0, and decayed more slowly afterwards (Figure 5A), in agreement with the results of steady-state βgal assays (Figure 6B). The distribution of three lysines in the strongest N-degron (clone 119: Lys3, 8, 15) (Figure 4) was similar to that in a relatively weak one (clone 4: Lys4, 12, 14) (Figure 6), emphasizing the importance of lysines at positions 3 and 15.

To address in more detail the relative contributions of the lysines at positions 3 and 15 to the activity of a K/N-based N-degron, site-directed mutagenesis was used to construct the otherwise identical N-degrons that contained either exclusively Lys3 and Lys15 (no lysine at a third, interior position), exclusively Lys3 or exclusively Lys15 (Figure 6A). The resulting Arg-$(K/N)_{14}$-e^{Δ}-βgal proteins were examined using both steady-state (Figure 6B) and pulse–chase assays (Figure 5C and D). The K/N-based N-degron that contained only Lys3 and Lys15 was active but considerably less so than the strongest K/N-based N-degron of clone 119, which contained Lys8 as well (Figures 4, 5C and 6). The elimination of either Lys3 or Lys15 further weakened the N-degron, so that the Lys3 and Lys15 versions of Arg-$(K/N)_{14}$-e^{Δ}-βgal were, respectively, slightly and moderately short-lived proteins (Figures 5C, D and 7).

Mechanistic implications

One finding of this work is that an N-degron can be greatly strengthened by spiking it with additional lysine residues. The resulting degradation signals remained specific: nearly all of the enhanced N-degrons were completely inactive in *ubr1*Δ cells, which lacked the E3 (recognition) component of the N-end rule pathway. In addition, the enhanced N-degrons could be inactivated by replacing their destabilizing N-terminal residue with Met, a stabilizing residue in the N-end rule.

The strength of a K/N-based N-degron depends on the arrangement of lysines in the 14-residue N-terminal region of the test protein. This dependence is both strong and complex (Figures 4 and 6). The patterns observed may reflect distinct conformational flexibilities of different K/N sequences *vis-à-vis* the relatively fixed spatial arrangement of the type 1 or 2 sites of Ubr1p (N-recognin) and its associated Ubc2p E2 enzyme (Madura *et al.*, 1993).

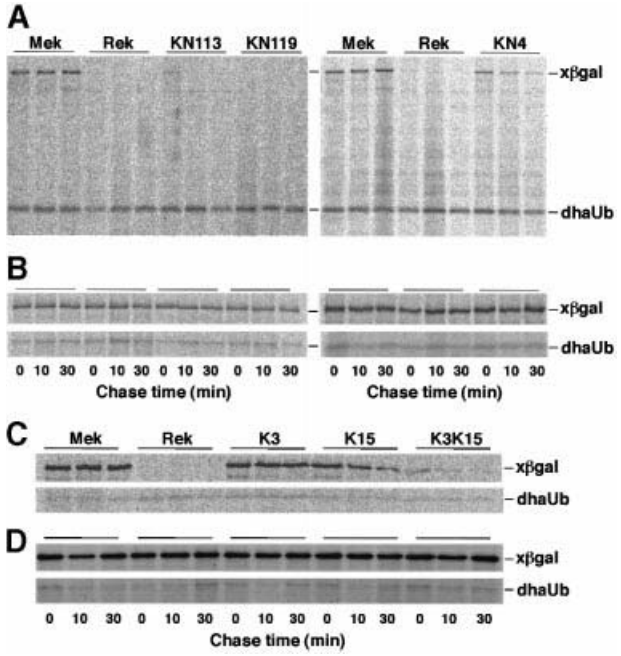


Fig. 5. Pulse–chase analysis of K/N-based N-degrons. (**A**) JD47-13C (*UBR1*) *S.cerevisiae* that expressed the UPR-based fusions Met-e^K-βgal (Mek) (DHFR-ha-Ub^{R48}-Met-e^K-βgal), R-e^K-βgal (Rek), clone 113 of Arg-$(K/N)_{14}$-e^Δ-βgal (KN113), clone 119 of Arg-$(K/N)_{14}$-e^Δ-βgal (KN119), clone 4 of Arg-$(K/N)_{14}$-e^Δ-βgal (KN4) and other clones (see Figures 1B and 4) were labeled with [^{35}S]methionine/cysteine for 5 min at 30°C, followed by a chase for 0, 10 and 30 min, extraction, immunoprecipitation with anti-ha and anti-βgal antibodies, SDS–PAGE and autoradiography (see Materials and methods). The bands of X-βgal (test protein) and DHFR-ha-Ub^{R48} (reference protein) are indicated on the right. (**B**) As in (A), but with the congenic JD55 (*ubr1*Δ) cells. (**C**) As in (A) but with the clones K3, K15 and K3K15 (see Figure 6). (**D**) As in (C) but with JD55 (*ubr1*Δ) cells.

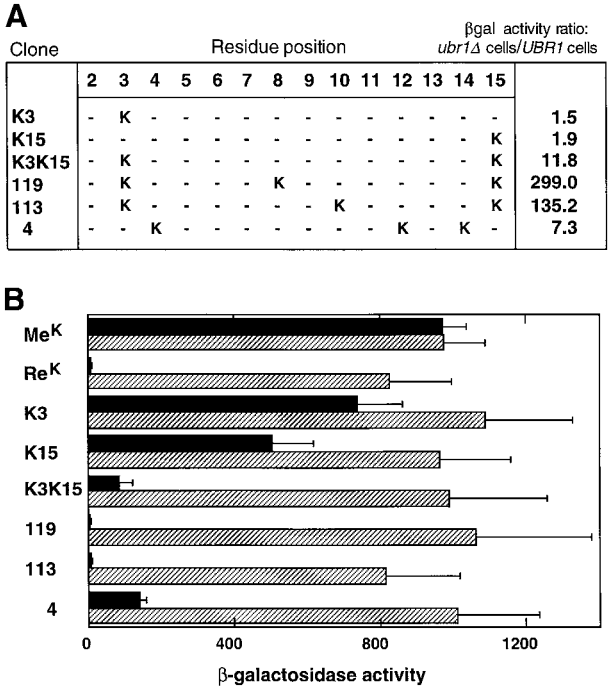


Fig. 6. Site-directed mutagenesis of N-degrons identified through the screen in the lysine–asparagine sequence space. (**A**) K/N sequences of clones 119, 113 and 4 and the K3/K15 derivatives of clone 119 (K3, K15 and K3K15). The relative activity of these N-degrons, defined as the ratio of βgal activities in the *ubr1*Δ versus *UBR1* cells expressing a given Arg-$(K/N)_{14}$-e^Δ-βgal protein, is indicated on the right. (**B**) The levels of βgal activity in extracts from *UBR1* cells (filled bars) and *ubr1*Δ cells (striped bars) cells expressing the indicated test proteins. Standard deviations (for triplicate measurements) are indicated.

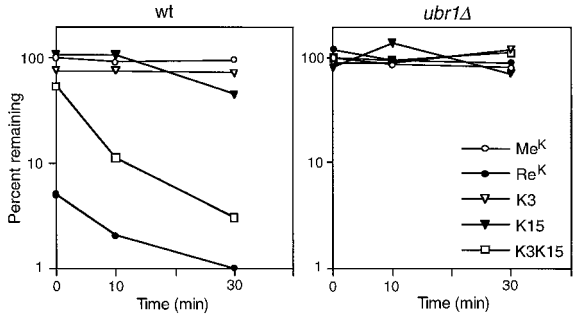


Fig. 7. Decay curves of K3 and/or K15 mutants of Arg-$(K/N)_{14}$-e^Δ-βgal in *UBR1 and ubr1*Δ *S.cerevisiae*. Pulse–chase patterns of Met-e^K-βgal (Me^K), Arg-e^K-βgal (Re^K), and Lys3 (K3), Lys15 (K15) and Lys3/Lys15 (K3K15) variants of Arg-$(K/N)_{14}$-e^Δ-βgal in Figure 5C and D were quantitated as described in the legend to Figure 3 and Materials and methods.

Some of the conclusions indicated by our data are described below.

(i) A completely uniform feature of all 68 K/N-based N-degrons was the absence of Lys from position 2 (Figure 4; data not shown), consistent with the fact that all of the previously examined N-degrons (Varshavsky, 1996) also lacked a strongly basic residue (Arg or Lys) at position 2. Using site-directed mutagenesis with clone 119 (Figure 4), we confirmed that lysine is not tolerated at position 2 of an N-degron (data not shown).

(ii) The strongest K/N-based N-degrons contained single Lys residues, surrounded by Asn residues (Figure 4), in contrast to the Lys-Arg-Lys (KRK) motif present in the e^K extension of N-degrons studied previously (Figure 1A). We conclude that a lysine-containing motif of three adjacent basic residues is not an essential feature of an N-degron.

(iii) The presence of Lys15 in all of the most active K/N-based N-degrons (Figure 4) is consistent with the earlier model (Bachmair and Varshavsky, 1989; Varshavsky, 1996) in which a targetable Lys residue should be sufficiently far along the chain from a destabilizing N-terminal residue to allow the formation of a loop that positions this lysine spatially close to the N-terminal residue. This model is also consistent with the finding that Lys3 is another, nearly invariant, component of a K/N-based N-degron. Specifically, Lys3 may be located at the uniquely favorable 'linear' distance from the N-terminal residue, allowing its proximity to the destabilizing N-terminal residue in the absence of loop formation. That this view is at best incomplete is indicated by the finding that either Lys3 alone or Lys15 alone cannot substitute for the combination of Lys3, Lys8 and Lys15 that defines the strongest K/N-based N-degron identified to date (clone 119 in Figures 4 and 6). Moreover, even Lys3 and Lys15 together, in the absence of Lys8, result in a much weaker N-degron that the three-lysine N-degron of clone 119 (Figure 6). One possibility is that Lys15 is the only lysine, among the three, that can function as the site of ubiquitylation, and that lysines at positions 3 and 8 are required largely for optimal conformational flexibility

of the 14-residue K/N region in this N-degron. It is unlikely that the much higher activity of the clone 119 (three-lysine) N-degron resulted simply from its higher positive charge, in comparison with the two-lysine N-degrons, because most of the other K/N sequences containing three or more lysines were much less active than the clone 119 N-degron (Figure 4).

We searched for N-degrons in a K/N region 14 residues long (Figures 1B and 4), in part because the search with a significantly longer K/N region would have precluded the screen from being exhaustive. Nevertheless, since the e^K extension of the earlier N-degrons is 40 residues long (Figure 1A), and since a part of the e^K extension was retained, downstream of the K/N region, in our test proteins (see Materials and methods), it remains to be determined whether a K/N-based N-degron could be made even stronger by replacing the entire e^K with a relevant K/N motif. For example, one could take advantage of the already identified 14-residue K/N sequence of clone 119 (Figure 4), and carry out an analogous screen for N-degrons in which the K/N sequence of clone 119 is fixed, while the downstream, e^K-derived sequence is replaced by random K/N motifs. In sum, the K/N strategy of this work is far from exhausted by the present screen, and can be used, for example, to produce even stronger N-degrons, to search for other K/N-based degradation signals and to probe the targeting mechanisms of the Ub system.

Concluding remarks

One implication of our results is the possibility of constructing much stronger N-degrons. These portable degradation signals can be used to render proteins of interest that are short-lived in either conditional or unconditional settings (Dohmen *et al.*, 1994; Worley *et al.*, 1998).

Another implication of our results stems from the demonstrated feasibility of defining a specific class of degradation signals in a sequence space of just two amino acids. Since the N-degrons have been defined previously in much more complex sequence contexts (Varshavsky, 1996), our findings suggest that other classes of degradation signal, and some of the other targeting signals as well, could also be identified and examined in this simple-sequence setting. Low-complexity sequences eliminate some of the informational 'noise' of natural sequences and thereby help to define the major determinants of structural specificity. The advantages of the simple-sequence approach have long been recognized by researchers who study the fundamentals of protein folding (Clarke, 1995). To our knowledge, the present work is the first to extend the simple-sequence approach to the realm of protein degradation.

Materials and methods

Strains, media and genetic techniques

Saccharomyces cerevisiae strains used in this work were JD47-13C (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112*) (Madura *et al.*, 1993), JD55 (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 ubr1-Δ1::HIS3*) (Madura and Varshavsky, 1994) and JD54 (*MAT***a** *ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3 112 GAL1::UBR1*) (Ghislain *et al.*, 1996). Rich (YPD) medium contained 1% yeast extract, 2% peptone (Difco) and 2% glucose. Synthetic media (Ausubel *et al.*, 1996) contained either 2% dextrose (glucose) (SD medium), 3% raffinose (SR medium) or 3% galactose (SG medium). To induce the P_{CUP1} promoter, $CuSO_4$ was added to a final concentration of 0.2 mM. Cells were incubated and assayed at 30°C. Transformation of *S.cerevisiae* was carried out using the lithium acetate method (Ausubel *et al.*, 1996).

Plasmid construction

The plasmids encoded Ub fusions of the UPR technique (Lévy *et al.*, 1996)] (see Results and discussion). The reference protein was mouse DHFR fused, through a 20-residue spacer containing the ha epitope, to the Ub^{R48} moiety bearing Arg instead of wild-type Lys at position 48 (Lévy *et al.*, 1996). The plasmids pDhaUbXeKβgal expressed, from the P_{CUP1} promoter, the fusions DHFR-ha-Ub^{R48}-X-e^K-βgal, where the junctional residue X was either Met, Tyr or His; e^K was a previously described 40-residue, *E.coli* Lac repressor-derived N-terminal extension (Bachmair and Varshavsky, 1989; Johnson *et al.*, 1992); the *E.coli* βgal moiety lacked the first 24 residues of wild-type βgal (Bachmair *et al.*, 1986). The βgal-coding fragment was produced by PCR amplification of the βgal open reading frame (ORF) in pUB23 (Bachmair *et al.*, 1986; Bachmair and Varshavsky, 1989), using primers BGALS1 (5′-CAGAGATCTCTTAATCGCCTTGCAGCA-3′) and RSAS2 (5′-CCCTCGAGGTCGACGGTATCG-3′), and the Expand PCR System (Boehringer, Indianapolis, IN). The *Bgl*II/*Xho*I-cut PCR product was ligated to the insert-lacking, P_{CUP1} promoter-containing *Bgl*II–*Xho*I fragment of pdUbXekUra3, a pRS314-based, low copy plasmid (F.Lévy and A.Varshavsky, unpublished data). ORFs that expressed DHFR-ha-Ub^{R48}-X-e^K-βgal containing a modified e^K region [single, double or triple insertions of the sequence Lys-Arg-Lys (KRK) (Figure 1A)] were produced using PCR. Specifically, pDhaUbXK1βgal, pDhaUbXK2βgal, pDhaUbXK3βgal, pDhaUbXKK1βgal and pDhaUbXKKK1βgal were constructed by replacing the *Bam*HI–*Bgl*II fragment encoding X-e^K-βgal with DNA fragments amplified using, respectively, the sense primers EKKS1 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGTTGCCC-3′), EKKS2 (5′-CACGGATCCGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCACTGGTG-3′), EKKS3 (5′-CACGGATCCGGAGCTTGGCTGTTGCCCGTCTCAAAGAGAAAGCTGGTGAAAAGAAAAACC-3′), EKKS4 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCACTGGTG-3′) and EKKKS1 (5′-CACGGATCCAAGAGAAAGGGAGCTTGGCTGAAGAGAAAGTTGCCCGTCTCAAAGAGAAAGCTGGTGAAAAGAAAAACC-3′). The primer EKAS1 (5′-GGAAGATCTCTGCATTAATGAATC-3′) and the plasmid pDhaUbMeKβgal served as the antisense primer and the template, respectively. The plasmids pDhaUbRK3eΔβgal, pDhaUbRK15-eΔβgal and pDhaUbRK3K15eΔβgal, which expressed, respectively, DHFR-ha-Ub^{R48}-Arg-Asn-Lys-$(Asn)_{12}$-e^{Δ}-βgal, DHFR-ha-Ub^{R48}-Arg-$(Asn)_{13}$-Lys-e^{Δ}-βgal and DHFR-ha-Ub^{R48}-Arg-Asn-Lys$(Asn)_{11}$-Lys-e^{Δ}-βgal, were constructed as follows. The term e^{Δ} denotes the sequence HGSGAWLLPVSLVRS, a 13-residue derivative of the 40-residue e^K (residues 2–14), followed by two residues Arg and Ser, encoded by the *Bgl*II site (Figure 1B). Double-stranded oligonucleotides that encompassed the 5′ end of the *Sac*II site, the end of Ub-Arg-$(Lys/Asn)_{14}$-coding sequences (see below), and the 3′ end of the *Bam*HI site were digested with *Sac*II and *Bam*HI. The resulting fragments were ligated to the *Sac*II–*Bam*HI vector-containing fragment of pDhaUbReΔβgal. The latter plasmid was produced by inserting annealed complementary oligonucleotides EDKS1 and EDAS1, which encoded the e^{Δ} sequence HGSGAWLLPVSLVRS (see above), into *Bam*HI/*Bgl*II-cut pDhaUbReK-βgal. All constructs were verified by nucleotide sequencing.

A screen for N-degrons in the lysine–asparagine sequence space

The library of pDhaUbReΔβgal-derived plasmids was constructed that expressed, from the P_{CUP1} promoter, Arg-$(K/N)_{14}$-e^{Δ}-βgal proteins that contained random-sequence 14-residue Lys/Asn(K/N)-inserts between the N-terminal Arg and the e^{Δ} moiety (Figure 1B). Arg-$(K/N)_{14}$-e^{Δ}-βgals were the products of co-translational cleavage of DHFR-ha-Ub^{R48}-Arg-$(\underline{K/N})_{14}$-e^{Δ}-βgals (Figure 1B). To produce the library, a method for cloning random-sequence oligonucleotides produced by mutually primed synthesis was used (Oliphant *et al.*, 1986; Ghislain *et al.*, 1996). A set of oligonucleotides 5′-CGC$\underline{\text{CCGCGG}}$TGGTAGG$(AAA/T)_{14}CAC\underline{\text{GGATCC}}$G-3′ that contained random permutations of 14 codons, either AAA (encoding Lys) or AAT (encoding Asn), flanked by the *Sac*II and *Bam*HI sites (underlined), was synthesized. The oligonucleotides were converted into their double-stranded counterparts with Klenow Pol I (Ausubel *et al.*, 1996), and were digested with *Sac*II and *Bam*HI, yielding a set of equal-length fragments containing randomly permuted $(AAA/T)_{14}$ inserts. The fragments were ligated to the *Sac*II–*Bam*HI vector-containing fragment of pDhaUbReΔβgal. The resulting library, in the plasmid termed pRKN14, was introduced into *E.coli* DH5α by electropor-

ation (Ausubel *et al.*, 1996). Digestion of the pool of recovered plasmids with appropriate restriction enzymes showed that >90% of transformants contained an oligonucleotide-derived insert. The pRKN14 library was transformed into *S.cerevisiae* JD54, in which Ubr1p was expressed from the P_{GAL1} promoter. Approximately 2×10^4 transformants growing on SD(–Trp) plates containing 0.2 mM $CuSO_4$ were replica-plated onto $CuSO_4$-containing SG(–Trp) plates. The XGal-based filter assay (Ausubel *et al.*, 1996) was used to screen for colonies that were white (low levels of βgal) on SG(–Trp) plates but blue (high levels of βgal) on SD(–Trp) plates. These colonies were grown up in liquid SG(–Trp), and the activity of βgal was determined as described below. Plasmid DNA was isolated from the positive transformants, amplified in *E.coli* and transformed into *S.cerevisiae* JD47-13C (*UBR1*) and JD55 (*ubr1*Δ). The metabolic stabilities of the corresponding Arg-$(K/N)_{14}$-e^{Δ}-βgal proteins in *ubr1*Δ versus wild-type (*UBR1*) *S.cerevisiae* were determined by measuring βgal activity, and also directly, by carrying out pulse–chase assays.

Measurement of βgal activity

Saccharomyces cerevisiae cells were added to 5 ml of SD(–Trp) containing 0.2 mM $CuSO_4$, and grown to A_{600} ~1. Cells were gently pelleted by centrifugation, lysed with glass beads in 20% glycerol, 1 mM dithiothreitol, 0.1 M Tris–HCl pH 8, and the activity of βgal was measured in the clarified extract using *o*-nitrophenyl-β-D-galactoside, as described (Ausubel *et al.*, 1996). The activity was normalized to the total protein concentration, determined using the Bradford assay (Bio-Rad, Hercules, CA).

Pulse–chase assays

Transformed JD47-13C and JD55 cells from 10 ml cultures (A_{600} 0.5–1) in SD(–Trp) containing 0.2 mM $CuSO_4$ were gently pelleted by centrifugation, and washed in the same medium. The cells were resuspended in 0.3 ml of SD(–Trp) containing $CuSO_4$ and labeled for 5 min at 30°C with 0.15 mCi (5.5 MBq) of [^{35}S]methionine/cysteine (Trans^{35}S-label, ICN, Costa Mesa, CA). The cells were harvested by centrifugation, resuspended in 0.3 ml of SD(–Trp), 10 mM L-methionine, 0.5 mg/ml cycloheximide and incubated further at 30°C. At each time point, 0.1 ml samples were withdrawn and added to 0.7 ml of the lysis buffer (1% Triton X-100, 0.15 M NaCl, 1 mM EDTA, 50 mM Na–HEPES pH 7.5) containing 1 mM phenylmethylsulfonyl fluoride. The cells were then lysed by vortexing with 0.5 ml of 0.5 mm glass beads four times for 1 min, with intermittent cooling on ice, followed by centrifugation at 12 000 *g* for 10 min. The volumes of supernatants were adjusted to equalize the amounts of 10% trichloroacetic acid-insoluble ^{35}S, followed by immunoprecipitation with a mixture of saturating amounts of monoclonal antibodies against the ha epitope (Babco, Berkeley, CA) and βgal (Promega, Madison, WI). The samples were incubated at 4°C for 2 h, with rotation, followed by the addition of 10 µl of protein A–Sepharose suspension (Repligen, Cambridge, MA), further incubation for 1 h, and centrifugation at 12 000 *g* for 30 s. The immunoprecipitates were washed four times with 0.8 ml of the lysis buffer plus 0.1% SDS, resuspended in SDS-sample buffer (Ausubel *et al.*, 1996), heated at 100°C for 3 min and fractionated by SDS–10% PAGE, followed by autoradiography and quantitation with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA), using the reference provided by the UPR technique (Figures 2 and 5) (Lévy *et al.*, 1996).

Pulse–chase assays with JD54 cells, in which *UBR1 was* expressed from the P_{GAL1} promoter, were carried out by growing 20 ml cultures to A_{600} 0.5–1 in SR(–Trp), collecting and resuspending the cells in 0.5 ml of the same medium containing 0.2 mM $CuSO_4$ and labeling with 0.3 mCi of [^{35}S]methionine/cysteine for 10 min at 30°C. The cells were washed twice with SR(–Trp), and transferred to 0.5 ml of SR (–Trp) containing 10 mM L-methionine and 1 mM cysteine. After a further 20 min incubation at 30°C, galactose was added (to a final concentration of 3%) to induce the expression of *UBR1 and* initiate the chase, whose time points are indicated in Figures 2D and 3D.

ID^5, the extent of initial decay at the end of the 5 min pulse, was calculated as follows: $ID^5 = \{1 - [X\beta gal]_0/[Mete^K\beta gal]_0\} \times 100\%$. This parameter (the upper index refers to the length of pulse) equals 100% minus the ratio of ^{35}S in an X-βgal to ^{35}S in the reference protein dha-Ub (DHFR-ha-Ub^{R48}) at the end of the pulse, normalized against the same ratio with metabolically stable Met-e^K-βgal (Lévy *et al.*, 1996). To denote the observed half-lives of a test protein at different regions of a non-exponential decay curve, a generalized half-life term $t_{0.5}^{y-z}$ was used, in which 0.5 denotes the parameter's half-life aspect and *y–z* denotes the relevant time interval, from *y* to *z* min of chase.

Acknowledgements

We thank the current and former members of the Varshavsky laboratory, particularly F.Lévy, A.Webster and Y.Xie, for helpful discussions. We also thank F.Du, J.Sheng, H.-R.Wang, Y.Xie, H.Rao and especially G.Turner for comments on the manuscript. This work was supported by grants to A.V. from the National Institutes of Health (DK39520) and the US Army Breast Cancer Research Program (DAMD179818042).

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Smith,J.A., Seidman,J.G. and Struhl,K. (eds) (1996) *Current Protocols in Molecular Biology*. Wiley-Interscience, New York, NY.

Bachmair,A. and Varshavsky,A. (1989) The degradation signal in a short-lived protein. *Cell*, **56**, 1019–1032.

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

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

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

Baumeister,W., Walz,J., Zühl,F. and Seemüller,E. (1998) The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, **92**, 367–380.

Chau,V., Tobias,J.W., Bachmair,A., Marriott,D., Ecker,D.J., Gonda,D.K. and Varshavsky,A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. *Science*, **243**, 1576–1583.

Clarke,N.D. (1995) Sequence 'minimization': exploring the sequence landscape with simplified sequences. *Curr. Opin. Biotechnol*., **6**, 467–472.

Coux,O., Tanaka,K. and Goldberg,A.L. (1996) Structure and functions of the 20S and 26S proteasomes. *Annu. Rev. Biochem.*, **65**, 801–817.

deGroot,R.J., Rümenapf,T., Kuhn,R.J. and Strauss,J.H. (1991) Sindbis virus RNA polymerase is degraded by the N-end rule pathway. *Proc. Natl Acad. Sci. USA*, **88**, 8967–8971.

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

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

Grigoryev,S., Stewart,A.E., Kwon,Y.T., Arfin,S.M., Bradshaw,R.A., Jenkins,N.A., Copeland,N.G. and Varshavsky,A. (1996) A mouse amidase specific for N-terminal asparagine. The gene, the enzyme and their function in the N-end rule pathway. *J. Biol. Chem.*, **271**, 28521–28532.

Hershko,A. and Ciechanover,A. (1998) The ubiquitin system. *Annu. Rev. Biochem.*, **76**, 425–479.

Hill,C.P., Johnston,N.L. and Cohen,R.E. (1993) Crystal structure of a ubiquitin dependent degradation substrate: a three-disulfide form of lysozyme. *Proc. Natl Acad. Sci. USA*, **90**, 4136–4140.

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

Hochstrasser,M. (1996) Ubiquitin-dependent protein degradation. *Annu. Rev. Genet.*, **30**, 405–439.

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

Johnson,E.S., Bartel,B.W. and Varshavsky,A. (1992) Ubiquitin as a degradation signal. *EMBO J.*, **11**, 497–505.

Koepp,D.M., Harper,J.W. and Elledge,S.J. (1999) How the cyclin became a cyclin: regulated proteolysis in the cell cycle. *Cell*, **97**, 431–434.

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

Kwon,Y.T., Kashina,A.S. and Varshavsky,A. (1999) 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.*, **19**, 182–193.

Laney,J.D. and Hochstrasser,M. (1999) Substrate targeting in the ubiquitin system. *Cell*, **97**, 427–430.

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

Lévy,F., Johnston,J.A. and Varshavsky,A. (1999) Analysis of a conditional degradation signal in yeast and mammalian cells. *Eur. J. Biochem.*, **259**, 244–252.

Madura,K. and Varshavsky,A. (1994) Degradation of Ga by the N-end rule pathway. *Science*, **265**, 1454–1458.

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

Oliphant,A.R., Nussbaum,A.L. and Struhl,K. (1986) Cloning of random-sequence oligodeoxynucleotides. *Gene*, **44**, 177–183.

Peters,J.M. (1998) SCF and APC: the Yin and Yang of cell cycle regulated proteolysis. *Curr. Opin. Cell Biol.*, **10**, 759–768.

Pickart,C.M. (1997) Targeting of substrates to the 26S proteasome. *FASEB J.*, **11**, 1055–1066.

Rechsteiner,M. (1998) The 26S proteasome. In Peters,J.M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 147–189.

Sadis,S.C.A. and Finley,D. (1995) Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.*, **15**, 4086–4094.

Scheffner,M., Smith,S. and Jentsch,S. (1998) The ubiquitin conjugation system. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY, pp. 65–98.

Sijts,A.J., Pilip,I. and Pamer,E.G. (1997) 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.*, **272**, 19261–19268.

Stewart,A.E., Arfin,S.M. and Bradshaw,R.A. (1995) The sequence of porcine protein NH_2-terminal asparagine amidohydrolase. A new component of the N-end rule pathway. *J. Biol. Chem.*, **270**, 25–28.

Tobery,T. and Siliciano,R.F. (1999) Induction of enhanced CTL-dependent protective immunity *in vivo* by N-end rule targeting of a model tumor antigen. *J. Immunol.*, **162**, 639–642.

Townsend,A., Bastin,J., Gould,K., Brownlee,G., Andrew,M., Coupar,B., Boyle,D., Chan,S. and Smith,G. (1988) Defective presentation to class I-restricted cytotoxic T lymphocytes in vaccinia-infected cells is overcome by enhanced degradation of antigen. *J. Exp. Med.*, **168**, 1211–1224.

Tyers,M. and Willems,A.R. (1999) One ring to rule a superfamily of E3 ubiquitin ligases. *Science*, **284**, 602–604.

Varshavsky,A. (1996) The N-end rule: functions, mysteries, uses. *Proc. Natl Acad. Sci. USA*, **93**, 12142–12149.

Varshavsky,A. (1997) The ubiquitin system. *Trends Biochem. Sci.*, **22**, 383–387.

Wilkinson,K. and Hochstrasser,M. (1998) The deubiquitinating enzymes. In Peters,J.-M., Harris,J.R. and Finley,D. (eds), *Ubiquitin and the Biology of the Cell*. Plenum Press, New York, NY.

Worley,C.K., Ling,R. and Callis,J. (1998) Engineering *in vivo* instability of firefly luciferase and *Escherichia coli* beta-glucuronidase in higher plants using recognition elements from the ubiquitin pathway. *Plant Mol. Biol.*, **37**, 337–347.

Received August 18, 1999; revised September 15, 1999; accepted September 16, 1999