REVIEW

The N-end rule pathway of protein degradation

Alexander Varshavsky*

Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. Similar but distinct versions of the N-end rule operate in all organisms examined, from mammals to fungi and bacteria. In eukaryotes, the N-end rule pathway is a part of the ubiquitin system. Ubiquitin is a 76-residue protein whose covalent conjugation to other proteins plays a role in many biological processes, including cell growth and differentiation. I discuss the current understanding of the N-end rule pathway.

Introduction

Many intracellular proteins are metabolically unstable, or can become unstable during their lifetime in a cell. The functions of intracellular proteolysis include the elimination of abnormal proteins, the maintenance of amino acid pools in cells affected by stresses such as starvation, and the generation of protein fragments that act as hormones, antigens or other effectors. Yet another function of proteolytic pathways is the selective destruction of proteins whose concentrations must vary with time, and alterations in the state of a cell. Metabolic instability is a property of many regulatory proteins. A short *in vivo* half-life[a] of a regulator provides a way to generate its spatial gradients and allows for rapid adjustments of its concentration (or subunit composition) through changes in the rate of its synthesis. A protein can also be conditionally unstable, i.e., long-lived or short-lived depending on the state of a cell. Conditionally short-lived regulators are often deployed as components of control circuits. One example is cyclins—a family of related proteins whose destruction at specific stages of the cell cycle regulates cell division and growth (Murray & Hunt 1993). In addition, many proteins are long-lived as components of larger complexes such as ribosomes and oligomeric proteins, but are metabolically unstable as free subunits. The short *in vivo* lifetimes of free subunits allow for a less stringent control over the relative rates of their synthesis, since a subunit produced in excess would not accumulate to a significant level.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (Varshavsky 1991). The essential component of one degradation signal—the first to be discovered—is a destabilizing N-terminal residue of a protein (Bachmair *et al.* 1986; Varshavsky 1992, 1996a). This signal is called the N-degron. A set of N-degrons containing different destabilizing residues yields 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 (Table 1 and Fig. 1). The N-end rule pathway is present in all organisms examined, including the bacterium *Escherichia coli* (Tobias *et al.* 1991; Shrader *et al.* 1993), the yeast (fungus) *Saccharomyces cerevisiae* (Bachmair & Varshavsky 1989), and mammalian cells (Gonda *et al.* 1989; Lévy *et al.* 1996) (Fig. 1).

The N-end rule was encountered in experiments that explored the metabolic fate of a fusion between Ub and a reporter protein such as *E. coli* β-galactosidase (βgal) in *S. cerevisiae* (Bachmair *et al.* 1986). In yeast and other eukaryotes, Ub-X-βgal is cleaved, cotranslationally or nearly so, by Ub-specific processing proteases at the Ub-βgal junction. This cleavage takes place regardless

* *Correspondence*: E-mail: varshavskya@starbase1.caltech.edu

[a] The *in vivo* degradation of many short-lived proteins, including the engineered N-end rule substrates, deviates from first-order kinetics (Baker & Varshavsky 1991). Therefore the term 'half-life', if applied to an entire decay curve, is a useful but often crude approximation. A more rigorous terminology for describing nonexponential decay was proposed by Lévy *et al.* (1996).

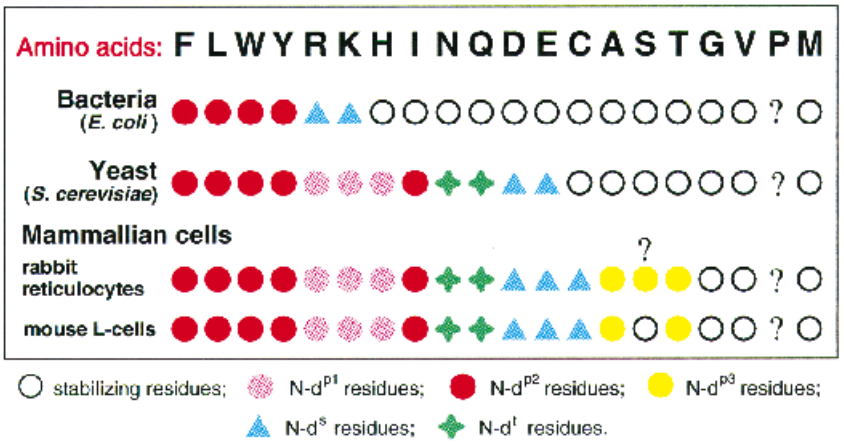

Figure 1 Comparison of eukaryotic and bacterial N-end rules. Open circles denote stabilizing residues. Purple and red circles denote, respectively, type 1 and type 2 primary destabilizing residues. Blue triangles denote secondary destabilizing residues. Green crosses denote tertiary destabilizing residues (Varshavsky 1996a). A question mark at Pro indicates its uncertain status (see the main text). A question mark above Ser indicates its uncertain status in the reticulocyte N-end rule (Lévy *et al.* 1996). Single-letter abbreviations for amino acids: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

of the identity of the residue X at the C-terminal side of the cleavage site, proline being the single exception. By allowing a bypass of the normal N-terminal processing of a newly formed protein, this finding (Fig. 2A) yielded an *in vivo* method for generating different residues at the N-termini of otherwise identical proteins—a technical advance that led to the N-end rule (Varshavsky 1992, 1996a).

In eukaryotes, the N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine (or lysines) of a substrate (Fig. 2B) (Bachmair & Varshavsky 1989; Johnson *et al.* 1990; Dohmen *et al.* 1994). The Lys residue is the site of formation of a multiubiquitin chain (Chau *et al.* 1989). Ubiquitin (Ub) is a 76-residue protein whose covalent conjugation to other proteins is involved in a multitude of processes—cell growth and differentiation, signal transduction, DNA repair, transmembrane traffic, and responses to stress, including the immune response. In many of these settings, Ub acts through routes that involve the processive degradation of Ub-protein conjugates (Hershko 1991; Jentsch 1992; Varshavsky 1996a).

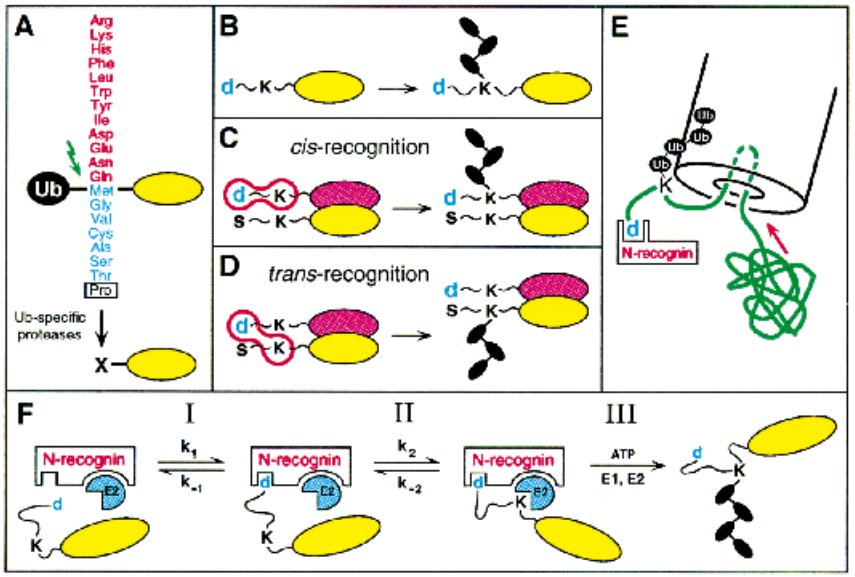

Table 1 The N-end rule in *E. coli* and *S. cerevisiae*

	Half-life of X-βgal	
Residue X	*E. coli*	*S. cerevisiae*
Arg	2 min	2 min
Lys	2 min	3 min
Phe	2 min	3 min
Leu	2 min	3 min
Trp	2 min	3 min
Tyr	2 min	10 min
His	> 10 h	3 min
Ile	> 10 h	30 min
Asp	> 10 h	3 min
Glu	> 10 h	30 min
Asn	> 10 h	3min
Gln	> 10 h	10 min
Cys	> 10 h	> 30 h
Ala	> 10 h	> 30 h
Ser	> 10 h	> 30 h
Thr	> 10 h	> 30 h
Gly	> 10 h	> 30 h
Val	> 10 h	> 30 h
Pro	?	> 5 h
Met	> 10 h	> 30 h

Approximate *in vivo* half-lives of X-βgal proteins in *E. coli* at 36 °C (Tobias *et al.* 1991) and in *S. cerevisiae* at 30 °C (Bachmair & Varshavsky 1989). A question mark at Pro indicates its uncertain status in the N-end rule (see the main text).

The binding of an N-end rule substrate by a targeting complex is followed by the formation of a substrate-linked multi-Ub chain (Dohmen *et al.* 1991). The ubiquitylated substrate is processively degraded by the 26S proteasome—an ATP-dependent, multisubunit protease (Rechsteiner *et al.* 1993). The N-end rule pathway is present in both the cytosol (Bachmair *et al.* 1986) and the nucleus (J.A. Johnston & A.V., unpublished data). In this paper, I summarize the current understanding of the N-end rule. For a more detailed review, see Varshavsky *et al.* (1997).

Components and evolution of the N-end rule pathway

The N-end rule is organized hierarchically. In *S. cerevisiae*, Asn and Gln are *tertiary* destabilizing N-terminal residues in that they function through their conversion, by enzymatic deamidation, into the *secondary* destabilizing N-terminal residues Asp and Glu, whose activity requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the *primary* destabilizing N-terminal residues (Gonda *et al.* 1989; Balzi *et al.* 1990; Baker & Varshavsky 1995). The primary destabilizing residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway (Fig. 3).

N-recognin (E3)

In *S. cerevisiae*, N-recognin is a 225 kDa protein (encoded by *UBR1*) that selects potential N-end rule

Figure 2 Mechanics of the N-end rule. (**A**) The ubiquitin fusion technique. Linear fusions of Ub to other proteins are cleaved at the last residue of Ub, making it possible to produce different residues at the N-termini of otherwise identical proteins (Bachmair *et al.* 1986; Gonda *et al.* 1989). Amino acid residues in blue and red are stabilizing and destabilizing, respectively, in the *S. cerevisiae* N-end rule (Bachmair & Varshavsky 1989). (**B**) The two-determinant organization of eukaryotic N-degrons. *d*, a destabilizing N-terminal residue. A chain of black ovals linked to the second-determinant lysine (*K*) denotes a multi-Ub chain. (**C**) *Cis* recognition of the N-degron in one subunit of a dimeric protein. The other subunit bears *s*, a stabilizing N-terminal residue. (**D**) *Trans* recognition, in which the first (*d*) and second (*K*) determinants of the N-degron reside in different subunits of a dimeric protein (Johnson *et al.* 1990). (**E**) The hairpin insertion model. A targeted N-end rule substrate (in green) bearing a multi-Ub chain is shown bound to the 26S proteasome through the chain. The position of a targeting complex containing N-recognin is unknown, and is left unspecified. Only the 20S core component of the 26S proteasome is shown. A red arrow indicates the direction of net movement of the substrate's polypeptide chain toward active sites in the interior of proteasome. By analogy with the arrangement of signal sequences during transmembrane translocation of proteins (Schatz & Dobberstein 1996), it is proposed that a region of the substrate upstream of its ubiquitylated lysine (K) does not move through the proteasome during the substrate's degradation, and may be released intact following a cleavage downstream of the lysine. Variants of this model may also be relevant to the targeting of proteins that bear internal or C-terminal degrons. (**F**) A model for the recognition of an N-end rule substrate (Bachmair & Varshavsky 1989). The reversible binding of N-recognin to a primary destabilizing N-terminal residue (*d*) of a substrate (step **I**) must be followed by a capture of the second-determinant lysine (*K*) of the substrate by a targeting complex containing a Ub-conjugating (E2) enzyme (step **II**). It is unknown whether the lysine is captured by E2 (as shown here) or by N-recognin. Ubiquitylation of the substrate commences once the targeting complex is bound to both determinants of the N-degron (step **III**). This model does not specify, among other things, the details of Ub conjugation (see the main text).

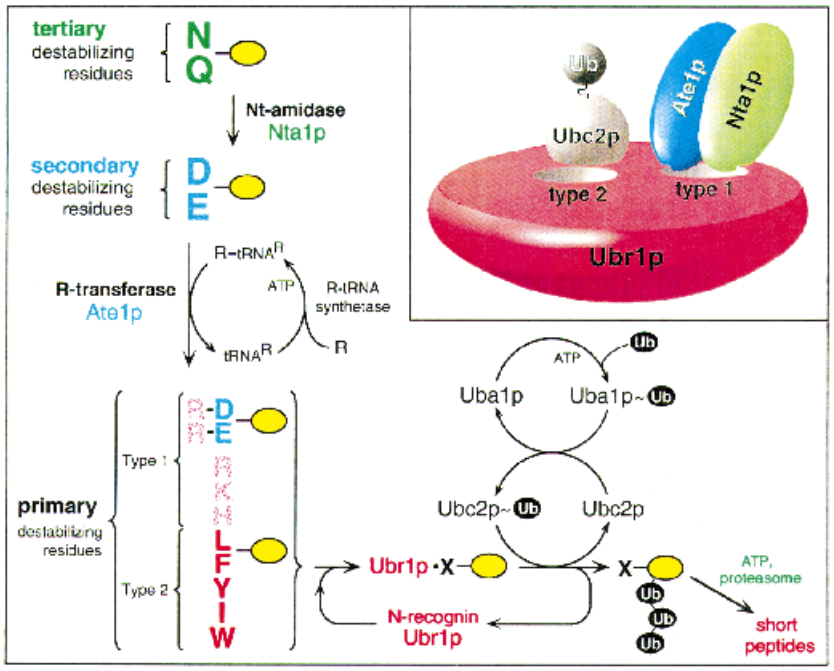

Figure 3 The *S. cerevisiae* N-end rule pathway. Type 1 and type 2 primary destabilizing N-terminal residues are in purple and red, respectively. Secondary and tertiary destabilizing N-terminal residues are in blue and green, respectively. The yellow ovals denote the rest of a protein substrate. The conversion of tertiary destabilizing residues **N** and **Q** into secondary destabilizing residues **D and E** is mediated by N-terminal amidohydrolase (Nt-amidase), encoded by *NTA1*. The conjugation of the primary destabilizing residue **R** to the secondary destabilizing residues **D** and **E** is mediated by Arg-tRNA-protein transferase (R-transferase), encoded by *ATE1*. A complex of N-recognin and the Ub-conjugating (E2) enzyme Ubc2p catalyses the conjugation of activated Ub, produced by the Ub-activating (E1) enzyme Uba1p, to a Lys residue of the substrate, yielding a substrate-linked multi-Ub chain. Uba1p~Ub and Ubc2p~Ub denote covalent (thioester-mediated) complexes of these enzymes with Ub. A multiubiquitylated substrate is degraded by the 26S proteasome. Inset: A model of the targeting complex. The 20 kDa Ubc2p E2 enzyme is depicted carrying activated Ub linked to Cys-88 of Ubc2p through a thioester bond. Both the 52 kDa Nta1p (Nt-amidase) and the 58 kDa Ate1p (R-transferase) bind to the 225 kDa Ubr1p (N-recognin) in proximity to the type 1 substrate-binding site of Ubr1p. In addition, Nta1p directly interacts with Ate1p (see the main text).

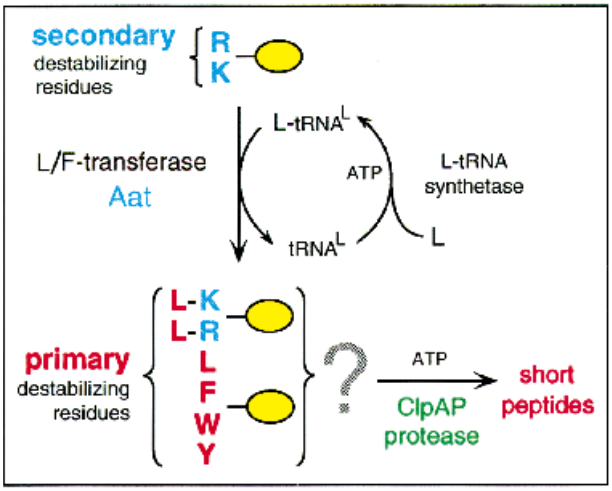

Figure 4 The *E. coli* N-end rule pathway. Primary destabilizing N-terminal residues **L**, **F**, **W** and **Y** are in red. Secondary destabilizing N-terminal residues **R** and **K** are in blue. The yellow ovals denote the rest of a protein substrate. Conjugation of the primary destabilizing residue **L** to the secondary destabilizing residues **R** and **K** is mediated by Leu, Phe-tRNA-protein transferase (L/F-transferase), encoded by *aat* (Tobias *et al.* 1991). *In vivo*, L/F-transferase appears to conjugate predominantly, if not exclusively, **L** (Shrader *et al.* 1993). The degradation of a substrate bearing a primary destabilizing N-terminal residue is carried out by the ATP-dependent protease ClpAP, encoded by *clpA* and *clpP*. A question mark denotes an ambiguity about the nature of N-recognin in *E. coli*.

substrates through the binding to their primary destabilizing N-terminal residues Phe, Leu, Trp, Tyr, Ile, Arg, Lys or His (Bartel *et al.* 1990; Varshavsky 1996a). N-recognin has at least two substrate-binding sites. The type 1 site is specific for the basic N-terminal residues Arg, Lys and His. The type 2 site is specific for the bulky hydrophobic N-terminal residues Phe, Leu, Trp, Tyr and Ile (Fig. 3). At present, these sites are defined through dipeptide-based competition experiments. Specifically, a dipeptide bearing a destabilizing N-terminal residue was found to inhibit the degradation of a test N-end rule substrate if that substrate's N-terminal residue was of the same type (1 or 2) as the dipeptide's N-terminal residue (Reiss *et al.* 1988; Gonda *et al.* 1989).

A genetic dissection of the type 1 and type 2 sites in *S. cerevisiae* N-recognin (Ubr1p) has shown that either of the sites can be mutationally inactivated without significantly perturbing the other site. Mutations that selectively inactivate the type 1 or the type 2 site are located within the ≈50 kDa N-terminal region of the 225 kDa N-recognin (A. Webster, M. Ghislain & A. V., unpublished data). E3α, the mammalian counterpart of *S. cerevisiae* N-recognin, has been characterized biochemically in extracts from rabbit reticulocytes (Hershko 1991). Another mammalian N-recognin, termed E3β, which apparently binds to substrates bearing N-terminal Ala and Thr (and possibly also Ser (Lévy *et al.* 1996)), has been described as well (Hershko 1991).

All eukaryotes examined have both Ub and the N-end rule pathway. Some, but not all, prokaryotes contain Ub (Wolf *et al.* 1993). The bacterium *E. coli* lacks Ub but does have an N-end rule pathway (Fig. 4) (Tobias *et al.* 1991). Screens for mutations that inactivate either the entire N-end rule pathway or its subset have identified three *E. coli* genes—*clpA*, *clpP*, and *aat* (Shrader *et al.* 1993). Aat is a Leu, Phe-tRNA-protein transferase (L/F-transferase). ClpA (81 kDa) and ClpP (21 kDa) form a ≈750 kDa complex, ClpAP, which exhibits ATP-dependent protease activity *in vitro* (Gottesman & Maurizi 1992), and is a functional counterpart of the eukaryotic 26S proteasome in the *E. coli* N-end rule pathway (Fig. 4).

ClpP exhibits a chymotrypsin-like protease activity *in vitro* (Gottesman & Maurizi 1992). ClpA is the ATP-binding component of ClpAP. *In vitro* studies have shown that ClpA can act as a chaperone in the activation of RepA, the replication initiator encoded by the plasmid P1 (Wickner *et al.* 1994). *In vivo* ramifications of these results, and in particular their relevance to the proteolytic function of ClpAP in the *E. coli* N-end rule pathway (Fig. 4), remain to be examined. ClpP associates not only with ClpA (forming ClpAP protease), but also with the ClpA homologues ClpB or ClpX, forming, respectively, ClpBP or ClpXP proteases (Gottesman *et al.* 1993; Wawrzynov *et al.* 1995). In contrast to ClpA, whose mutational elimination stabilizes the normally short-lived N-end rule substrates (Tobias *et al.* 1991), the elimination of either ClpB or ClpX appears not to perturb the *E. coli* N-end rule pathway (O. Lomovskaya & A. V., unpublished data).

N-terminal amidases

The *S. cerevisiae* N-terminal amidohydrolase (Nt-amidase), encoded by *NTA1*, is a 52 kDa enzyme which deamidates Asn or Gln if, and only if, they are located at the N-terminus of a polypeptide (Figs 3 and 5B) (Baker & Varshavsky 1995; Grigoryev *et al.* 1996). Null *nta1* mutants are unable to degrade N-end rule substrates that bear N-terminal Asn or Gln. The deduced sequence of Nta1p is not similar to those of the other known amidotransferases, save for the sequence Gly Ile-Cys-Met that is a part of an 11-residue region conserved among some, but not all, amidotransferases. The conserved cysteine of this sequence is required for the enzymatic activity of Nta1p (Grigoryev *et al.* 1996).

Stewart *et al.* (1995) purified a porcine Nt-amidase that deamidates N-terminal Asn (N) but not Gln (Q), and isolated a cDNA that encodes this enzyme. Grigoryev *et al.* (1996) isolated and characterized a ≈17 kb gene, termed *Ntan1*, that encodes a mouse homologue of the porcine amidase, termed Nt^N-amidase. The ≈1.4 kb *Ntan1* mRNA is expressed in all of the tested mouse tissues and cell lines. The recently produced *ntan1*Δ mouse strains are viable (Y.T. Kwon and A.V., unpublished data). Their phenotypic analysis is under way.

Both Asn and Gln are destabilizing residues in the mammalian N-end rule (Fig. 1). Further, both N-terminal Asn and Gln of the test proteins are deamidated in mammalian cell extracts (Gonda *et al.* 1989; S. Grigoryev & A.V., unpublished data). Therefore there must exist yet another mammalian Nt-amidase (Nt^Q-amidase), which can deamidate N-terminal Gln (Fig. 5A).

Aminoacyl-tRNA-protein transferases

The *S. cerevisiae* Arg-tRNA-protein transferase (R-transferase), encoded by *ATE1*, is a 58 kDa enzyme that utilizes Arg-tRNA to arginylate the N-termini of polypeptides (but not free amino acids) that bear Asp or Glu (Figs 3 and 5B). Null *ate1* mutants are unable to

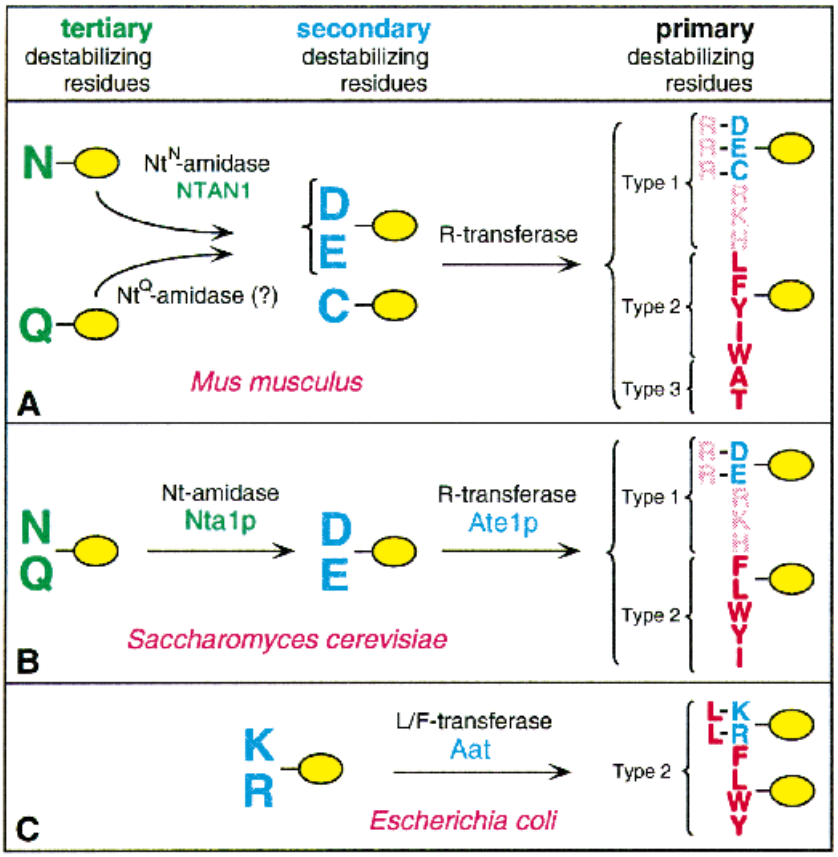

Figure 5 Comparison of enzymatic reactions that underlie the activity of tertiary and secondary destabilizing residues in different organisms. (**A**) Mouse (*Mus musculus*) L-cells and rabbit (*Oryctolagus cuniculus*) reticulocytes (Lévy *et al.* 1996; Gonda *et al.* 1989). (**B**) The yeast *Saccharomyces cerevisiae* (Bachmair & Varshavsky 1989). (**C**) The bacterium *Escherichia coli* (Tobias *et al.* 1991). The *E. coli* N-end rule lacks tertiary destabilizing residues. The postulated mammalian Nt^Q-amidase (a question mark in **A**) remains to be identified.

degrade N-end rule substrates that bear N-terminal Asn, Gln, Asp or Glu—the tertiary and secondary destabilizing N-terminal residues (Balzi *et al.* 1990). In contrast to *S. cerevisiae*, where only Asp and Glu are secondary destabilizing residues, in mammals, Cys is a secondary destabilizing residue as well (Gonda *et al.* 1989) (Table 1 and Fig. 1). It is not known whether the arginylation of N-terminal Asp, Glu and Cys in mammals is catalysed by an R-transferase whose specificity is broader than that of its yeast counterpart, or whether N-terminal Cys is arginylated by a distinct R-transferase (Fig. 5).

Approximately 2 h after a crush injury to the rat sciatic nerve, an extract was prepared from a segment of the nerve immediately upstream of the crush site. This extract was found to conjugate a ≈10-fold higher amount of the added ^{3}H-arginine to the N-termini of unidentified endogenous proteins than an otherwise identical extract from the same region of an unperturbed sciatic nerve (Dayal *et al.* 1990), suggesting a crush-induced increase in the level of N-end rule substrates and/or a post-crush induction of the N-end rule pathway. No post-crush increase in arginylation was observed with extracts from the optic nerve, which does not regenerate after a crush injury, in contrast to the sciatic nerve (Dayal *et al.* 1990).

R-transferase appears to be confined to eukaryotes, whereas Leu, Phe-tRNA-protein transferase (L/F-transferase) is present in bacteria such as *E. coli* but is apparently absent from eukaryotes. *E. coli* L/F-transferase is a 27 kDa enzyme encoded by the gene *aat* (Shrader *et al.* 1993). *In vivo*, L/F-transferase conjugates mainly if not exclusively the Leu to N-terminal Arg or Lys of a polypeptide substrate (Shrader *et al.* 1993) (Fig. 4). *E. coli* mutants lacking *aat* are unable to degrade N-end rule substrates that bear N-terminal Arg or Lys. These data (Tobias *et al.* 1991) identified L/F-transferase as a component of the *E. coli* N-end rule pathway.

Ubiquitin-conjugating enzymes

The initial interaction between an N-end rule substrate and N-recognin is of moderate affinity (the inferred K_d of roughly 10 μM; Varshavsky *et al.* 1997), but becomes much stronger if an internal lysine of the

substrate is captured by a targeting complex containing a Ub-conjugating (E2) enzyme and N-recognin (E3). This capture initiates a processive synthesis of a lysine-linked multi-Ub chain. The E2 enzymes utilize activated Ub, produced by the Ub-activating (E1) enzyme, to catalyse the formation of isopeptide bonds between the C-terminal Gly 76 of Ub and ε-amino groups of lysines in acceptor proteins (Fig. 3) (Pickart 1988; Hershko 1991; Jentsch 1992).

In at least some Ub-dependent systems (Scheffner *et al.* 1995), including apparently the N-end rule pathway (V. Chau & A.V., unpublished data), the pathway-specific Ub ligase—a complex of a recognin (E3) and an E2 enzyme—shifts the activated Ub moiety (which is initially linked to a Cys residue of the E1 enzyme) through a relay of Ub thioesters before conjugating Ub to a Lys residue of a targeted substrate. In a substrate-linked multi-Ub chain, the C-terminal glycine of one Ub moiety is joined to an internal lysine of the adjacent Ub moiety, resulting in a chain of Ub-Ub conjugates. In a multi-Ub chain linked to an N-end rule substrate, only Lys-48 of Ub was found to be joined to another Ub moiety within a chain (Chau *et al.* 1989). Recently, multi-Ub chains mediated by Lys-63, Lys-29, Lys-11 or Lys-6 of Ub have been described as well (Arnason & Ellison 1994; Johnson *et al.* 1995; Spence *et al.* 1995; Baboshina & Haas 1996). It is not known whether these chains play a role in the N-end rule pathway.

In *S. cerevisiae*, the ubiquitylation of N-end rule substrates requires the Ubc2p E2 enzyme (Dohmen *et al.* 1991). Processes known to be perturbed by mutations in *UBC2* include the N-end rule pathway, DNA repair, induced mutagenesis, cell cycle control, and sporulation (Jentsch 1992, and references therein). The N-end rule pathway is inactive in both *ubr1*Δ and *ubc2*Δ mutants (Madura *et al.* 1993), but the overall effect of *ubc2*Δ on cell growth and sporulation is much more severe than that of *ubr1*Δ, indicating that the functions of Ubc2p are not confined to the N-end rule pathway.

The N-end rule as a witness of evolution

The organization of N-end rules, with their tertiary, secondary and primary destabilizing residues, is a feature more conserved in evolution than either the Ub dependence of an N-end rule pathway or the identity of enzymatic reactions that mediate the hierarchy of destabilizing residues. For example, in a bacterium such as *E. coli*, which lacks the Ub system, the N-end rule has both secondary and primary destabilizing residues (it lacks tertiary residues) (Figs 1, 4 and 5C). The identities of secondary destabilizing residues in *E. coli* (Arg and Lys) are different from those in eukaryotes (Figs 1 and 5). Bacterial and eukaryotic enzymes that implement the coupling between secondary and primary residues are also different: L/F-transferase in *E. coli* and R-transferase in eukaryotes. Note, however, that bacterial L/F-transferase and eukaryotic R-transferase catalyse reactions of the same type (conjugation of an amino acid to an N-terminal residue of a polypeptide) and utilize the same source of activated amino acid (aminoacyl-tRNA) (Fig. 5).

The apparent confinement of R-transferase to eukaryotes and of L/F-transferase to prokaryotes suggests that secondary destabilizing residues were recruited late in the evolution of N-end rule, after the divergence of prokaryotic and eukaryotic lineages. The lack of sequence similarity between the yeast Nt-amidase and the mammalian Nt^N-amidase, as well as the more narrow specificity of the mammalian enzyme (Fig. 5A, B) suggest that tertiary destabilizing residues Asn and Gln became a part of the N-end rule much later yet, possibly after the divergence of metazoan and fungal lineages. If so, the N-end rule pathway may be an especially informative witness of evolution: the ancient origins of this proteolytic system, the simplicity and discreteness of changes in the rule books of N-end rules among different species, and the diversity of proteins that either produce or target the N-degron should facilitate phylogenetic deductions—once the components of this pathway become characterized across a broad range of organisms.

Code vs. hardware

A given N-end rule is defined operationally—for a set of proteins such as X-βgals that differ exclusively by their N-terminal residues. Existing evidence (Bachmair & Varshavsky 1989) suggests that the ranking aspect of an N-end rule, i.e., an ordering of relative destabilizing activities among 20 fundamental amino acids, is invariant from one protein reporter to another in a given intracellular compartment. (The case of N-end rule substrates bearing N-terminal Pro presents an apparent exception to this conjecture; see below.) By contrast, the actual *in vivo* half-lives may differ greatly among *different* proteins bearing one and the same N-terminal residue (Bachmair & Varshavsky 1989). The cause of these differences is the multicomponent nature of underlying N-degrons (Fig. 2B). For example, in eukaryotes, an N-degron comprises not only a destabilizing N-terminal residue of a

protein but also its internal lysine (or lysines), whose quality as a determinant can range from high to nonexistent.

A priori, one and the same N-end rule can be implemented through vastly different assortments of targeting hardware. At one extreme, each destabilizing N-terminal residue may be bound by a distinct N-recognin. Conversely, a single N-recognin may be responsible for the entire rule book of destabilizing residues in a given N-end rule. The actual N-end rule pathways lie between these extremes, and happen to have a hierarchic rather than 'linear' structure (Figs 3–5).

Targeting complex of the N-end rule pathway

The known components of the *S. cerevisiae* N-end rule pathway that mediate steps prior to the proteolysis of a targeted substrate by the 26S proteasome are Nt-amidase (Nta1p), R-transferase (Ate1p), N-recognin (Ubr1p), a Ub-conjugating (E2) enzyme (Ubc2p), and the Ub-activating (E1) enzyme (Uba1p) (Fig. 3) (Varshavsky 1996a). In addition to direct (immunoprecipitation-based) evidence for the physical association between N-recognin and Ubc2p (Madura *et al.* 1993), there is also circumstantial (overexpression-based) evidence for the existence of a complex between N-recognin, R-transferase and Nt-amidase (Baker & Varshavsky 1995). Recently, a high–affinity interaction between Nta1p and Ate1p was demonstrated directly; other data suggest that both Nta1p and Ate1p interact with Ubr1p (M. Ghislain, A. Webster & A.V., unpublished results). In a quaternary Ubc2p-Ubr1p-Nta1p-Ate1p complex suggested by these data, Ate1p and Nta1p interact with each other and Ubr1p (Fig. 3).

Other, perhaps more transient, components of the targeting complex in *S. cerevisiae* are likely to include the 114 kDa Uba1p (E1 enzyme, which must be bound to Ubc2p during the E1→E2 transfer of activated Ub moiety), and also Arg-tRNA synthetase. The latter possibility is suggested by the finding that, in mammals, Arg-tRNA synthetase (whose product, Arg-tRNA, is a co-substrate of R-transferase) copurifies with R-transferase (Ciechanover *et al.* 1988). It is also likely that the targeting complex interacts with the 26S proteasome *in vivo*, for example, during the transfer of a multiubiquitylated N-end rule substrate to substrate-binding sites of the proteasome. The proteolytic machine that implements the N-end rule is thus a strikingly diverse assembly of enzymes and binding factors whose total mass is close to that of the large ribosomal subunit. However, even a transient existence of this 'metacomplex' is conjectural at present, the alternative possibility being a sequential formation of transient subcomplexes that produce a substrate-linked multi-Ub chain and relay a substrate toward the 26S proteasome.

The effects of overexpressing Nt-amidase and/or R-transferase in *S. cerevisiae* not only suggested the existence of the Nta1p-Ate1p-Ubr1p-Ubc2p complex but also led to the prediction that Nta1p and Ate1p are associated with Ubr1p in proximity to its type 1 substrate-binding site (Fig. 3) (Baker & Varshavsky 1995). The 'proximity' aspect of the postulated complex was invoked to account for the markedly different effects of overexpressed R-transferase on the degradation of N-end rule substrates bearing type 1 vs. type 2 primary destabilizing N-terminal residues (Baker & Varshavsky 1995). In the diagram of Fig. 3, the physical proximity of the bound R-transferase to the type 1 site of N-recognin is presumed to decrease the steric accessibility of this site to an N-end rule substrate that bears a type 1 residue such as Arg and approaches the type 1 binding site of N-recognin directly from the bulk solvent. By contrast, a substrate that acquired Arg through arginylation by the N-recognin-bound R-transferase would be able to reach the (nearby) type 1 binding site of N-recognin directly—without dissociating into the bulk solvent first—a feature known as substrate 'channeling' in multistage enzymatic reactions (Negrutskii & Deutscher 1991). The mechanics of channeling may involve the diffusion of an N-end rule substrate in proximity to surfaces of the targeting complex, analogous to the mechanism of a bifunctional enzyme dihydrofolate reductase-thymidylate synthetase, where the channeling of dihydrofolate apparently results from its movement across the surface of the protein (Knighton *et al.* 1994).

The N-degron and pre-N-degron

Nascent proteins contain N-terminal Met (fMet in prokaryotes), which is a stabilizing residue in the known N-end rules (Fig. 1). Thus, the N-degron of an N-end rule substrate must be produced from a pre-N-degron. In an engineered N-end rule substrate, a pre-N-degron contains the N-terminal Ub moiety whose removal by Ub-specific proteases yields the protein's N-degron (Fig. 2A). This design of a pre-N-degron is unlikely to be relevant to physiological N-end rule substrates, because natural Ub fusions (including the precursors of Ub) either contain a stabilizing residue at the Ub-protein junction or bear a mutant Ub moiety that is

retained *in vivo* (Özkaynak *et al.* 1987; Finley *et al.* 1989; Watkins *et al.* 1993). The known Met-aminopeptidases remove N-terminal Met if, and only if, the second residue of a protein is stabilizing in the yeast N-end rule (Fig. 1). The structural basis of this selectivity is the size of a residue's side chain (Sherman *et al.* 1985; Arfin & Bradshaw 1988; Li & Chang 1995). Specifically, the side chains of the residues that are destabilizing in the yeast N-end rule are larger than those of stabilizing residues. The exception is Met—a bulky hydrophobic but stabilizing residue (Fig. 1).

Can there be just one or a few residues between N-terminal Met and the site of cleavage that produces an N-degron? If so, a short (≤10 residues) N-terminal sequence might contain both the recognition motif and the cleavage site(s) for a relevant (unknown) processing protease. Screens for such sequences, carried out in *S. cerevisiae* (Sadis *et al.* 1995; Ghislain *et al.* 1996), did identify short (≤10 residues) N-terminal regions that conferred Ubr1p-dependent metabolic instability on a reporter protein. Most of the sequences identified by these screens were not similar to each other, possibly because a very large number of 10-residue N-terminal extensions can produce an N-degron *in vivo*, analogous to a large number of N-terminal sequences that can function as signals for protein translocation across the ER membrane (Kaiser *et al.* 1987).

Analysis of one N-terminal extension identified by Ghislain *et al.* (1996) has shown that it targets a reporter protein for degradation while retaining its N-terminal Met (M. Gonzalez, F. Lévy, M. Ghislain & A.V., unpublished data). This finding suggests that N-recognin binds not only to N-degrons but also to a degron that consists of an entirely internal sequence motif. By contrast, two other examined (directly sequenced) extensions were found to be cleaved *in vivo* after N-terminal Met, yielding destabilizing N-terminal residues (Sadis *et al.* 1995; Ghislain *et al.* 1996). In sum, we are just beginning to understand the processing reactions that yield a destabilizing N-terminal residue in a non-polyprotein context.

Mechanics of N-degron

Stochastic capture model

Studies with βgal- and DHFR-based N-end rule substrates (Bachmair & Varshavsky 1989; Chau *et al.* 1989; Johnson *et al.* 1990; Hill *et al.* 1993) suggested a stochastic view of the N-degron, in which specific lysines of an N-end rule substrate can be assigned a probability of being utilized as a ubiquitylation site. This probability depends on the time-averaged spatial position and mobility of a protein's lysine. For some, and often for most of the lysines in an N-end rule substrate, the probability of serving as a ubiquitylation site would be negligible because of the lysine's lack of mobility and/or its distance from a destabilizing N-terminal residue. In this 'stochastic capture' model (Fig. 2F), the folded conformation of a substrate would be expected to slow down or preclude the search for a Lys residue, unless it is optimally positioned in the folded substrate.

The bipartite design of N-degron (Fig. 2B) is also likely to be characteristic of other Ub-dependent degradation signals—present in a multitude of naturally short-lived proteins that include cyclins (Murray & Hunt 1993), IκBα, and c-Jun (Pahl & Baeuerle 1996). The first component of these degrons is the internal region of a protein (instead of its N-terminal residue) that is specific for each degradation signal. The second component is an internal lysine (or lysines). A degron may also contain regulatory determinants whose modification (e.g. phosphorylation/dephosphorylation) can modulate the activity of this degron (Pahl & Baeuerle 1996; Nishizawa *et al.* 1993).

Cis-trans recognition and subunit-specific degradation of oligomeric proteins

The two determinants of N-degron can be recognized either in *cis* or in *trans* (Fig. 2C, D) (Johnson *et al.* 1990; F. Lévy & A. V., unpublished data). Experiments that revealed the *trans*-recognition have also brought to light a remarkable feature of the N-end rule pathway: only those subunits of an oligomeric protein that contain the ubiquitylation site (but not necessarily a destabilizing N-terminal residue) are actually degraded (11).

What might be the mechanism of subunit-specific proteolysis? A 'simple' model is suggested by the binding of a substrate-linked multi-Ub chain to a component of the proteasome. Specifically, a subunit of an oligomeric substrate bound to the proteasome through a subunit-linked multi-Ub chain may be the only subunit that undergoes further mechanochemical processing by chaperone-like, ATP-dependent components of the 26S proteasome. These components mediate the unfolding and translocation steps that cause a movement of the subunit toward active sites in the proteasome's interior, and in the process dissociate this subunit from the rest of oligomeric substrate. In this mechanism, the initial binding of N-recognin to another subunit—that which bears a destabilizing

N-terminal residue but not the lysine determinant (Fig. 2C)—may be either too transient (lasting, in a 'productive' engagement, only long enough for a lysine to be captured on a nearby subunit) or sterically unfavourable for the delivery of this subunit to the interior of the proteasome.

Since other Ub-dependent degradation signals appear to be organized similarly to the N-degron (a 'primary' recognition determinant plus an internal lysine or lysines), subunit selectivity is likely to be a general feature of proteolysis by the Ub system (Varshavsky 1996a). Examples of physiologically relevant subunit-selective proteolysis include the degradation of p53 in a complex with the papilloma viral protein E6 (Scheffner *et al.* 1995) and the degradation of a cyclin in a complex with a cyclin-dependent kinase (Murray & Hunt 1993).

The hairpin insertion model and the function of multiubiquitin chain

Formation of a substrate-linked multi-Ub chain produces an additional binding site (or sites) for components of the proteasome. The resulting increase in affinity, i.e. a decrease in the rate of dissociation of the proteasome–substrate complex, can be used to facilitate proteolysis. Suppose that a rate-limiting step which leads, several steps later, to the first proteolytic cleavage of the proteasome-bound substrate is an unfolding (driven by thermal fluctuations) of a relevant region of the substrate. If so, an increase in stability of the proteasome–substrate complex, brought about by the multi-Ub chain, should facilitate the substrate's degradation, because the longer the allowed 'waiting' time, the greater the probability of a required unfolding event. Another (not mutually exclusive) possibility is that a substrate-linked multi-Ub chain acts as a proximity trap for partially unfolded states of a substrate. This might be achieved through reversible interactions of the chain's Ub moieties with regions of the substrate that undergo local unfolding. A prediction common to both models is that the degradation of a substrate whose conformation poses less of a kinetic impediment to the proteasome should be less dependent on Ub and ubiquitylation than the degradation of an otherwise similar but more stably folded substrate.

How is a proteasome-bound, ubiquitylated protein directed to the interior of the proteasome? This problem is analogous to that in studies of transmembrane channels for protein translocation (Simon & Blobel 1991; Schatz & Dobberstein 1996). Could the solutions be similar in these systems, reflecting, perhaps, a common ancestry of translocation channels and proteasomes? The model in Fig. 2E proposes, by analogy with translocation systems, a 'hairpin' insertion mechanism for the initiation of proteolysis by the 26S proteasome. A biased random walk ('thermal ratchet') that is likely to underlie the translocation of proteins across membranes (Simon & Blobel 1991) may also be responsible for the movement of the substrate's polypeptide chain through the proteasome, with cleavage products diffusing out from the proteasome's distal end and thereby contributing to the net bias in the chain's bidirectional saltations through the proteasome channel. One prediction of the hairpin insertion model for an N-end rule substrate whose N-degron's determinants are located upstream of the hairpin insertion site is that the substrate's N-terminal region (Fig. 2E) is likely be cleaved-off at later stages of targeting, and is therefore likely to be spared from the proteasome-mediated degradation.

Two findings indicate that the unfolding of a targeted N-end rule substrate is a prerequisite for its degradation by the 26S proteasome. Methotrexate—a folic acid analogue and high-affinity ligand of DHFR—can inhibit the degradation of an N-end rule substrate such as Arg-DHFR by the N-end rule pathway (Johnston *et al.* 1995). This result suggests that a critical post-ubiquitylation step faced by the proteasome includes a 'sufficient' conformational perturbation of the proteasome-bound substrate. Furthermore, it was shown that the N-end rule-mediated degradation of a 17 kDa N-terminal fragment of the 70 kDa Sindbis virus polymerase is not precluded by the conversion of all of the fragment's 10 Lys residues into Arg residues, which cannot be ubiquitylated (T. Rümenapf, J. Strauss & A.V., unpublished data). Thus, the ubiquitylation requirement of previously studied N-end rule substrates may be a consequence of their relatively stable conformations. The binding of a largely unfolded substrate (such as a fragment of Sindbis polymerase) by the targeting complex of the N-end rule pathway may be sufficient for the delivery of the substrate to the proteasome's active sites in the absence of a multi-Ub chain. In the language of models in Fig. 2E, F, the 'waiting' time for a bound and conformationally unstable substrate may be short enough not to require the formation of a dissociation-slowing device such as a multi-Ub chain.

The N-end rule without ubiquitin

No Ub-like covalent modification of N-end rule substrates has been detected in *E. coli*, in contrast to ubiquitylation of the same substrates in eukaryotes.

Moreover, the conversion of ubiquitylation-site lysines of an N-end rule substrate into arginines rendered the substrate long-lived in eukaryotes but did not impair its degradation in *E. coli* (Tobias *et al.* 1991). Thus, *E. coli* not only lacks a homologue of eukaryotic Ub, but also lacks the requirement for a lysine-specific modification of a substrate. Bacteria may contain proteins whose function in the N-end rule pathway is Ub-like but involves a noncovalent, lysine-independent binding to a targeted substrate. The proposed role of a substrate-linked multi-Ub chain in 'marking' a subunit of a protein for selective destruction leads to another testable conjecture: if a subunit-marking device is absent from the *E. coli* N-end rule pathway, the latter may be incapable of degrading an oligomeric protein 'one subunit at a time'.

Substrates and functions of the N-end rule pathway

The N-end rule and osmoregulation in yeast

A synthetic lethal screen was used to isolate an *S. cerevisiae* mutant, termed *sln1* (for 'synthetic lethal of N-end rule'), whose viability requires the presence of *UBR1* (Ota & Varshavsky 1992). *SLN1* has been found to encode a eukaryotic homologue of two-component regulators—a large family of proteins previously encountered only in bacteria (Ota & Varshavsky 1993). The properties of *S. cerevisiae* Sln1p are consistent with it being a sensor component of the osmoregulatory (HOG) pathway—a MAP kinase cascade (Maeda *et al.* 1994). Since an otherwise lethal hypomorphic mutation in *SLN1* can be suppressed by the presence of Ubr1p (N-recognin) (Ota & Varshavsky 1993), it is likely that one or more of the proteins (e.g., kinases) whose activity is down-regulated by Sln1p can also be down-regulated through their degradation by the N-end rule pathway. The relevant physiological N-end rule substrate(s) remains to be identified.

The N-end rule and the import of peptides

Alagramam *et al.* (1995) have found that *ubr1*Δ yeast cells are unable to import di- and tripeptides. They have also shown that *ubr1*Δ cells, unlike the congenic *UBR1* cells, contain virtually no *PTR2* mRNA that encodes a peptide transporter, an integral plasma membrane protein. Recent results (C. Byrd & A.V., unpublished data) indicated that the control of *PTR2* expression by Ubr1p (N-recognin) involves the Ub-conjugating (E2) enzyme Ubc2p, a known component of the N-end rule pathway (Fig. 3). The Ubc4p E2 enzyme can partially compensate for the absence of Ubc2p; deletion of both *UBC2* and *UBC4* results in cells that do not express Ptr2p and are unable to import peptides, similarly to *ubr1*Δ cells. Ubc4p has not been previously identified as a component of the N-end rule pathway.

The findings of Alagramam *et al.* (1995), and the observed dependence of peptide import on two specific ubiquitin-conjugating enzymes can be accounted for by a model in which the expression of the Ptr2p transporter is regulated by a short-lived repressor that is degraded by the N-end rule pathway. One prediction of this model is that a mutational inactivation of the repressor would bypass the requirement for Ubr1p in the import of peptides. Using a screen for such mutants, we isolated a gene called *CUP9* (C. Byrd & A.V., unpublished data). Its product is a homeodomain-containing, short-lived protein whose degradation is carried out largely by the N-end rule pathway. Overexpression of Cup9p inhibits the import of peptides. Conversely, *cup9*Δ cells express Ptr2p and import peptides at higher rates than *CUP9* cells. Moreover, *cup9*Δ cells can import peptides in the absence of *UBR1*, whereas the import by *CUP9* cells requires *UBR1*. These findings (C. Byrd & A.V., unpublished data) strongly suggest that Cup9p is the postulated short-lived repressor which controls the rate of peptide import by regulating the expression of the Ptr2p transporter. Remarkably, an earlier study (Knight *et al.* 1994) identified Cup9p as a protein required for an aspect of resistance to copper toxicity in *S. cerevisiae*. Thus, one and the same physiological substrate of the N-end rule pathway functions as both a repressor of peptide import and a regulator of copper homeostasis.

Gα subunit of G protein

Overexpression of the N-end rule pathway was found to inhibit the growth of haploid but not diploid cells (Madura & Varshavsky 1994). This ploidy-dependent toxicity was traced to the enhanced degradation of Gpa1p, the Gα subunit of the G protein that regulates cell differentiation in response to mating pheromone. The half-life of newly formed Gα at 30 °C is ≈50 min in wild-type cells, ≈10 min in cells overexpressing the N-end rule pathway, and >10 h in cells lacking the pathway. The degradation of Gα is preceded by its multiubiquitylation (Madura & Varshavsky 1994). Like other Gα subunits of G proteins, the *S. cerevisiae* Gpa1p bears a conjugated N-terminal myristoyl moiety, which appears to be retained on Gpa1p during its targeting for degradation. A deletion of the first 88 residues

of Gpa1p greatly accelerates its decay but retains the dependence of Gpa1p degradation on Ubr1p (K. Madura, unpublished data). These data suggest that Ubr1p recognizes a feature of Gα that is distinct from the N-degron. Another, N-degron-based model invokes a *trans*-targeting mechanism (Fig. 2C, D).

Physiological implications of the Ubr1p-dependent degradation of Gα remain to be understood. Because the metabolic stability of Gα is expected to be influenced by its functional state—Gα can be GTP- or GDP-bound, covalently modified, or associated with Gβγ, the pheromone receptor, and other Gα ligands, the degradation of Gα in yeast may function either to augment or to inhibit cell's responses to a pheromone. A G_s-type Gα is short-lived in mouse cells as well (Levis & Bourne 1992), consistent with the possibility that Gα subunits of other organisms are also degraded by the N-end rule pathway. The activation of mouse Gα shortens its *in vivo* half-life (Levis & Bourne 1992), suggesting an adaptation-related function of Gα degradation. Further, Obin *et al*. (1994) described the ATP-dependent degradation of all three subunits of the bovine retinal G protein in reticulocyte extract. (It is not known whether Gβ and/or Gγ subunits of the *S. cerevisiae* G protein are also metabolically unstable.) Hondermarck *et al*. (1992) (see also Taban *et al*. 1996) reported that differentiation of rat pheochromocytoma PC12 cells is inhibited by dipeptides bearing destabilizing N-terminal residues. (These compounds have been shown to inhibit the N-end rule pathway in *S. cerevisiae* (Baker & Varshavsky 1991); their efficacy as N-end rule inhibitors in mammalian cells remains to be evaluated.) Given the findings with Gα (Madura & Varshavsky 1994), one interpretation of these results (Hondermarck *et al*. 1992) is that inhibitors of the N-end rule pathway may suppress cell differentiation through a metabolic stabilization of the relevant Gα subunits in PC12 cells.

Sindbis virus RNA Polymerase and other viral proteins

The Sindbis virus RNA polymerase, also called nsP4 (nonstructural Protein 4), is produced by an endoproteolytic cleavage of the viral precursor polyprotein nsP1234 (Strauss & Strauss 1994). The nsP4 protein bears N-terminal Tyr (a primary destabilizing residue; Figs 1 and 5A), and is degraded by the N-end rule pathway in reticulocyte extract (deGroot *et al*. 1991). Tyr is an N-terminal residue of other alphaviral RNA polymerases as well (Strauss & Strauss 1994), suggesting that these homologues of Sindbis polymerase are also degraded by the N-end rule pathway. Whereas the bulk of newly formed nsP4 is rapidly degraded, a fraction of nsP4 in infected cells is long-lived, presumably within a replication complex that contains viral and host proteins (Strauss & Strauss 1994, and references therein). This model may be generally applicable, in that physiological N-end rule substrates—including alphaviral RNA polymerases and Gα subunits of G proteins—are likely to exist in several states that differ by covalent modifications of a substrate and/or its associations with other ligands, and that consequently also differ by the rates at which various forms of a substrate are degraded by the N-end rule pathway.

There are many potential N-end rule substrates derived from viral polyproteins (Dougherty & Semler 1993). One of them is the integrase of the human immunodeficiency virus (HIV), produced by cleavages within the *gag-pol* precursor polyprotein. The processed integrase bears N-terminal Phe (Dougherty & Semler 1993), a strongly destabilizing residue in the N-end rule (Fig. 5A). Therefore it is possible that—similarly to the Sindbis virus RNA polymerase—at least a fraction of HIV integrase is short-lived *in vivo*.

c-Mos, a proto-oncoprotein

This 39 kDa Ser/Thr-kinase is expressed predominantly in male and female germ cells. Sagata and colleagues have identified c-Mos as a physiological substrate of the N-end rule pathway that is targeted for degradation through its N-terminal Pro residue (Nishizawa *et al*. 1992, 1993). Met-Pro-Ser-Pro, the encoded N-terminal sequence of *Xenopus* c-Mos, is conserved among all vertebrates examined (Nishizawa *et al*. 1992). Since the N-terminal Met-Pro peptide bond is readily cleaved by the major cytosolic Met-aminopeptidases (Arfin & Bradshaw 1988), the initially second-position Pro is expected to appear at the N-terminus of nascent c-Mos cotranslationally or nearly so.

The activity of the Pro-based N-degron in c-Mos is inhibited through the phosphorylation of Ser-2 (Ser-3 in the c-Mos ORF) (Nishizawa *et al*. 1992, 1993). During the maturation of *Xenopus* oocytes, c-Mos is phosphorylated partially and reversibly, and therefore remains short-lived. Later, at the time of germinal vesicle breakdown and the arrest of mature oocytes (eggs) at the second meiotic metaphase, c-Mos becomes long-lived, owing to its nearly stoichiometric phosphorylation at Ser-2 (Watanabe *et al*. 1991). Fertilization or mechanical activation of a *Xenopus* egg releases the meiotic arrest through the induced degradation of c-Mos—caused by a nearly complete dephosphorylation of phosphoserine-2 (Nishizawa *et al*. 1992, 1993).

Consistent with this model of the N-degron in c-Mos, the replacement of Ser-2 with Asp or Glu (whose negative charge mimics that of the phosphoryl group) rendered c-Mos long-lived, whereas the replacement of Ser-2 with Ala yielded a constitutively unstable c-Mos (Nishizawa *et al.* 1992). Lys-33 (Lys-34 in the c-Mos ORF) is a major ubiquitylation site of the c-Mos N-degron (Nishizawa *et al.* 1993).

In contrast to N-terminal Pro in the context of c-Mos, the N-terminal Pro followed by the sequence His-Gly Ser-. . . (this is the context of engineered N-end rule substrates such as X-βgal and X-DHFR (Varshavsky 1992)) did not confer a short half-life on a reporter protein in either yeast or mammalian cells (F. Lévy, T. Rümenapf & A.V., unpublished data). One interpretation of these results is that the N-degron of c-Mos, whose conserved N-terminal sequence is Pro-Ser-Pro-. . ., has a 'degron-enabling' internal determinant additional to, and perhaps specific for, the N-terminal Pro. The c-Mos N-degron is the first example of N-degron whose activity is regulated by phosphorylation (Nishizawa *et al.* 1992).

Compartmentalized proteins retrotransported to the cytosol

In contrast to cytosolic and nuclear proteins, the proteins that function in (or pass through) the ER, Golgi, and related compartments often bear destabilizing N-terminal residues—the consequence of cleavage specificity of signal peptidases, which remove signal sequences from proteins translocated into the ER (Bachmair *et al.* 1986). We have suggested that one function of the N-end rule pathway may be the degradation of previously compartmentalized proteins that 'leak' or are transported into the cytosol from compartments such as ER (Bachmair *et al.* 1986; Varshavsky 1992). Remarkably, it has recently been found that at least some compartmentalized proteins can be retrotransported to the cytosol through a route that requires specific ER proteins. US11, the ER-resident transmembrane protein encoded by cytomegalovirus, causes the newly translocated MHC class I heavy chain to be selectively retrotransported back to the cytosol, where the heavy chain is degraded by a proteasome-dependent pathway (Wiertz *et al.* 1996). Similarly, CPY*, a defective vacuolar carboxypeptidase of *S. cerevisiae*, is retrotransported to the cytosol shortly after entering the ER, and is degraded in the cytosol by a Ub/proteasome-dependent pathway that requires the Ubc7p Ub-conjugating enzyme (Hiller *et al.* 1996). The expected N-terminal residue of the translocated and processed MHC class I heavy chain is Gly—a stabilizing residue (Fig. 1). The expected N-terminal residue of the wild-type CPY carboxypeptidase whose signal sequence had been cleaved off is Ile—a primary destabilizing residue (Fig. 1). Whether the N-end rule pathway plays a role in the degradation of retrotransported proteins remains to be determined.

Potential N-end rule substrates

Several cytosolic and nuclear proteins are known to bear destabilizing N-terminal residues, but have not been shown, thus far, to be degraded by the N-end rule pathway. Among them are the λ phage cII protein, the *S. cerevisiae* Cup1p protein, the catalytic subunits of calpains (calcium-dependent proteases), and several histone-like, micronucleus-specific proteins of *Tetrahymena*. These putative N-end rule substrates are discussed by Varshavsky *et al.* (1997).

Applications of N-degron

The portability and modular organization of N-degrons make possible a variety of applications whose common feature is the conferring of a constitutive or conditional metabolic instability on a protein of interest. These applications are discussed elsewhere (Varshavsky 1995, 1996a,b).

Concluding remarks

Although many things have been learnt about the N-end rule since its discovery 10 years ago, the answers to several key questions remain unknown. For example, the detailed mechanics of targeting is not understood. The biochemical dissection of the N-end rule pathway reconstituted *in vitro* from defined (cloned) components will be essential for attaining this goal. Crystallographic-quality structural information about N-recognin and the entire targeting complex will be required as well. The recently emerged possibility that N-recognin may target not only N-degrons but also other degradation signals adds yet another level of complexity which will have to be addressed.

Genetic screens for proteins degraded by the N-end rule pathway are our best hope for bringing to light physiological N-end rule substrates. It is already clear that at least some of these substrates are conditionally unstable—for example, partitioned between a short-lived free substrate and a long-lived complex of the substrate with other proteins. In addition, for some substrates, the rate-limiting step in their degradation

may be a processing (cleavage) event that produces an N-degron from a pre-N-degron. If so, a significant fraction of extant substrate molecules may bear a stabilizing N-terminal residue. Given these obstacles to identifying physiological N-end rule substrates, they are likely to be more numerous than is apparent at the present time.

Acknowledgements

I am grateful to the current and former members of the laboratory, whose work on the N-end rule is described in this review. I thank colleagues whose names are cited in the text for their permission to discuss unpublished data. Our studies are supported by grants from the NIH and the Association for the Cure of the Cancer of the Prostate.

This article is a modified and updated version of an earlier review (Varshavsky 1996b).

Abbreviations

Ub, ubiquitin; βgal, β-galactosidase; DHFR, dihydrofolate reductase; Nt^N-amidase, amidohydrolase specific for N-terminal Asn; R-transferase, Arg-tRNA-protein transferase; L/F-transferase, Leu, Phe-tRNA-protein transferase.

References

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

Arfin, S.M. & Bradshaw, S.A. (1988) Cotranslational processing and protein turnover in eukaryotic cells. *Biochemistry* **27**, 7979–7984.

Arnason, T.A. & Ellison, M.J. (1994) Stress resistance in *S. cerevisiae* is stronglycorrelated with assembly of a novel type of multiubiquitin chain. *Mol. Cell. Biol.* **14**, 7876–7883.

Baboshina, O.V. & Haas, A.L. (1996) Novel multiubiquitin chain linkages catalyzed by the conjugating enzymes E2EPF and RAD6 are recognized by 26S proteasome subunit 5. *J. Biol. Chem.* **271**, 2822–2831.

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

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

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

Baker, R.T. & Varshavsky, A. (1991) Inhibition of the N-end rule pathway in living cells. *Proc. Natl. Acad. Sci. USA* **88**, 1090–1094.

Balzi, E., Choder, M., Chen, W., Varshavsky, A. & Goffeau, A. (1990) Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. *J. Biol. Chem.* **265**, 7464–7471.

Bartel, B., Wünning, I. & Varshavsky, A. (1990) The recognition component of the N-end rule pathway. *EMBO J.* **9**, 3179–3189.

Chau, V., Tobias, J.W., Bachmair, A. *et al.* (1989) A multiubiquitin chain is confined to a specific lysine in a targeted short–lived protein. *Science* **243**, 1576–1583.

Ciechanover, A., Ferber, S., Ganoth, D., Elias, S., Hershko, A. & Arfin, S. (1988) Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. *J. Biol. Chem.* **263**, 11155–11167.

Dayal, V.K., Chakraborty, G., Sturman, J.A. & Ingoglia, N.A. (1990) The site of amino acid addition to posttranslationally modified proteins of regenerating rat sciatic nerves. *Biochim. Biophys. Acta* **1038**, 172–177.

deGroot, R.J., Rümenapf, T., Kuhn, R.J., Strauss, E.G. & 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. & Varshavsky, A. (1994) Heat-inducible degron: a method for constructing temperature-sensitive mutants. *Science* **263**, 1273–1276.

Dohmen, R.J., Madura, K., Bartel, B. & Varshavsky, A. (1991) The N-end rule is mediated by the Ubc2 (Rad6) ubiquitin-conjugating enzyme. *Proc. Natl. Acad. Sci. USA* **88**, 7351–7355.

Dougherty, W.G. & Semler, B.L. (1993) Expression of virus-encoded proteinases: functional and structural similarities with cellular enzymes. *Microbiol. Rev.* **57**, 781–822.

Finley, D., Bartel, B. & Varshavsky, A. (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature* **338**, 394–401.

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

Gonda, D.K., Bachmair, A., Wünning, I., Tobias, J.W., Lane, W.S. & Varshavsky, A., (1989) Universality and structure of the N-end rule. *J. Biol. Chem.* **264**, 16700–16712.

Gottesman, S. & Maurizi, M.R. (1992) Regulation by proteolysis: energy-dependent proteases and their targets. *Microbiol. Rev.* **56**, 592–621.

Gottesman, S., Clark, W.P., de Crecy-Lagard, V. & Maurizi, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. *J. Biol. Chem.* **268**, 22618–22626.

Grigoryev, S., Stewart, A.E., Kwon, Y.T., Arfin, S.M., Bradshaw, R.A., Copeland, N.J. & 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. (1991) The ubiquitin pathway for protein degradation. *Trends Biochem. Sci.* **16**, 265–268.

Hill, C.P., Johnston, N.L. & 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.

Hiller, M.M., Finger, A., Scheiger, M. & Wolf, D.H. (1996) Endoplasmic reticulum associated degradation of a mutated soluble vacuolar enzyme, carboxypeptidase Y, occurs via the ubiquitin-proteasome pathway. *Science* **273**, 1725–1728.

Hondermarck, H., Sy, J., Bradshaw, R.A. & Arfin, S.M. (1992) Dipeptide inhibitors of ubiquitin-mediated protein turnover prevent growth factor-induced neurite outgrowth in rat

pheochromocytoma PC12 cells. *Biochem. Biophys. Res. Commun.* **189**, 280–288.
Jentsch, S. (1992) The ubiquitin-conjugation system. *Annu. Rev. Genet.* **26**, 179–207.
Johnson, E.S., Ma, P.C.M., Ota, I.M. & Varshavsky, A. (1995) A proteolytic pathway that recognizes ubiquitin as a degradation signal. *J. Biol. Chem.* **270**, 17442–17456.
Johnson, E.S., Gonda, D.K. & Varshavsky, A. (1990) *Cis-trans* recognition and subunit-specific degradation of short-lived proteins. *Nature* **346**, 287–291.
Johnston, J.A., Johnson, E.S., Waller, P.R.H. & Varshavsky, A. (1995) Methotrexate inhibits proteolysis of dihydrofolate reductase by the N-end rule pathway, *J. Biol. Chem.* **270**, 8172–8178.
Kaiser, C.A., Preuss, D., Grisafi, P. & Botstein, D. (1987) Many random sequences functionally replace the secretion signal sequence of yeast invertase. *Science* **235**, 312–317.
Knight, S.A.B., Tamai, K.T., Kosman, D.J. & Thiele, D.J. (1994) Identification and analysis of a *Saccharomyces cerevisiae* copper homeostasis gene encoding a homeodomain protein. *Mol. Cell. Biol.* **14**, 7792–7804.
Knighton, D.R., Kan, C.C., Howland, E. *et al.* (1994) Structure of and kinetic channeling in bifunctional dihydrofolate reductase-thymidylate synthase, *Nature Struct. Biol.* **1**, 186–194.
Levis, M.J. & Bourne, H.R. (1992) Activation of the α subunit of G_s in intact cells alters its abundance, rate of degradation, and membrane avidity. *J. Cell. Biol.* **119**, 1297–1307.
Lévy, F., Johnsson, N., Rümenapf, T. & Varshavsky, A. (1996) Using ubiquitin to follow the metabolic fate of a protein. *Proc. Natl. Acad. Sci. USA* **93**, 4907–4912.
Li, X. & Chang, Y.H. (1995) Amino-terminal protein processing in *Saccharomyces cerevisiae* is an essential function that requires two distinct methionine aminopeptidases. *Proc. Natl. Acad. Sci. USA* **92**, 12357–12361.
Madura, K. & Varshavsky, A. (1994) Degradation of Gα by the N-end rule pathway, *Science* **265**, 1454–1458.
Madura, K., Dohmen, R.J. & Varshavsky, A. (1993) N-recognin/Ubc2 interactions in the N-end rule pathway. *J. Biol. Chem.* **268**, 12046–12054.
Maeda, T., Wurgler-Murphy, S.M. & Sato, H. (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**, 242–245.
Murray, A. & Hunt, T. (1993) *The Cell Cycle.* New York: W.H. Freeman & Co., pp. 60–62.
Negrutskii, B.S. & Deutscher, M.P. (1991) Channeling of aminoacyl-tRNA for protein synthesis *in vivo*, *Proc. Natl. Acad. Sci. USA* **88**, 4991–4995.
Nishizawa, M., Furuno, N., Okazaki, K., Tanaka, H., Ogawa, Y. & Sagata, N. (1993) Degradation of Mos by the N-terminal proline-dependent ubiquitin pathway on fertilization of *Xenopus* eggs: possible significance of natural selection for Pro-2 in Mos. *EMBO J.* **12**, 4021–4027.
Nishizawa, M., Okazaki, K., Furuno, N., Watanabe, N. & Sagata, N. (1992) The 'second-codon rule' and autophosphorylation govern the stability and activity of Mos during the meiotic cell cycle in *Xenopus* oocytes. *EMBO J.* **11**, 2433–2446.
Obin, M., Nowell, T. & Taylor, A. (1994) The photoreceptor G-protein transducin is a substrate for ubiquitin-dependent proteolysis. *Biochem. Biophys. Res. Communication* **200**, 1169–1176.
Ota, I.M. & Varshavsky, A. (1992) A gene encoding a putative tyrosine phosphatase suppresses lethality of an N-end rule-dependent mutant. *Proc. Natl. Acad. Sci. USA* **89**, 2355–2359.
Ota, I.M. & Varshavsky, A. (1993) A yeast protein similar to bacterial two-component regulators. *Science* **262**, 566–569.
Özkaynak, E., Finley, D., Solomon, M.J. & Varshavsky, A. (1987) The yeast ubiquitin genes: a family of natural gene fusions. *EMBO J.* **6**, 1429–1440.
Pahl, H.L. & Baeuerle, P.A. (1996) Control of gene expression by proteolysis. *Curr. Opin. Cell Biol.* **8**, 340–347.
Pickart, C. (1988) Ubiquitin carrier proteins. In: *Ubiquitin* (ed. M. Rechsteiner), pp. 77–99. Plenum Press, NY.
Rechsteiner, M., Hoffman, L. & Dubiel, W. (1993) The multicatalytic and 26S proteases. *J. Biol. Chem.* **268**, 6065–6068.
Reiss, Y., Kaim, D. & Hershko, A. (1988) Specificity of binding of N-terminal residue of proteins to ubiquitin-protein ligase. Use of amino acid derivatives to characterize specific binding sites. *J. Biol. Chem.* **263**, 2693–2698.
Sadis, S., Atienza, C. & Finley, D. (1995) Synthetic signals for ubiquitin-dependent proteolysis. *Mol. Cell. Biol.* **15**, 4086–4095.
Schatz, G. & Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* **271**, 1519–1526.
Scheffner, M., Nuber, U. & Huibregtse, J.M. (1995) Protein ubiquitination involving an E1-E2-E3 enzyme ubiquitin thioester cascade. *Nature* **373**, 81–83.
Sherman, F., Stewart, J.W. & Tsunasawa, S. (1985), Methionine or not methionine at the beginning of a protein? *BioEssays* **3**, 27–31.
Shrader, T.E., Tobias, J.W. & Varshavsky, A. (1993) The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. *J. Bact.* **175**, 4364–4374.
Simon, S.M. & Blobel, G. (1991) A protein-conducting channel in the endoplasmic reticulum. *Cell* **69**, 371–380.
Spence, J., Sadis, S., Haas, A.L. & Finley, D. (1995) A ubiquitin mutant with specific defects in DNA repair and multiubiquitination. *Mol. Cell Biol.* **15**, 1265–1273.
Stewart, A.E., Arfin, S.M. & Bradshaw, R.A. (1995) The sequence of porcine protein N-terminal asparagine amidohydrolase: a new component of the N-end rule pathway. *J. Biol. Chem.* **270**, 25–28.
Strauss, J.H. & Strauss, E.G. (1994) The alphaviruses: gene expression, replication, and evolution. *Microbiol. Rev.* **58**, 491–562.
Taban, C.H., Hondermarck, H., Bradshaw, R.A. & Biolly, B. (1996) Effect of a dipeptide inhibiting ubiquitin-mediated protein degradation nerve-dependent limb regeneration in the newt. *Experientia* **52**, 865–70.
Tobias, J.W., Shrader, T.E., Rocap, G. & Varshavsky, A. (1991) The N-end rule in bacteria. *Science* **254**, 1374–1377.
Varshavsky, A., Byrd, C., Davydov, I.V. *et al.* (1997) The N-end rule pathway. In: *Ubiquitin and the Biology of the Cell*, (eds D. Finley & J.-M. Peters) Plenum Press, NY. (in press).
Varshavsky, A. (1996a) The N-end rule. *Cold Spring Harbor Symp. Quant. Biol.* **60**, 461–478.
Varshavsky, A. (1996b) The N-end rule: functions, mysteries, uses. *Proc. Natl. Acad. Sci. USA* **93**, 12142–12149.
Varshavsky, A. (1995) Codominance and toxins: a path to drugs

of nearly unlimited selectivity. *Proc. Natl. Acad. Sci. USA* **92**, 3663–3667.
Varshavsky, A. (1992) The N-end rule. *Cell* **69**, 725–735.
Varshavsky, A. (1991) Naming a targeting signal. *Cell* **64**, 13–15.
Watanabe, N., Hunt, T., Ikawa, Y. & Sagata, N. (1991) Independent inactivation of MPF and cytostatic factor (Mos) upon fertilization of *Xenopus* eggs. *Nature* **352**, 247–248.
Watkins, J.F., Sung, P., Prakash, L. & Prakash, S. (1993) The *Saccharomyces cerevisiae* DNA repair gene *RAD23* encodes a nuclear protein containing ubiquitin-like domain required for biological function. *Mol. Cell. Biol.* **13**, 7757–7765.
Wawrzynov, A., Wojtkowiak, D., Marzalec, J., *et al.* (1995) The ClpX heat-shock protein of *Escherichia coli*, the ATP-dependent specificity component of the ClpP-ClpX protease, is a novel molecular chaperone. *EMBO J.* **14**, 1867–1877.
Wickner, S., Gottesman, S., Skowyra, D., Hoskins, J., McKenney, K. & Maurizi, M.R. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl. Acad. Sci.* USA **91**, 12218–12222.
Wiertz, E.J.H.J., Jones, T.R., Sun, L., Bogyo, M., Geuze, H.J. & Ploegh, H.L. (1996) The human cytomegalovirus US11 gene product dislocates MHC class I heavy chains from the endoplasmic reticulum to the cytosol. *Cell* **84**, 769–779.
Wolf, S., Lottspeich, F. & Baumeister, W. (1993) Ubiquitin found in the archaebacterium *Thermoplasma acidophilum*. *FEBS Lett.* **326**, 42–44.