NIH Public Access

Author Manuscript

Annu Rev Biochem. Author manuscript; available in PMC 2013 March 28.

Published in final edited form as:
Annu Rev Biochem. 2012 ; 81: 261–289. doi:10.1146/annurev-biochem-051710-093308.

The N-End Rule Pathway

Takafumi Tasaki[1], **Shashikanth M. Sriram**[1], **Kyong Soo Park**[2,3], and **Yong Tae Kwon**[1,2]

[1]Center for Pharmacogenetics and Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, Pennsylvania 15261;

[2]World Class University Program, Department of Molecular Medicine and Biopharmaceutical Sciences, Graduate School of Convergence Science and Technology and College of Medicine, Seoul National University, Seoul 110-799, Korea

[3]Department of Internal Medicine, College of Medicine, Seoul National University, Seoul 110-799, Korea

Abstract

The N-end rule pathway is a proteolytic system in which N-terminal residues of short-lived proteins are recognized by recognition components (N-recognins) as essential components of degrons, called N-degrons. Known N-recognins in eukaryotes mediate protein ubiquitylation and selective proteolysis by the 26S proteasome. Substrates of N-recognins can be generated when normally embedded destabilizing residues are exposed at the N terminus by proteolytic cleavage. N-degrons can also be generated through modifications of posttranslationally exposed pro-N-degrons of otherwise stable proteins; such modifications include oxidation, arginylation, leucylation, phenylalanylation, and acetylation. Although there are variations in components, degrons, and hierarchical structures, the proteolytic systems based on generation and recognition of N-degrons have been observed in all eukaryotes and prokaryotes examined thus far. The N-end rule pathway regulates homeostasis of various physiological processes, in part, through interaction with small molecules. Here, we review the biochemical mechanisms, structures, physiological functions, and small-molecule-mediated regulation of the N-end rule pathway.

Keywords

N-degron; arginylation; ubiquitin; proteolysis

INTRODUCTION

The selectivity in regulated proteolysis is governed by timely generation and recognition of specific degrons on substrates. Degrons on short-lived proteins in eukaryotes are recognized by ubiquitin (Ub) ligases, which mediate the conjugation of Ub to an internal Lys of the substrate, resulting in ATP-dependent degradation by the 26S proteasome. In bacteria, which lack Ub, specific adaptor proteins recognize and deliver protein substrates to proteolytic machinery. The first degron identified in short-lived proteins is a single N-terminal residue, which is targeted by the N-end rule pathway (1). Substrates of N-recognins are generated through N-terminal Met excision or endoproteolytic cleavages of proteins

Copyright © 2012 by Annual Reviews. All rights reserved

yok5@pitt.edu.

DISCLOSURE STATEMENT The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

associated with covalent modifications of posttranslationally exposed N-terminal residues. The N-end rule pathway has long been thought to target a limited number of regulatory proteins of the Ub-proteasome system (UPS). However, recent studies suggest that the majority of cellular proteins may carry N-terminal degradation determinants, at least transiently, through protein-specific and global posttranslational modifications under certain physiological states.

The N-end rule pathway was discovered in 1986 when Varshavsky and colleagues (1) found that engineered substrates carrying certain N-terminal residues, which were generated from Ub fusion proteins following cleavage by deubiquitylating enzymes, were rapidly degraded in *Saccharomyces cerevisiae* cells (Figure 1). A series of genetic analyses in *S. cerevisiae* identified the N-recognin Ubr1 and proteins involved in the generation of N-degrons. Ubr1, a 200-kDa-RING E3 ligase, binds a primary destabilizing residue and mediates protein ubiquitylation and subsequent degradation by the proteasome (2, 3). Substrates of Ubr1 include positively charged (Arg, Lys, and His; type 1) and bulky hydrophobic (Phe, Trp, Tyr, Leu, and Ile; type 2) primary destabilizing residues (Figure 1). A destabilizing residue is part of N-degrons, and successful degradation through Ubr1 requires additional sequence features, such as an internal Lys residue (the site of a polyubiquitylation) and an unstructured N-terminal extension (4). In the yeast N-end rule pathway, Arg is the principal degron and can be generated through post-translational modifications, such as arginylation and deamidation, of pro-N-degrons (Asn, Gln, Asp, and Glu). A recent study identified an alternative N-end rule pathway in *S. cerevisiae* in which acetylated N-terminal residues, which occur in the majority of cellular proteins, act as N-degrons (5, reviewed in References 6–8). Hereafter, we refer to the arginylation-based N-end rule pathway as the classical N-end rule pathway or, simply, the N-end rule pathway.

The N-end rule pathway has been identified in all species examined, ranging from mammals (9) and plants (10, 11) to yeasts (1) and bacteria (12). Several excellent reviews have recently discussed the hierarchical structures and basic mechanisms of the N-end rule pathway in eukaryotes and prokaryotes (6–8, 13, 14). Here, we review the biochemical details, substrates, and physiological functions of posttranslational modifications involved in generation of N-degrons and discuss the functions and mechanisms of various eukaryotic and bacterial N-recognins, with an emphasis on the emerging UBR box N-recognin family in mammalian development. The topics also include how components of the N-end rule pathway sense and react to small molecules, such as oxygen, nitric oxide, heme, and peptides with destabilizing residues, by controlling cellular concentrations of their substrates. Finally, we discuss how N-degrons can be engineered to induce proteolysis of other proteins and to serve as an affinity ligand.

GENERATION OF N-DEGRONS BY CONJUGATION OF AMINO ACIDS

In the classical N-end rule pathway of eukaryotes and bacteria, conjugation of destabilizing amino acids to pro-N-degrons is the major way to generate primary destabilizing residues (Figure 1). This modification, mediated by evolutionarily conserved aminoacyl-tRNA transferases (the enzyme EC2.3.2), enables pro-N-degrons to be conditionally recognized by N-recognins (15–19). Most substrates of the classical N-end rule pathway identified thus far carry the N-terminal residues derived from the aminoacyl moiety of aminoacyl-tRNAs or pro-N-degrons, whose activity requires aminoacyl-tRNA transferases. Interestingly, recent studies suggest that mammalian and bacterial N-recognins have been structurally optimized to the degrons derived from aminoacyl-tRNAs (Arg in eukaryotes and Leu and Phe in prokaryotes), highlighting the importance of amino acid conjugation at a licensing step prior to irreversible proteolysis (6, 7, 20, 21). Below, we discuss the catalytic mechanisms,

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

enzymatic specificities, evolution, and functions of aminoacyl-tRNA transferases that generate N-degrons.

Arginylation in the Eukaryotic N-End Rule Pathway

In eukaryotes, the N-terminal Arg is the structurally preferred degron for the UBR box of N-recognins (6, 20, 21). The degron Arg can be generated by *ATE1*-encoded arginyl (R)-transferases, which transfer Arg from Arg-tRNA to the N-terminal α-amino group of acceptor substrates (Figure 2) (15, 18, 22–25). In *S. cerevisiae*, a single R-transferase, encoded by *Ate1* with no known functions, conjugates Arg to the secondary residues Asp and Glu (Figure 1*b*) (16). By contrast, the mammalian *ATE1* gene expresses at least six isoforms through alternative splicing of pre-mRNA, including those with either of two homologous exons (18, 26, 27). The physiological importance of protein arginylation has been established by the discovery that ATE1-deficient mouse embryos die owing to defects in cardiac and vascular development (24). Although ATE1 isoforms remain poorly characterized in donor and acceptor specificities, tissue distribution, and physiological functions (18, 26, 27), biochemical analyses indicate that Cys as well as Asp and Glu are substrates of arginylation in mammals (reviewed in Reference 4).

Substrates of arginylation include structurally related mammalian RGS (regulator of G-protein signaling) proteins (28–30). These regulators of G protein–signaling function as GTPase-activating proteins for heterotrimeric G protein α-subunits of the i and q classes (Supplemental Figure 1; follow the Supplemental Material link from the Annual Reviews home page at http://www.annualreviews.org). The degradation of these substrates involves N-terminal Met excision by Met aminopeptidases, which exposes the pro-N-degron Cys2 at the N terminus. The N-terminal Cys2 is subsequently arginylated by ATE1 to generate the degron Arg of N-recognins. Although the exact chemical nature of Cys2 as an arginylation substrate remains murky, various biochemical analyses suggest that the N-terminal Cys2 is not a direct substrate of arginylation but is converted to an arginylation-permissive acceptor following oxidation into Cys sulfinic acid [$CysO_2$(H)] or Cys sulfonic acid [$CysO_3$(H)], a structural mimic of the arginylation substrate Asp (Supplemental Figure 1). Interestingly, Cys-dependent degradation of RGS5 is inhibited when molecular oxygen or nitric oxide is deleted (29, 24, 30). On the basis of hypoxia-sensitive proteolysis of RGS proteins and cardiovascular defects of ATE1-deficient mice, it was proposed that oxidation of N-terminal Cys represents an oxygen sensor in cardiovascular development and signaling (4, 24, 30). This conjecture was further supported by recent discoveries that the N-end rule pathway of the plant *Arabidopsis* functions as an oxygen sensor through regulated proteolysis of the hypoxia-sensitive transcription factor family carrying the pro-N-degron Cys2 (31, 32). In normoxia, the ethylene response factor group VII transcription factors, including hypoxia-responsive element 1 and 2 (HRE1 and HRE2) and related to AP2.12 (RAP2.12), are downregulated through proteasomal degradation in a manner depending on the pro-N-degron Cys2 (31, 32). In hypoxia, however, these hypoxia-sensitive transcription factors are accumulated, resulting in transcriptional induction of genes that promote anaerobic metabolism and survival of hypoxia. As hypoxia-inducible factor-1 (HIF-1), a known oxygen sensor in animals, is absent in plants, the Cys branch of the N-end rule pathway may represent an oxygen-sensing mechanism in plants. The *Arabidopsis* and human genomes encode at least 206 and 502 proteins, respectively, with the Met-Cys motif (4, 31, 32). Thus, these Met-Cys proteins may represent a unique proteome, whose functions include sensing oxygen and other cellular stresses through oxidation and arginylation of the pro-N-degron Cys.

Polyubiquitination of an ideal N-end rule substrate requires a Lys residue as a site of polyubiquitination and an unstructured N-terminal extension (4). Thus, N-terminal arginylation does not necessarily lead to proteasomal degradation but can serve as a non-

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

proteolytic modification controlling cellular processes. Calreticulin is an endoplasmic reticulum (ER) resident chaperone whose the signal peptide is removed upon translocation, exposing the pro-N-degron Asp18 at the N terminus of the mature protein (33). A portion of the mature protein in the ER lumen is arginylated upon retrotranslocation into the cytosol and regulates intracellular relocalization of calreticulin to the stress granule upon cellular stresses, apparently without involving acute proteolysis (33). The protein β-actin, one of most abundant cellular proteins, also undergoes non-proteolytic arginylation at the N-terminal Asp2 or Asp3 following N-terminal Met excision, which plays a role in actin filament properties, actin polymerization, and lamella formation in motile cells (reviewed in Reference 34). In addition to the N terminus, posttranslational arginylation has been observed in the side chain of various residues of many cellular proteins (35). Recent genetic studies revealed various physiological functions of ATE1 in cardiovascular development (24), spermatogenesis (36), metabolism (36), and neural crest migration (37) in animals; seed ripening and germination, shoot and leaf development, and leaf senescence in the plant *Arabidopsis* (10, 38; reviewed in Reference 14); and in apoptosis and viability in the fly *Drosophila* (39). It remains to be determined which of these functions are directly relevant to arginylation-induced proteolysis.

Leucylation and Phenylalanylation in the Prokaryotic N-End Rule Pathway

N-terminal Leu and Phe residues on bacterial proteins are the primary destabilizing residues that can be generated by conjugation of destabilizing amino acids derived from aminoacyl-tRNAs (Figure 1*c*) (16, 19). Two classes of aminoacyl transferases are known to mediate leucylation and phenylalanylation in the N-end rule pathway: leucyl/pheylalanyl-tRNA-protein (L/F)-transferases and leucyl-tRNA-protein (L)-transferases (Figure 1*c*). The *Escherichia coli* L/F-transferase, encoded by *aat*, transfers Leu or Phe to the acceptors Arg and Lys, which are type 1 primary residues in eukaryotes (Figure 2) (19). This acceptor specificity was originally determined in the 1971 study by Leibowitz and Soffer (40) and was later confirmed by crystal structures of the L/F-transferase (41). However, recent identification of the first substrate of the bacterial N-end rule pathway, PATase (putrescine aminotransferase), lead to an unexpected discovery that the L/F-transferase conjugates Leu or Phe to the N-terminal Met (not Arg or Lys) of PATase (42). Because Met-β-galactosidase is not a substrate of the L/F-transferase and is stable in *E. coli* (12, 42), the substrate specificity of the L/F-transferase needs to be further investigated. The second aminoacyl transferase of the bacterial N-end rule pathway is Bpt L-transferase. In contrast to the Aat L/F-transferase, the Bpt L-transferase transfers Leu to N-terminal Asp and Glu, which are arginylation acceptors in eukaryotes (Figure 2) (16). Consistent with R-transferase-like acceptor specificity, the Bpt L-transferase is sequelogous [similar in sequence (43)] to eukaryotic R-transferases but possesses the donor tRNA specificity of prokaryotic Aat L/F-transferase. In the *Vibrio vulnificus* genome, the Aat and Bpt transferases are expressed from a single operon in which *bpt* begins within the stop codon of *aat* (16). The Aat-Bpt tandem gene structure is conserved in a significant portion of β-proteobacteria and γ-proteiobacteria but not in *E. coli* and other enterobacteria, a newest group of bacteria where *bpt* has been lost (T. Tasaki and Y.T. Kwon, unpublished data).

The protozoan parasite *Plasmodium falciparum*, which causes malaria in humans, expresses yet another type of transferases, termed ATEL1 (16). ATEL1 is sequelogous to the prokaryotic Aat L/F-transferase but has the enzymatic specificity of eukaryotic ATE1 R-transferases, which conjugates Arg to N-terminal Asp and Glu (Figure 2) (16). Thus, eukaryotic R-transferases and prokaryotic Aat L/F-transferase may have a common ancestor but are evolutionally divergent. Indeed, the C-terminal halves of mammalian R-transferases and bacterial L/F-transferases share a structurally conserved domain, which mediates the synthesis of the interchain peptide of the peptidoglycan and is unique to aminoacyl-tRNA

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

transferases of the FemABX family (Figure 2) (44, 45). The conserved GCN5-related N-acetyltransferase folds, which catalyze peptidyltransferase reactions using aminoacyl-tRNA, are confined to the R-transferase, L/F-transferase, and FemABX families (45). The catalytic mechanism of peptide-bond formation by the L/F-transferase was determined using crystal structures of the *E. coli* L/F-transferase in complex with a donor substrate, phenylalanyl adenosine, and an acceptor substrate, the α-casein peptide RYLGYL bearing N-terminal Arg (Figure 3*a*,*b*) (41). The L/F-transferase mediates phenylalanyl transfer from phenylalanyl adenosine to the peptide RYLGYL, yielding the product peptide FRYLGYL. The N-terminal Phe of the product peptide occupies a hydrophobic pocket with a confined C-shaped edge. The size and shape of the hydrophobic pocket suits hydrophobic residues lacking the branched β-carbon, such as Leu and Phe, but excludes hydrophilic or charged residues (45). It has been proposed that the access of the acceptor protein bearing an N-terminal Arg, associated with conformational changes in the L/F-transferase, induces the hydrogen bond breakage between Gln188 and Glu156 and the electron relays between Asp186 and Gln188. This electron relay facilitates the nucleophilic attack of the α-amino group of Arg on the neighboring carbonyl carbon of the esterified aminoacyl-tRNAs, leading to the formation of a peptide bond between Arg and Leu/Phe of the corresponding aminoacyl-tRNA (41).

GENERATION OF N-DEGRONS VIA DEAMIDATION

The tertiary destabilizing residues Asn and Gln on eukaryotic proteins can induce proteasomal degradation of otherwise stable proteins through deamidation into the secondary destabilizing residues Asp and Glu (Figure 1*a*,*b*) (4, 46, 47). In *S. cerevisiae*, these pro-N-degrons are deamidated by the N-terminal amidohydrolase Nta1 ($Nt^{N,Q}$-amidase) with no known functions, a member of the Nitrilase superfamily (46, 48). In mammals and other multicellular eukaryotes, N-terminal Asn and Gln are respectively deamidated by *NTAN1*-encoded Nt^{N}-amidase and *NTAQ1*-encoded Nt^{Q}-amidase, which are not sequelogous to each other or their yeast counterparts (47–50). Protein sequences and enzymatic properties indicate that these amidases have different evolutionary origins and were independently recruited to the UPS via arginylation (47–50). Despite low sequelogy, they share the principles of substrate recognition and catalysis involving a conserved Cys residue critical for the deamidation activity (46, 49). The physiological function of deamidation was first identified when NTAN1-deficient mice were found to exhibit altered learning, memory, and social behavior (50–52) which were, in part, attributed to misregulation in magnetism-induced proteasomal degradation of the microtubule-associated protein 2 in hippocampal neurons (53, 54).

The cocrystal of human NTAQ1 with a bound peptide (PDB code 3C9Q) illustrated a monomeric globular protein with a three-layer sandwich architecture (Figure 3*c*,*d*). Its active-site region is structurally similar to that of the human transglutaminase factor XIII, despite lack of sequence similarity (48). Transglutaminase catalyzes formation of an isopeptide bond between the free ε-amino group of a Lys residue and the γ-carboxamide group of a Gln residue (55). The catalytic triad (Cys314, His373, and Asp396) of factor XIII transglutaminase corresponds to the evolutionarily conserved Cys28, His81, and Asp97 of NTAQ1, which are located near the bound peptide (48). It was therefore speculated that NTAQ1 may mediate deamidation of Gln through a transglutamination-like reaction in which a water molecule, instead of the ε-amino group of a Lys, attacks the N-terminal Gln (48).

Substrates of deamidation include the *Drosophila* apoptosis inhibitor, DIAP1. This protein has a short half-life of ~30 min, in part, through degradation by its RING domain-mediated autoubiquitylation (56, 57). In addition to autoubiquitylation, the N-end rule pathway

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

provides an additional layer of regulation in the turnover of DIAP1 (Supplemental Figure 2*a*) (57). In *Drosophila* cells, DIAP1 can be cleaved after Asp20 by effector caspases (e.g., DrICE and DCP-1), exposing the pre-N-degron Asn21 on Asn21-DIAP1 (57). The N-terminal Asn21 is subsequently deamidated into Asp by Ntan1, which, in turn, is conjugated with Arg by the *Ate1*-encoded R-transferase, resulting in ubiquitylation and proteasomal degradation through N-recognins (57, 58). This pro-N-degron is well conserved in different insect species over an evolutionary distance of ~300 million years (59). Consistent with N-end rule degradation of DIAP1, inactivation of a single *Ntan1* or *Ate1* allele was found to inhibit head involution defective (HID)-induced apoptosis in the developing eyes of transgenic flies overexpressing the DIAP1 antagonist HID (57). Intriguingly, an analogous analysis resulted in an opposite effect, i.e., accelerated apoptosis when apoptosis was induced by another DIAP1 antagonist, Reaper (57), suggesting that the N-end rule pathway is part of a complicated network maintaining homeostasis in apoptosis.

The pro-N-degron Asn21 is also required for optimal activity of DIAP1 as an inhibitor of apoptosis (56, 60). For example, whereas overexpression of full-length DIAP1 or Asn21-DIAP1 efficiently inhibits Reaper-induced apoptosis, metabolically more stable Met21-DIAP1 does not show such activity (56, 57). According to the N-end rule, the nonproteolytic activity of Asn21 requires its deamidation into Asp and subsequent arginylation to generate the degron Arg. Ditzel et al. (56) proposed a model in which an N-recognin is recruited by the degron Arg of arginylated DIAP1 to form a heterodimeric E3 complex targeting an effector caspase that also has been recruited to DIAP1 through the BIR domain (Supplemental Figure 2*b*).

GENERATION OF N-DEGRONS BY ACETYLATION

In *S. cerevisiae*, acetylation of N-terminal residues can generate N-degrons, which are targeted by the N-recognin Doa10 in conjunction with the Ubc6 or Ubc7 E2 enzyme (5). N-terminal acetylation is one of the most common protein modifications in eukaryotes, occurring on ~57% of yeast proteins and ~84% human proteins but rarely on prokaryotic proteins (61). This global modification is catalyzed by ribosome-associated N-terminal acetyltransferases (NATs), which transfer acetyl groups from acetyl-coenzyme A to the N-terminal α-amino group. Acetylation starts when the nascent chain with the initiator Met emerges from the ribosome at a length of ~25 residues or somewhat later (~50 residues) if the initiator Met has to be removed by methionine aminopeptidase. Thus, one important factor in N-terminal acetylation is the primary structure of the N-terminal region if NAT and acetyl coenzyme A are not rate limiting (62).

Three major NAT complexes (NatA, NatB, and NatC) conserved from yeast to humans are thought to be responsible for the majority of acetylation, whereas other NATs (NatD, NatE, and NatF) remain poorly characterized (63). NatA acetylates N-terminal Ala, Ser, Thr, Val, Gly, and Cys following N-terminal Met excision by methionine aminopeptidases. NatB and NatC acetylate the N-terminal Met when the second residue is either acidic or hydrophobic, respectively. Eukaryotic proteins with N-terminal Ser (90%) and Ala (50%) are the most frequently acetylated, and these residues along with Met, Gly, and Thr account for over 95% of N-terminal acetylated residues (61, 64). Despite these substrate specificities of NATs, only a subset of proteins with these consensus residues are acetylated (63). Polevoda & Sherman (63, 64) reported that NatA and NatB have tendency to avoid basic (Lys, Arg, and His) and Pro residues at any positions on the nascent chain during translation, whereas NatC avoids acidic residues. Therefore, the overall efficiency of acetylation may be the net effect of acetylation-permissive and inhibitory residues near the N terminus. It remains unknown how many of N-terminally acetylated proteins undergo proteasomal degradation through N-recognin.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

RECOGNITION OF N-DEGRONS BY CANONICAL N-RECOGNINS

Substrate selectivity in the N-end rule pathway is governed by recognition of N-degrons by N-recognins, which induce, in eukaryotes, protein ubiquitylation and proteolysis through the proteasome. The mechanism of substrate selectivity was revealed by the discovery that the UBR box, conserved in many N-recognins, is the substrate recognition domain (65–67). Below, we discuss the enzymatic specificities, structures, functions, and substrates of N-recognins carrying the UBR box, with an emphasis on the mammalian N-recognin family in proteolytic and nonproteolytic physiological processes, including the pathogenesis of human genetic diseases.

The N-Recognin Family Containing the UBR Box

UBR boxes of *S. cerevisiae* Ubr1 and mammalian UBR1 and UBR2 (Ub ligase N-recognin1 and -2) bind type 1 residues with the dissociation constant of low micromolars (65–67). The mammalian genome encodes at least seven UBR box proteins, UBR1–UBR7 (Figure 4*a*) (65). With the exception of UBR4, the UBR box family members contain signatures of the substrate recognition components of the UPS (Figure 4*a*) (65). We here refer to UBR1/E3α, UBR2, and UBR3 as canonical, owing to their sequelogy, size (about 200 kDa), and conserved domains, including the UBR box (type 1 binding site), N domain(type 2 binding site), RING finger (ubiquitylation domain), and autoinhibitory domain (which sterically blocks the UBR box and N domain via intramolecular interaction) (65, 68). UBR4 through UBR7, referred to as noncanonical UBR box proteins, are evolutionarily divergent and nonsequelogous to each other. On the basis of binding and degradation assays, UBR1, UBR2, UBR4, and UBR5 are classified as N-recognins, and UBR3, UBR6, and UBR7 as non-N-recognins (65, 66).

Molecular Principles of N-End Rule Interaction

Recently determined crystal structures of UBR boxes of canonical N-recognins revealed the molecular principles of N-end rule interactions for type 1 degrons (7, 20, 21, 66; reviewed in References 6 and 7). The UBR box consists of two zinc fingers: a typical Cys_2His_2 motif containing a zinc ion and an atypical binuclear Cys_6His_1 motif containing two zinc ions (7, 20, 21). UBR boxes bind to positively charged N-terminal residues through a negatively charged, shallow binding groove (Figure 5*a,b*). The recognition of N-end rule substrates initiates with hydrogen bonding with the free α-amino group of the N-terminal residue, a unique structure conserved in all proteins. Once engaged with a genuine substrate while scanning the N termini, N-recognin establishes a substrate-selective interaction through hydrogen bonds with the positively charged side chains; this interaction is further supported by additional hydrogen bonds with the side chain of the second residues and the backbone atoms of the first three residues. Overall, N-end rule interactions are largely confined to the first two residues as the side chain at position three stays away from the surface of the UBR box (7, 20, 21), enabling N-recognins to select substrates on the basis of destabilizing N-terminal residues.

Canonical N-recognins have a second, structurally distinct substrate-binding domain, the N domain, which binds type 2 degrons (Phe, Trp, Tyr, Leu, and Ile) (Figure 4*a*) (66). The N domain appears to be derived from and spalogous (similar in structure) to bacterial N-recognin ClpS, ATP-dependent Clp protease adaptor protein, which recognizes the primary destabilizing residues (Phe, Trp, Tyr, and Leu) (69).

The Classical N-End Rule Pathway in *S. cerevisiae*

In yeasts, a single canonical N-recognin, Ubr1, recognizes type 1 (Arg, Lys, and His) and type 2 (Phe, Trp, Tyr, Leu, and Ile) N termini in concert with the Ub-conjugating enzyme

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Ubc2/Rad6 (Figure 1*b*) (2, 3). Substrates of Ubr1 include Scc1, a subunit of the cohesin complex, which holds the sister chromatids connected together during metaphase (70; reviewed in Reference 8). At the metaphase-to-anaphase transition, separase cleavage of Scc1 produces a C-terminal fragment with the degron Arg, which is degraded through Ubr1 for efficient chromatin separation. Notably, analogous cleavage events by separases in many eukaryotes generate destabilizing N-terminal residues on C-terminal Scc1 fragments (Arg in *S. cerevisiae*, Asn in *S. pombe*, Glu in mammals, and Cys in *D. melanogaster*) (8), indicating a strong evolutionary pressure to regulate SCC1 proteins by the N-end rule pathway. Ubr1 can also recognize internal degrons (non-N-degrons). This class of substrates includes the homeodomain protein Cup9 (a transcriptional repressor of the Ptr2 peptide transporter) (71–73), GPA1 (the Gα-subunit that controls signal transduction during mating) (74), and Mgt1 (the O6-alkylguanine-DNA-alkyltransferase) (75). In addition to selective proteolysis on the basis of protein-specific degrons of normally folded substrates, Ubr1 mediates protein quality control through degradation of misfolded proteins in conjunction with chaperons, such as Hsp70 and Sse1 (76, 77). Interestingly, the processivity of Ubr1 for some of these substrates was shown to be accelerated by the Ufd4 HECT domain E3 (78, 75, 79). Ufd4 is the recognition component of the Ub fusion degradation pathway that targets the noncleavable N-terminal Ub moiety as a degron (75, 78, 79). Ufd4 in complex with Ubr1 acts as an E4-like processive-enhancement cofactor for Lys48-linked ubiquitylation by Ubr1 once Ubr1 recognizes a substrate (8). This synergistic cotargeting was verified in ubiquitylation on the basis of both N-degrons and internal degrons (75, 78, 79).

Redundant Functions of Canonical N-Recognins in Mammalian Development

UBR1-deficient mice are alive but exhibit pleiotrophic phenotypes, such as defective growth, muscle protein degradation, and fat metabolism, as well as moderate hypoglycemia (80). The mutants also exhibit pancreatic abnormalities with defective secretion of digestive enzymes from exocrine cells, reminiscent of a human genetic disorder, Johanson-Blizzard syndrome (Online Mendelian Inheritance in Man code 243800; http://www.ncbi.nlm.nih.gov/omim) (81, 82). In humans, mutations in *UBR1* is the primary cause of this autosomal recessive disorder, characterized by exocrine pancreatic insufficiency and additional clinical features with aplasia or hypoplasia of the alae nasi, oligodontia, sensorineural hearing loss, hypothyroidism, scalp defects, mental retardation, and developmental delay (80, 82). *UBR1* (Online Mendelian Inheritance in Man code 605981) spanning a 161-kb region of chromosome 15q15-q21.1 is highly expressed in acinar cells of the pancreas (82). Several nonsense, splice site, and frameshift mutations as well as missense mutations of *UBR1* were identified (82, 83 and the references therein), among which complete or near-complete null mutations correlate to the Johanson-Blizzard syndrome. Thus, a UBR1-deficiency in humans and mice involves common pathogenic mechanisms with impaired excretion of zymogens and irreversible acinar cell damage.

UBR2-deficient mice develop distinct phenotypes depending on the genetic background and gender (84). In the 129/SvJ genetic background, the mutants of both genders die before adulthood. In contrast, the mutants in a mixed backgroud between the 129/SvJ and C57BL6 strains show infertility in males and partial lethality in females (84). Infertility is caused by arrest of spermatocytes at meiotic prophase I, resulting in apoptosis of spermatocytes and degeneration of testicular tubules (84). Substrates of UBR2 in spermatocytes include histone H2A (85). In vitro UBR2 functions as a scaffolding E3 ligase that mediates monoubiquitylation and polyubiquitylation of H2A by promoting the interaction between the E2 enzyme HR6B/RAD6 and H2A (85). In pachytene spermatocytes, ubiquitylated H2A is implicated in the transcriptional silencing of sex chromosomes. UBR2-deficient spermatocytes fail to maintain a normal level of ubiquitylated H2A and to induce transcriptional silencing of many genes linked to the X and Y chromosomes (85). Thus, in

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

meiotic spermatocytes, a portion of UBR2 molecules (*a*) is associated with meiotic chromatin and (*b*) regulates chromatin dynamics and gene expression through ubiquitylation of H2A.

UBR1 and UBR2 are similar in size (200 kDa), have 46% sequence identity, have conserved domains, and exhibit similar specificity to type 1 and type 2 residues (84). Knockout of *UBR1* in mice resulted in tissue-specific, partial inactivation of the N-end rule pathway (80). Cells lacking both UBR1 and UBR2 still contain significant activities for degradation of N-end rule substrates, perhaps because of functional redundancy between N-recognins (65). In contrast to single-knockout mice, embryos lacking both UBR1 and UBR2 die at about embryonic day 11.0, which is associated with impaired neurogenesis and cardiovascular development (86). During neurogenesis in the neural tube, double-mutant neural progenitors differentiate prematurely, resulting in depletion of self-renewing progenitors, consistent with impaired Notch-1 signaling, which controls the balance between proliferation and differentiation of neural cells (86).

OTHER N-RECOGNIN FAMILY MEMBERS

Canonical N-recognins (yeast Ubr1 and its sequelogs) mediate selective proteolysis on the basis of type 1 and type 2 destabilizing N-terminal residues. Recent studies identified new N-recognins that are nonsequelogous to canonical N-recognins but can recognize various N-degrons generated through protein-specific or global modifications. Below, we discuss the biochemical mechanisms, structures, functions, and substrates of recently identified N-recognins from mammals, plants, flies, yeasts, and bacteria.

Noncanonical N-Recognins in Mammals

Mammalian UBR4, an extraordinarily big protein (570 kDa) with no known ubiquitylation domain, has the UBR box and binds to both type 1 and type 2 residues (Figure 4*a*) (65). Mammalian UBR5 [alternatively called EDD (E3 ligase identified by differential display) or hHYD (homolog of the *Drosophila melanogaster* hyperplastic disc)] and its *Drosophila* sequelog selectively bind type 1 residues (65) and have a UBR box, a UBA (Ub association) domain (which can bind to a polyubiquitin chain), a HECT domain, a nuclear localization signal, and PABC domain (65). Knockdown of *UBR4* in cells lacking UBR1 and UBR2 almost completely impaired the type 2 pathway but not the type 1 pathway as determined by degradation of model substrates and the human immunodeficiency virus type 1 integrase bearing the N-degron Phe (65).

Mice lacking UBR4 die during embryogenesis with defects in angiogenesis (T. Tasaki & Y.T. Kwon, unpublished data). Mammalian UBR4 has been implicated in integrin-mediated signaling and cell adhesion (87), E7-mediated cellular immortalization (88, 89), and neuronal migration and development (90). Proteins that interact with mammalian UBR4 include molecules involved in the progression between G1 and S phases, such as the pRb retinoblastoma tumor suppressor (87), the E7 oncoprotein of human and bovine papillomavirus (88, 89), and calmodulin (87). Mice lacking UBR5 also die during embryogenesis with defects in the development of yolk sac vasculature (91). Mammalian UBR5 plays a role in various processes, such as progesterone-regulated cell proliferation, DNA repair, tumorigenesis, and smooth muscle differentiation (91), in part, through ubiquitylation of its known substrates, TopBP1 (92), the PABP-interacting protein 2 (Paip2) (93), and CDK9 (94), and through interaction with a number of cellular proteins: p53 (95); CIB/KIP (calcium- and integrin-binding protein/DNA-dependent protein kinase-interacting protein) (96); importin α5 (a component of the nuclear import complex) (96); and the progesterone receptor, PR (to potentiate progestin-mediated gene transactivation) (96).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Mammalian UBR4 and UBR5 have their homologs in various multicellular eukaryotes but not in fungi (65), and their physiological functions are described in the Figure 4*a* caption.

The Plant N-End Rule Pathway

The hierarchical structure, degrons, and components of the N-end rule pathway in the plant *Arabidopsis* are largely similar to those in mammals except for the genes encoding and specificities of N-recognins (10, 14). Two N-recognins, PRT1 and PRT6, have been identified in plants. PRT1, a 410-residue RING finger protein with a ZZ domain, is nonsequelogous to known N-recognins and does not have the UBR box and the N domain (97). The mutation of *PRT1* stabilizes model substrates bearing aromatic hydrophobic residues (Phe, Trp, and Tyr) at the N termini but not aliphatic hydrophobic residues, such as Leu and Ile (97). PRT6 is a 2006-residue RING finger E3 ligase sequelogous to yeast Ubr1 and has the UBR box but apparently not the N domain (98). *Arabidopsis prt6* mutants are impaired in degradation of type 1, but not type 2, substrates (98). Thus, the mutations in PRT1 and PRT6 selectively stabilize the N-terminal aromatic and basic destabilizing residues, respectively, with-out affecting substrates with a subset of type 2 degrons, Leu and Ile. It remains unknown whether PRT1 and PRT6 functionally overlap in recognition of the Leu and Ile N-degrons or whether there is yet another N-recognin.

The Acetylation-Based N-End Rule Pathway

A recent study showed that an acetylated N-terminal residue can act as an N-degron in the *S. cerevisiae* N-end rule pathway (5). N-terminal acetylation is irreversible and typically occurs cotranslationally at the retained N-terminal Met residue or at a newly exposed N-terminal amino acid (Ala, Val, Ser, Thr, or Cys) following N-terminal Met excision by Met aminopeptidases (99). The Doa10 E3 ligase in complex with the Ubc6 or Ubc7 E2 enzyme was identified as an N-recognin that targets at least eight substrates carrying acetylated N-degrons (5). Doa10 is a transmembrane protein, which mediates ubiquitination of misfolded proteins on the cytosolic surface of the ER membrane (100). It is unknown whether the acetylated N-degron-based proteolytic system targets the majority or only a few cellular proteins.

The Bacterial N-End Rule Pathway

In *E. coli*, N-terminal Leu, Phe, Trp, and Tyr are primary destabilizing residues for the bacterial N-recognin ClpS (Figures 1*c* and 4*b*) (12, 69, 101–103). ClpS mediates N-end rule degradation without Ub-like molecules (Figure 4*b*) (12, 69, 104); this is in contrast to mammals, where N-recognins use Ub as a secondary degron for proteasomal delivery (Supplemental Figure 3*b*). ClpS recognizes N-end rule substrates through a preformed, deep hydrophobic pocket within which the N-terminal side chain of substrates is completely buried, and the α-amino group and the first peptide bond make additional contacts through hydrogen bonding (Figure 5*c*,*d*) (102, 103). ClpS-bound substrates are delivered through a multistep process to the ClpAP complex, a ring-shaped proteolytic machinery that functions like the 26S proteasome (Supplemental Figure 3*a*) (104). The ClpP proteolytic complex, a counterpart of the 20S core particle, is a stack of two rings of heptameric complexes in which the active sites are exposed on the internal surface. The size of the chamber is relatively small in diameter (10 Å) so the AAA+ protein ClpA, which forms a hexameric ring, unfolds and feeds ClpS-bound substrates into the ClpP chamber (105). Docking and release of substrate-bound ClpS to ClpAP involve the interaction of ClpS with both the substrate and ClpA on the N- and C-terminal domains of ClpS, respectively (104, 106). While holding the substrate on the N-terminal domain, the C-terminal domain of ClpS docks on a highly mobile N-terminal domain of ClpA. This induces the movement of the ClpA N-terminal domain, placing ClpS-bound substrate in proximity to the ClpAP pore. Substrate-bound ClpS now undergoes a conformational change in the N-terminal region, resulting its

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

contact with a second unknown site on ClpA. This triggers a conformational change in ClpA and/or ClpS, which in turn results in the delivery of the substrate to the ClpAP complex.

Putative N-Recognins: *Drosophila* Inhibitors of Apoptosis

The apoptosis inhibitors (DIAP1 and DIAP2) and their mammalian homologs (XIAP, cIAPs, and ML-IAP) are RING E3s that suppress undesirable apoptotic activities by inhibiting the functions of initiator and effecter caspases (57, 107, 108). Ditzel & Meier with their coworkers (56, 107, 108) and Yoo et al. (109) proposed that DIAP1 acts as an N-recognin-like component that recognizes N-terminal Ala of peptidase-cleaved caspases as a major determinant in binding and subsequently mediates proteasomal degradation. DIAP1-dependent downregulation of caspases requires the binding of the DIAP1 BIR (baculovirus IAP repeat) domain to the N-terminally exposed IBM (IAP-binding motif) of caspases (110). The IBM is a tetrapeptide consensus sequence [Ala-X-(P/A)-Y, X and Y are hydrophobic] in which N-terminal unmodified Ala is essential for binding with the DIAP1 BIR, contributing to degradation of caspases and IAP antagonists and acting as an N-degron-like element. The IBMs of caspases are typically generated through cleavage by caspases and other proteases during apoptotic induction (111).

The crystal structures of the DIAP1 BIR2 domain in complex with a Hid or Grim peptide show that the BIR2-IBM interaction is similar to that of the UBR box with type 1 residues (Figure 5*e*,*f*) (112). The BIR2 binds the IBM on the surface of a hydrophobic groove formed by a ~70-residue zinc-finger motif (112). The binding involves two hydrogen bonds between the N-terminal α-amino group of the IBM and two negatively charged BIR2 residues (Asp277 and Gln282); this is supported by additional two hydrogen bonds between the first peptide bond of the IBM and surrounding BIR2 residues (112). The N-terminal side chain fits tightly in the binding pocket of the BIR domain. As N-terminal Ala binds to the BIR domain, the next six residues interact with the surface of the shallow groove (112). Thus, N-terminal Ala is an anchoring component in the BIR-IBM interaction, functioning as an N-degron-like determinant. This action mode in ubiquitylation of caspases and IAP antagonists defines DIAP1 and many other IAPs as N-recognin-like components in regulated proteolysis. It remains to be determined whether N-terminal Ala of the IBM is a strong degron.

UBIQUITIN ACTIVATION AND CONJUGATION IN THE N-END RULE PATHWAY

Ub can be activated by two Ub-activating enzymes, UBA1 and UBA6 (113–116) (Supplemental Figure 4). UBA1 is the major E1 Ub-activating enzyme responsible for the majority of Ub conjugation to E2 enzymes and substrates (113). UBA6, containing all three conserved domains of UBA1, is an alternative E1 enzyme, representing 5%–10% of UBA1 in abundance. In contrast to UBA1, which works with a broad range of E2s, UBA6 has a designated E2, USE1 (115). Homologs of human UBA6 and USE1 are found in mammals, zebra fish, and sea urchins, but not in worms, flies, and yeast. Mouse embryos lacking UBA6 (also termed E1-L2) die before midgestation (114), indicating the essential role of UBA6 in mammalian development. The physiological E3s and substrates of the UBA6-USE1 system have remained unknown until the recent discovery (116) that UBA6 mediates Ub activation and conjugation for the canonical N-recognins, UBR1 and UBR2, as well as for a non-N-recognin, UBR3. These E3 ligases mediate ubiquitylation with either UBA1-activated UBE2 (alternatively called RAD6) or UBA6-activated USE1. Substrates of UBA6-activated N-recognins include RGS4, whose ubiquitylation can be regulated by both cytosol-specific UBA6 and ubiquitous UBA1 (116).

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

REGULATION OF THE N-END RULE PATHWAY BY SMALL MOLECULES

The functions of the N-end rule pathway are regulated by various mechanisms, including transcription (80), posttranslational modifications (phosphorylation and ubiquitylation) (117, 118), and interactions with the Ub ligase Ufd4 (78, 75) and small molecules (71, 73, 117, 119). Below, we discuss how components of the N-end rule pathway interact with and react to small peptides and heme by adjusting cellular concentrations of short-lived proteins.

Regulation of the N-End Rule Pathway by Short Peptides

In *S. cerevisiae*, extracellular dipeptides and tripeptides are mainly imported by the transporter Ptr2 whose level is tightly regulated by the N-end rule pathway in response to changing nutritional availability (Supplemental Figure 5) (71, 73). When extracellular peptides are limited, the transcriptional repressor Cup9, a substrate of Ubr1, is accumulated and shuts down the transcription of Ptr2 to prevent unnecessary synthesis of the transporter machinery (71–73). However, if cells are exposed to a peptide-rich environment, Ubr1 binds to destabilizing N-terminal residues of constitutively imported small peptides and undergoes an allosteric conformational change, exposing its Cup9-binding site, which is otherwise masked by the C-terminal autoinhibitory domain through intramolecular interaction. Following activation by peptides, Ubr1 mediates the degradation of Cup9 [$t_{1/2}$ (half-life) < 1 min], leading to transcriptional derepression of Ptr2 and accelerated peptide import (72). This allosteric activation of Ubr1 is maximally induced when both type 1 and type 2 peptides bind to Ubr1 (72). Using this positive-feedback circuit, cells sense the levels of extra-cellular peptides and maintain the homeostasis in peptide uptake.

S. cerevisiae cells sense free amino acids, in particular Trp and Leu, as an indication of a protein-rich environment. Ubr1-dependent degradation of Cup9 links the availability of extracellular free amino acids to transcriptional induction of Ptr2 through the amino acid sensing SPS (Ssy1-Ptr3-Ssy5) pathway (119). Ssy1 is an integral membrane protein that can bind to amino acids. Amino acid–bound Ssy1, together with the peripheral membrane protein Ptr3, activates the chymotrypsin-like endoprotease Ssy5, which requires the casein kinases Yck1 and Yck2 (120). Activated Ssy5 in turn cleaves cytoplasmic retention sequences of the transcription factors Stp1 and Stp2, releasing them to induce the transcription of Ptr2 and other proteins involved in peptide transport. In this process, the kinases Yck1 and Yck2 phosphorylate Ubr1 at Ser300, and this is critical for efficient ubiquitylation of Cup9 and thus the induction of Ptr2 (Supplemental Figure 5*b*) (118).

Regulation of the N-End Rule Pathway by Hemin

Heme is an iron (Fe^{2+})-containing protoporphyrin IX, and hemin is its ferric (Fe^{3+}) counterpart. Hemoproteins, including globins, oxidoreductases (e.g., catalase, peroxidase, P450s, cytochrome oxidases, and nitric oxide synthases), and cytochromes, play a role in electron transfer, redox modification, and sensing of diatomic gases (e.g., oxygen and nitric oxide) (117). A recent study showed that hemin binds to components of the N-end rule pathway and inhibits their functions in the degradation of short-lived proteins (117). Upon binding to hemin, the functions of R-transferases in yeasts and mammals are dually downregulated, in part, by inactivation of enzymatic activities, which involves the formation of a disulfide bond between two adjacent Cys residues (Cys71 and Cys72 in mouse ATE1), and, in part, by metabolic destabilization through the UPS (117). In addition, hemin binds and inhibits in vitro the aminoacylation activity of the human arginyl-tRNA synthetase (the enzyme EC6.1.1.19) that synthesizes Arg-$tRNA^{Arg}$, a cofactor required for arginylation by R-transferases (121). Hemin also binds and prevents yeast Ubr1 from dipeptide-induced allosteric conformational changes, which would otherwise unmask the Cup9-binding site of Ubr1 from inhibition by the autoinhibitory domain (117). As a consequence, hemin-bound

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Ubr1 cannot properly ubiquitylate Cup9 even in the presence of dipeptides (117). These results suggest that the N-end rule pathway contributes to homeostasis of heme/hemin and other redox-related molecules by regulating concentrations of proteins involved in redox-dependent processes.

USING ENGINEERED N-DEGRONS AS A MOLECULAR TOOL

Engineered N-degrons have been used to control the metabolic stabilities of various proteins through the N-end rule pathway. Short peptides or small-molecule ligands carrying destabilizing amino acids have been used to inhibit, promote, or probe the interactions of N-recognins with substrates or to identify new N-recognins and their interactors. As the basic principles of N-end rule interactions are conserved across species in eukaryotes and prokaryotes (7, 20, 21, 103), the approaches described below should be generally applicable for a broad range of proteins.

Engineering N-Degrons to Control Protein Stability

The first engineered N-degron, the **X**-e^k extension, is an ~40-residue *E. coli* Lac repressor fragment bearing an N-terminal destabilizing residue (**X**) (1). The Ub-fusion technique (1, 122) was employed to generate **X** through deubiquitylating enzyme-mediated cotranslational cleavage of Ub-**X**-e^k-protein at the Ub-X junction (Supplemental Figure 6*a*). When fused with β-galactosidase ($t_{1/2} > 20$ h), the Arg-e^k extension induced rapid proteolysis in yeast cells with a half-life of ~2 min (Supplemental Figure 6*b*) (1). The **X**-e^k extension was used as a portable degron to control the concentration of *S. cerevisiae* ARD1 expressed from the *GAL10* promoter (123). Arg-e^k-ARD, ARD carrying Arg-e^k extension, was normally maintained at a physiologically functional level but was rapidly depleted by turning off the promoter. Ghislain et al. (124) employed a different strategy in which the function of a protein carrying a portable degron was controlled by the expression of Ubr1. Using this strategy, Arg-e^k-Spr54 was accumulated when Ubr1 expression from the GAL1 promoter was repressed and was rapidly depleted when Ubr1 expression was restored. Taxis et al. (125) reported an inducible N-degron using the *Tobacco etch virus* protease, which cleaves the Q**X** junction of the ENLYFQ**X** consensus sequence (X is an N-degron). This cleavage exposed an N-degron on the C-terminal fragment that was conditionally degraded by *Tobacco etch virus* protease expressed from the GAL1 promoter. Dohmen & Varshavsky (126, 127) developed a portable heat-inducible N-degron that can induce temperature-sensitive (ts) degradation through the N-end rule pathway (Supplemental Figure 6*c*). The ts N-degron is based on 21-kDa mouse dihydroforate reductase bearing the Arg degron. The ts Arg dihydroforate reductase is long-lived at 23°C but rapidly degraded at 37°C ($t_{1/2} < 10$ min) owing to a Pro-to-Leu mutation at position 66, which causes a ts conformational change at the N terminus (126–128). The ts Arg dihydroforate reductase was shown to induce conditional degradation of other proteins that were fused to their N termini without making individual ts mutants (126–128). Bernal & Venkitaraman (129) applied the heat-inducible degron strategy to avian *Gallus gallus* DT40 cells, which have high homologous recombination efficiency and grow at ambient temperatures. Portable degrons were also used to develop short-lived fluorescent reporters of gene expression and protein localization. Li et al. (130) engineered enhanced green fluorescent protein (EGFP) with the PEST degron of mouse ornithine decarboxylase to produce moderately short-lived EGFP ($t_{1/2} <$ ~2 h). Dantuma and coworkers (131, 132) tagged an engineered N-degron at the EGFP N terminus to generate fluorescent N-end rule substrates with half-lives of 10 min to 2 h. These substrates were used as reporters of the UPS to study proteasome inhibitors, ER stress, valosin-containing protein (p97), Ran-binding protein-2, and the N-end rule in chloroplasts (131, 133–135). Lee et al. (136) developed a real-time maker of caspase in cultured cells by inserting a caspase cleavage site between the N-degron and EGFP. Finally, Hackett et al.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

(137) developed cyan fluorescent protein-based fluorescent N-end rule substrates to study various transcriptional activities in *S. cerevisiae*.

Using Engineered N-Degrons as Molecular Probes to Dissect the Function of N-Recognins

The 12-mer X peptides derived from the Sindbis virus polymerase nsP4, an N-end rule substrate (138), was used as an affinity ligand to characterize the interaction of N-recognins with the N termini, leading to elucidation of the peptide-induced allosteric conformational change of yeast Ubr1 in peptide transport (72) and identification of substrate recognition domains of mammalian N-recognins (65, 66). An affinity-based proteomic study using X peptides reported the isolation of endogenous UBR4 and UBR5 from mammalian and *Drosophila* cells (65). The two substrate-binding sites of N-recognins were also exploited as a target of multivalent inhibitors (139–141). In a test-of-concept study (139), a heterotetramer comprising Arg-e^k-βgal and Leu-e^k-βgal, which expose type 1 and type 2 residues from the X-e^k extension, significantly inhibited the degradation of N-end rule substrates in yeast cells. A recent study reported a nonprotenaceous compound, RF-C11 (831 kDa), in which single amino acids (Arg and Phe) were linked by a C11 hydrocarbon chain (140). RF-C11 significantly inhibited the degradation of N-end rule substrates in vivo and in vitro, with higher efficacy compared to monovalent (e.g., dipeptides) or homodivalent controls (e.g., RR-C11 with two Arg ligands), and cardiac hypertrophic responses in cardiomyocytes (140).

THE N-DEGRON CODE: EVOLUTION AND BEYOND

Below, we present a model to reconstitute a series of evolutionary events that led to the current N-degron code comprising 13 amino acids in the genetic code. In this model, the eukaryotic N-end rule pathway was born with the N-recognin Ubr1 that utilizes the UBR box to target type 1 substrates as part of the UPS, whereas ancient bacteria adopted a distinct folding of ClpS to recognize type 2 degrons. Substrate specificities of the N-recognin were expanded when the *ClpS* gene of a prokaryotic endosymbiont was incorporated into *Ubr1*, perhaps about a billion years ago when prokaryotic endosymbionts in eukaryotic cells were converted to organelles, such as the mitochondrion, and possibly the ER as well (142). Because the zinc-finger motif of the UBR box is superior in structure and regulation (e.g., redox modification and allosteric modulation by N-end rule ligands), the *ClpS*-encoded N domain became structurally dependent on the UBR box. In parallel, R-transferase that had evolved from bacterial aminoacyl-tRNA transferases became functionally linked to Ubr1 and thus provided a licensing step prior to irreversible degradation by the proteasome. As mammals evolved, the N-end rule pathway added yet another regulatory mechanism by recruiting $Nt^{N,Q}$-amidase with a eukaryotic origin, which mediates deamidation of the pro-N-degrons Asn and Gln. An analogous process occurred in mammals with Nt^{N}-amidase and Nt^{Q}-amidase through convergence evolution. Although this model explains fairly well the hierarchical structures of the N-end rule pathways, there are several questions remaining to be answered. For example, the majority of known substrates carry the degron Arg or pro-N-degrons (Asn, Gln, Cys, Glu, and Asp) that work through the degron Arg. By sharp contrast, all the primary destabilizing residues (Lys, His, Trp, Phe, Tyr, Leu, and Ile) except for Arg are highly underrepresented in physiological substrates. One intriguing possibility is that these under-represented primary residues may represent not degrons but ligands (e.g., short peptides) that bind to and modulate the functions of N-recognins. As such, the existence of natural non-peptide compounds that induce allosteric conformational changes in N-recognins as part of a negative feedback loop is cautiously predicted.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

CONCLUDING REMARKS

Since 1986, the N-end rule pathway has long remained as an orphan proteolytic system without clear functions or substrates. Systematic genetic and biochemical dissection over the past decade revealed numerous insights into its components, functions, and mechanisms. The N-end rule pathway is now emerging as a major cellular proteolytic system in which the bulk of cellular proteins potentially carry N-terminal destabilizing residues at least temporarily during their life cycles. What should we focus on in the coming decade? One outstanding issue is the identity of substrates of the classical N-end rule pathway, which lists 13 out of 20 principal amino acids as degradation determinants, and for this reason, the pathway is anticipated to target a large number of substrates. Despite extensive efforts for the past 25 years, only a limited number of substrates have been identified, perhaps because the majority of substrates are conditionally generated under specific physiological processes, for example, apoptotic induction, cellular stresses, or cell cycle transition. Many N-degrons exposed after proteolytic cleavage may exist only briefly and at low concentrations prior to rapid degradation, making it difficult to detect such events under standard assay conditions. A reliable method for genome-wide screening is urgently needed to better understand the N-end rule pathway. One possible approach would be global protein stability profiling with genome-wide cDNA libraries to identify short-lived proteins whose ubiquitylation is sensitive to N-end rule inhibitors (30). It is also important to link the knowledge from known substrates to phenotypes identified from genetic studies in animals. *S. cerevisiae* has only one N-recognin, but mammals have at least four N-recognins whose functions may be overlapping in specific tissues, which makes it challenging to dissect the function of a specific N-recognin. It is still unclear how these mammalian N-recognins cooperate in degradation of N-end rule substrates and whether they recognize N-terminal and second residues of physiological substrates using similar molecular principles. We anticipate that the structures of large fragments containing both the UBR box and the N domain will be elucidated in few years. It is important to understand how small-molecule ligands induce allosteric conformational changes in two binding sites of N-recognins; this information will provide a structural basis for the design of potent inhibitors.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Dong Hoon Han and William T. Kwon for editorial assistance. Work in the authors' laboratories was supported by the National Institutes of Health grant (HL083365 to Y.T.K.) and World Class University (R31-2008-000-10103-0 to Y.T.K. and K.S.P.) and World Class Institute (WCI 2009-002 to Bo Yeon Kim) programs through the National Research Foundation funded by the Ministry of Education, Science and Technology, Korea. This article is dedicated to my father who is under post-stroke rehabilitation (Y.T.K.). Takafumi Tasaki is now affiliated with the Medical Research Institute, Kanazawa Medical University, Ishikawa, 920-0293, Japan.

Glossary

N-end rule: a rule that relates the in vivo half-life of a given protein to the destabilizing activity of its N-terminal residue

N-recognin: a recognition component of the N-end rule pathway that recognizes N-degrons

N-degron: a class of degrons in which the N-terminal destabilizing residue is the major degradation determinant in substrate recognition

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

E3 ligase	a protein that recognizes a specific substrate and accelerates the transfer of ubiquitin from an E2 enzyme to the substrate
Pro-N-degron	an N-terminal degradation determinant whose modification can generate an N-degron
UBR box	an ~70-residue zinc-finger motif that acts as a substrate recognition domain for type 1 substrates of the N-end rule pathway
RGS	regulator of G protein–signaling
ER	endoplasmic reticulum
DIAP1	*Drosophila* inhibitor of apoptosis 1
N domain	the type 2 substrate-recognition domain of N-recognins; this domain has a secondary structure similar to that of the ClpS N-recognin of prokaryotes
Sequelog	a protein whose nucleotide or amino acid sequence is similar, to a significant extent, to another sequence

LITERATURE CITED

1. Bachmair A, Finley D, Varshavsky A. In vivo half-life of a protein is a function of its amino-terminal residue. Science. 1986; 234:179–86. [PubMed: 3018930]
2. Bartel B, Wünning I, Varshavsky A. The recognition component of the N-end rule pathway. EMBO J. 1990; 9:3179–89. [PubMed: 2209542]
3. Varshavsky A. The N-end rule: functions, mysteries, uses. Proc. Natl. Acad. Sci. USA. 1996; 93:12142–49. [PubMed: 8901547]
4. Tasaki T, Kwon YT. The mammalian N-end rule pathway: new insights into its components and physiological roles. Trends Biochem. Sci. 2007; 32:520–28. [PubMed: 17962019]
5. Hwang CS, Shemorry A, Varshavsky A. N-terminal acetylation of cellular proteins creates specific degradation signals. Science. 2010; 327:973–77. [PubMed: 20110468]
6. Sriram SM, Kim BY, Kwon YT. The N-end rule pathway: emerging functions and molecular principles of substrate recognition. Nat. Rev. Mol. Cell Biol. 2011; 12:735–47. [PubMed: 22016057]
7. Sriram SM, Kwon YT. The molecular principles of N-end rule recognition. Nat. Struct. Mol. Biol. 2010; 17:1164–65. [PubMed: 20924402]
8. Varshavsky A. The N-end rule pathway and regulation by proteolysis. Protein Sci. 2011; 20:1298–345.
9. Gonda DK, Bachmair A, Wünning I, Tobias JW, Lane WS, Varshavsky A. Universality and structure of the N-end rule. J. Biol. Chem. 1989; 264:16700–12. [PubMed: 2506181]
10. Graciet E, Walter F, Maoiléidigh DO, Pollmann S, Meyerowitz EM, et al. The N-end rule pathway controls multiple functions during *Arabidopsis* shoot and leaf development. Proc. Natl. Acad. Sci. USA. 2009; 106:13618–23. [PubMed: 19620738]
11. Potuschak T, Stary S, Schlögelhofer P, Becker F, Nejinskaia V, Bachmair A. *PRT1* of *Arabidopsis thaliana* encodes a component of the plant N-end rule pathway. Proc. Natl. Acad. Sci. USA. 1998; 95:7904–8. [PubMed: 9653113]
12. Tobias JW, Shrader TE, Rocap G, Varshavsky A. The N-end rule in bacteria. Science. 1991; 254:1374–77. [PubMed: 1962196]
13. Dougan DA, Truscott KN, Zeth K. The bacterial N-end rule pathway: expect the unexpected. Mol. Microbiol. 2010; 76:545–58. [PubMed: 20374493]
14. Graciet E, Wellmer F. The plant N-end rule pathway: structure and functions. Trends Plant Sci. 2010; 15:447–53. [PubMed: 20627801]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

15. Balzi E, Choder M, Chen WN, Varshavsky A, Goffeau A. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. 1990; 265:7464–71. [PubMed: 2185248]
16. Graciet E, Hu RG, Piatkov K, Rhee JH, Schwarz EM, Varshavsky A. Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc. Natl. Acad. Sci. USA. 2006; 103:3078–83. [PubMed: 16492767]
17. Kaji H, Novelli GD, Kaji A. A soluble amino acid–incorporating system from rat liver. Biochim. Biophys. Acta. 1963; 76:474–77. [PubMed: 14097412]
18. Kwon YT, Kashina AS, Varshavsky A. Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. Mol. Cell. Biol. 1999; 19:182–93. [PubMed: 9858543]
19. Shrader TE, Tobias JW, Varshavsky A. The N-end rule in *Escherichia coli*: cloning and analysis of the leucyl, phenylalanyl-tRNA-protein transferase gene *aat*. J. Bacteriol. 1993; 175:4364–74. [PubMed: 8331068]
20. Choi WS, Jeong BC, Joo YJ, Lee MR, Kim J, et al. Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat. Struct. Mol. Biol. 2010; 17:1175–81. [PubMed: 20835240]
21. Matta-Camacho E, Kozlov G, Li FF, Gehring K. Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat. Struct. Mol. Biol. 2010; 17:1182–87. [PubMed: 20835242]
22. Ciechanover A, Ferber S, Ganoth D, Elias S, Hershko A, Arfin S. Purification and characterization of arginyl-tRNA-protein transferase from rabbit reticulocytes. Its involvement in post-translational modification and degradation of acidic NH2 termini substrates of the ubiquitin pathway. J. Biol. Chem. 1988; 263:11155–67. [PubMed: 2841321]
23. Kaji H. Further studies on the soluble amino acid incorporating system from rat liver. Biochemistry. 1968; 7:3844–50. [PubMed: 5722252]
24. Kwon YT, Kashina AS, Davydov IV, Hu RG, An JY, et al. An essential role of N-terminal arginylation in cardiovascular development. Science. 2002; 297:96–99. [PubMed: 12098698]
25. Soffer RL, Horinishi H. Enzymic modification of proteins. I. General characteristics of the arginine-transfer reaction in rabbit liver cytoplasm. J. Mol. Biol. 1969; 43:163–75. [PubMed: 5811819]
26. Hu RG, Brower CS, Wang H, Davydov IV, Sheng J, et al. Arginyltransferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem. 2006; 281:32559–73. [PubMed: 16943202]
27. Rai R, Kashina A. Identification of mammalian arginyltransferases that modify a specific subset of protein substrates. Proc. Natl. Acad. Sci. USA. 2005; 102:10123–28. [PubMed: 16002466]
28. Davydov IV, Varshavsky A. RGS4 is arginylated and degraded by the N-end rule pathway in vitro. J. Biol. Chem. 2000; 275:22931–41. [PubMed: 10783390]
29. Hu RG, Sheng J, Qi X, Xu Z, Takahashi TT, Varshavsky A. The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature. 2005; 437:981–86. [PubMed: 16222293]
30. Lee MJ, Tasaki T, Moroi K, An JY, Kimura S, et al. RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 2005; 102:15030–35. [PubMed: 16217033]
31. Gibbs DJ, Lee SC, Isa NM, Gramuglia S, Fukao T, et al. Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature. 2011; 479:415–18. [PubMed: 22020279]
32. Licausi F, Kosmacz M, Weits DA, Giuntoli B, Giorgi FM, et al. Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature. 2011; 479:419–22. [PubMed: 22020282]
33. Carpio MA, López Sambrooks C, Durand ES, Hallak ME. The arginylation-dependent association of calreticulin with stress granules is regulated by calcium. Biochem. J. 2010; 429:63–72. [PubMed: 20423325]
34. Saha S, Kashina A. Posttranslational arginylation as a global biological regulator. Dev. Biol. 2011; 358:1–8. [PubMed: 21784066]

NIH-PA Author Manuscript

35. Wang J, Han X, Saha S, Xu T, Rai R, et al. Arginyltransferase is an ATP-independent self-regulating enzyme that forms distinct functional complexes in vivo. Chem. Biol. 2011; 18:121–30. [PubMed: 21276945]

36. Brower CS, Varshavsky A. Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE. 2009; 4:e7757. [PubMed: 19915679]

37. Kurosaka S, Leu NA, Zhang F, Bunte R, Saha S, et al. Arginylation-dependent neural crest cell migration is essential for mouse development. PLoS Genet. 2010; 6:e1000878. [PubMed: 20300656]

38. Yoshida S, Ito M, Callis J, Nishida I, Watanabe A. A delayed leaf senescence mutant is defective in arginyl-tRNA:protein arginyltransferase, a component of the N-end rule pathway in *Arabidopsis*. Plant J. 2002; 32:129–37. [PubMed: 12366806]

39. Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, et al. The Berkeley *Drosophila* Genome Project gene disruption project: single *P*-element insertions mutating 25% of vital *Drosophila* genes. Genetics. 1999; 153:135–77. [PubMed: 10471706]

40. Leibowitz MJ, Soffer RL. Enzymatic modification of proteins. VII. Substrate specificity of leucyl, phenylalanyl-transfer ribonucleic acid-protein transferase. J. Biol. Chem. 1971; 246:5207–12. [PubMed: 4937659]

41. Watanabe K, Toh Y, Suto K, Shimizu Y, Oka N, et al. Protein-based peptide-bond formation by aminoacyl-tRNA protein transferase. Nature. 2007; 449:867–71. [PubMed: 17891155]

42. Ninnis RL, Spall SK, Talbo GH, Truscott KN, Dougan DA. Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in *Escherichia coli*. EMBO J. 2009; 28:1732–44. [PubMed: 19440203]

43. Varshavsky A. ‘Spalog’ and ‘sequelog’: neutral terms for spatial and sequence similarity. Curr. Biol. 2004; 14:R181–83. [PubMed: 15028230]

44. Rai R, Mushegian A, Makarova K, Kashina A. Molecular dissection of arginyltransferases guided by similarity to bacterial peptidoglycan synthases. EMBO Rep. 2006; 7:800–5. [PubMed: 16826240]

45. Suto K, Shimizu Y, Watanabe K, Ueda T, Fukai S, et al. Crystal structures of leucyl/phenylalanyl-tRNA-protein transferase and its complex with an aminoacyl-tRNA analog. EMBO J. 2006; 25:5942–50. [PubMed: 17110926]

46. Baker RT, Varshavsky A. Yeast N-terminal amidase. A new enzyme and component of the N-end rule pathway. J. Biol. Chem. 1995; 270:12065–74. [PubMed: 7744855]

47. Grigoryev S, Stewart AE, Kwon YT, Arfin SM, Bradshaw RA, et al. A mouse amidase specific for N-terminal asparagine. The gene, the enzyme, and their function in the N-end rule pathway. J. Biol. Chem. 1996; 271:28521–32. [PubMed: 8910481]

48. Wang H, Piatkov KI, Brower CS, Varshavsky A. Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell. 2009; 34:686–95. [PubMed: 19560421]

49. Cantor JR, Stone EM, Georgiou G. Expression and biochemical characterization of the human enzyme N-terminal asparagine amidohydrolase. Biochemistry. 2011; 50:3025–33. [PubMed: 21375249]

50. Kwon YT, Balogh SA, Davydov IV, Kashina AS, Yoon JK, et al. Altered activity, social behavior, and spatial memory in mice lacking the NTAN1p amidase and the asparagine branch of the N-end rule pathway. Mol. Cell. Biol. 2000; 20:4135–48. [PubMed: 10805755]

51. Balogh SA, Kwon YT, Denenberg VH. Varying intertrial interval reveals temporally defined memory deficits and enhancements in NTAN1-deficient mice. Learn. Mem. 2000; 7:279–86. [PubMed: 11040259]

52. Balogh SA, McDowell CS, Kwon YT, Denenberg VH. Facilitated stimulus-response associative learning and long-term memory in mice lacking the NTAN1 amidase of the N-end rule pathway. Brain Res. 2001; 892:336–43. [PubMed: 11172781]

53. Goto Y, Taniura H, Yamada K, Hirai T, Sanada N, et al. The magnetism responsive gene *Ntan1* in mouse brain. Neurochem. Int. 2006; 49:334–41. [PubMed: 16600435]

54. Hirai T, Taniura H, Goto Y, Ogura M, Sng JC, Yoneda Y. Stimulation of ubiquitin-proteasome pathway through the expression of amidohydrolase for N-terminal asparagine (Ntan1) in cultured

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

rat hippocampal neurons exposed to static magnetism. J. Neurochem. 2006; 96:1519–30. [PubMed: 16539681]

55. Pedersen LC, Yee VC, Bishop PD, Le Trong I, Teller DC, Stenkamp RE. Transglutaminase factor XIII uses proteinase-like catalytic triad to crosslink macromolecules. Protein Sci. 1994; 3:1131–35. [PubMed: 7920263]
56. Ditzel M, Broemer M, Tenev T, Bolduc C, Lee TV, et al. Inactivation of effector caspases through nondegradative polyubiquitylation. Mol. Cell. 2008; 32:540–53. [PubMed: 19026784]
57. Ditzel M, Wilson R, Tenev T, Zachariou A, Paul A, et al. Degradation of DIAP1 by the N-end rule pathway is essential for regulating apoptosis. Nat. Cell Biol. 2003; 5:467–73. [PubMed: 12692559]
58. Herman-Bachinsky Y, Ryoo HD, Ciechanover A, Gonen H. Regulation of the *Drosophila* ubiquitin ligase DIAP1 is mediated via several distinct ubiquitin system pathways. Cell Death Differ. 2007; 14:861–71. [PubMed: 17205079]
59. Tenev T, Ditzel M, Zachariou A, Meier P. The antiapoptotic activity of insect IAPs requires activation by an evolutionarily conserved mechanism. Cell Death Differ. 2007; 14:1191–201. [PubMed: 17347664]
60. Yan N, Wu JW, Chai J, Li W, Shi Y. Molecular mechanisms of DrICE inhibition by DIAP1 and removal of inhibition by Reaper, Hid and Grim. Nat. Struct. Mol. Biol. 2004; 11:420–28. [PubMed: 15107838]
61. Polevoda B, Arnesen T, Sherman F. A synopsis of eukaryotic Nalpha-terminal acetyltransferases: nomenclature, subunits and substrates. BMC Proc. 2009; 3(Suppl. 6):S2. [PubMed: 19660095]
62. Driessen HP, de Jong WW, Tesser GI, Bloemendal H. The mechanism of N-terminal acetylation of proteins. CRC Crit. Rev. Biochem. 1985; 18:281–325. [PubMed: 3902358]
63. Polevoda B, Sherman F. N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. 2003; 325:595–622. [PubMed: 12507466]
64. Polevoda B, Sherman F. Nalpha-terminal acetylation of eukaryotic proteins. J. Biol. Chem. 2000; 275:36479–82. [PubMed: 11013267]
65. Tasaki T, Mulder LC, Iwamatsu A, Lee MJ, Davydov IV, et al. A family of mammalian E3 ubiquitin ligases that contain the UBR box motif and recognize N-degrons. Mol. Cell. Biol. 2005; 25:7120–36. [PubMed: 16055722]
66. Tasaki T, Zakrzewska A, Dudgeon DD, Jiang Y, Lazo JS, Kwon YT. The substrate recognition domains of the N-end rule pathway. J. Biol. Chem. 2009; 284:1884–95. [PubMed: 19008229]
67. Xia Z, Webster A, Du F, Piatkov K, Ghislain M, Varshavsky A. Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. J. Biol. Chem. 2008; 283:24011–28. [PubMed: 18566452]
68. Kwon YT, Reiss Y, Fried VA, Hershko A, Yoon JK, et al. The mouse and human genes encoding the recognition component of the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 1998; 95:7898–903. [PubMed: 9653112]
69. Erbse A, Schmidt R, Bornemann T, Schneider-Mergener J, Mogk A, et al. ClpS is an essential component of the N-end rule pathway in *Escherichia coli*. Nature. 2006; 439:753–56. [PubMed: 16467841]
70. Rao H, Uhlmann F, Nasmyth K, Varshavsky A. Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature. 2001; 410:955–59. [PubMed: 11309624]
71. Byrd C, Turner GC, Varshavsky A. The N-end rule pathway controls the import of peptides through degradation of a transcriptional repressor. EMBO J. 1998; 17:269–77. [PubMed: 9427760]
72. Du F, Navarro-Garcia F, Xia Z, Tasaki T, Varshavsky A. Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain. Proc. Natl. Acad. Sci. USA. 2002; 99:14110–15. [PubMed: 12391316]
73. Turner GC, Du F, Varshavsky A. Peptides accelerate their uptake by activating a ubiquitin-dependent proteolytic pathway. Nature. 2000; 405:579–83. [PubMed: 10850718]
74. Madura K, Varshavsky A. Degradation of Gα by the N-end rule pathway. Science. 1994; 265:1454–58. [PubMed: 8073290]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

75. Hwang CS, Shemorry A, Varshavsky A. Two proteolytic pathways regulate DNA repair by cotargeting the Mgt1 alkylguanine transferase. Proc. Natl. Acad. Sci. USA. 2009; 106:2142–47. [PubMed: 19164530]

76. Eisele F, Wolf DH. Degradation of misfolded protein in the cytoplasm is mediated by the ubiquitin ligase Ubr1. FEBS Lett. 2008; 582:4143–46. [PubMed: 19041308]

77. Heck JW, Cheung SK, Hampton RY. Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. Proc. Natl. Acad. Sci. USA. 2010; 107:1106–11. [PubMed: 20080635]

78. Hwang CS, Shemorry A, Auerbach D, Varshavsky A. The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 2010; 12:1177–85. [PubMed: 21076411]

79. Johnson ES, Ma PC, Ota IM, Varshavsky A. A proteolytic pathway that recognizes ubiquitin as a degradation signal. J. Biol. Chem. 1995; 270:17442–56. [PubMed: 7615550]

80. Kwon YT, Xia Z, Davydov IV, Lecker SH, Varshavsky A. Construction and analysis of mouse strains lacking the ubiquitin ligase UBR1 (E3α) of the N-end rule pathway. Mol. Cell. Biol. 2001; 21:8007–21. [PubMed: 11689692]

81. Rezaei N, Sabbaghian M, Liu Z, Zenker M. Eponym: Johanson-Blizzard syndrome. Eur. J. Pediatr. 2011; 170:179–83. [PubMed: 20556422]

82. Zenker M, Mayerle J, Lerch MM, Tagariello A, Zerres K, et al. Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). Nat. Genet. 2005; 37:1345–50. [PubMed: 16311597]

83. Fallahi GH, Sabbaghian M, Khalili M, Parvaneh N, Zenker M, Rezaei N. Novel *UBR1* gene mutation in a patient with typical phenotype of Johanson-Blizzard syndrome. Eur. J. Pediatr. 2011; 170:233–35. [PubMed: 20556423]

84. Kwon YT, Xia Z, An JY, Tasaki T, Davydov IV, et al. Female lethality and apoptosis of spermatocytes in mice lacking the UBR2 ubiquitin ligase of the N-end rule pathway. Mol. Cell. Biol. 2003; 23:8255–71. [PubMed: 14585983]

85. An JY, Kim EA, Jiang Y, Zakrzewska A, Kim DE, et al. UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. Proc. Natl. Acad. Sci. USA. 2010; 107:1912–17. [PubMed: 20080676]

86. An JY, Seo JW, Tasaki T, Lee MJ, Varshavsky A, Kwon YT. Impaired neurogenesis and cardiovascular development in mice lacking the E3 ubiquitin ligases UBR1 and UBR2 of the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 2006; 103:6212–17. [PubMed: 16606826]

87. Nakatani Y, Konishi H, Vassilev A, Kurooka H, Ishiguro K, et al. p600, a unique protein required for membrane morphogenesis and cell survival. Proc. Natl. Acad. Sci. USA. 2005; 102:15093–98. [PubMed: 16214886]

88. DeMasi J, Chao MC, Kumar AS, Howley PM. Bovine papillomavirus E7 oncoprotein inhibits anoikis. J. Virol. 2007; 81:9419–25. [PubMed: 17596310]

89. Huh KW, DeMasi J, Ogawa H, Nakatani Y, Howley PM, Münger K. Association of the human papillomavirus type 16 E7 oncoprotein with the 600-kDa retinoblastoma protein-associated factor, p600. Proc. Natl. Acad. Sci. USA. 2005; 102:11492–97. [PubMed: 16061792]

90. Shim SY, Wang J, Asada N, Neumayer G, Tran HC, et al. Protein 600 is a microtubule/endoplasmic reticulum-associated protein in CNS neurons. J. Neurosci. 2008; 28:3604–14. [PubMed: 18385319]

91. Saunders DN, Hird SL, Withington SL, Dunwoodie SL, Henderson MJ, et al. *Edd*, the murine hyperplastic disc gene, is essential for yolk sac vascularization and chorioallantoic fusion. Mol. Cell. Biol. 2004; 24:7225–34. [PubMed: 15282321]

92. Honda Y, Tojo M, Matsuzaki K, Anan T, Matsumoto M, et al. Cooperation of HECT-domain ubiquitin ligase hHYD and DNA topoisomerase II–binding protein for DNA damage response. J. Biol. Chem. 2002; 277:3599–605. [PubMed: 11714696]

93. Yoshida M, Yoshida K, Kozlov G, Lim NS, De Crescenzo G, et al. Poly(A) binding protein (PABP) homeostasis is mediated by the stability of its inhibitor, Paip2. EMBO J. 2006; 25:1934–44. [PubMed: 16601676]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

94. Cojocaru M, Bouchard A, Cloutier P, Cooper JJ, Varzavand K, et al. Transcription factor IIS cooperates with the E3 ligase UBR5 to ubiquitinate the CDK9 subunit of the positive transcription elongation factor B. J. Biol. Chem. 2011; 286:5012–22. [PubMed: 21127351]
95. Ling S, Lin WC. EDD inhibits ATM-mediated phosphorylation of p53. J. Biol. Chem. 2011; 286:14972–82. [PubMed: 21383020]
96. Henderson MJ, Russell AJ, Hird S, Muñoz M, Clancy JL, et al. EDD, the human hyperplastic discs protein, has a role in progesterone receptor coactivation and potential involvement in DNA damage response. J. Biol. Chem. 2002; 277:26468–78. [PubMed: 12011095]
97. Stary S, Yin XJ, Potuschak T, Schlögelhofer P, Nizhynska V, Bachmair A. PRT1 of *Arabidopsis* is a ubiquitin protein ligase of the plant N-end rule pathway with specificity for aromatic amino-terminal residues. Plant Physiol. 2003; 133:1360–66. [PubMed: 14551326]
98. Garzón M, Eifler K, Faust A, Scheel H, Hofmann K, et al. *PRT6*/At5g02310 encodes an *Arabidopsis* ubiquitin ligase of the N-end rule pathway with arginine specificity and is not the *CER3* locus. FEBS Lett. 2007; 581:3189–96. [PubMed: 17572409]
99. Frottin F, Martinez A, Peynot P, Mitra S, Holz RC, et al. The proteomics of N-terminal methionine cleavage. Mol. Cell Proteomics. 2006; 5:2336–49. [PubMed: 16963780]
100. Swanson R, Locher M, Hochstrasser M. A conserved ubiquitin ligase of the nuclear envelope/ endoplasmic reticulum that functions in both ER-associated and Matalpha2 repressor degradation. Genes. Dev. 2001; 15:2660–74. [PubMed: 11641273]
101. Schmidt R, Zahn R, Bukau B, Mogk A. ClpS is the recognition component for *Escherichia coli* substrates of the N-end rule degradation pathway. Mol. Microbiol. 2009; 72:506–17. [PubMed: 19317833]
102. Schuenemann VJ, Kralik SM, Albrecht R, Spall SK, Truscott KN, et al. Structural basis of N-end rule substrate recognition in *Escherichia coli* by the ClpAP adaptor protein ClpS. EMBO Rep. 2009; 10:508–14. [PubMed: 19373253]
103. Wang KH, Roman-Hernandez G, Grant RA, Sauer RT, Baker TA. The molecular basis of N-end rule recognition. Mol. Cell. 2008; 32:406–14. [PubMed: 18995838]
104. Román-Hernández G, Hou JY, Grant RA, Sauer RT, Baker TA. The ClpS adaptor mediates staged delivery of N-end Rule substrates to the AAA+ ClpAP protease. Mol. Cell. 2011; 43:217–28. [PubMed: 21777811]
105. Effantin G, Ishikawa T, De Donatis GM, Maurizi MR, Steven AC. Local and global mobility in the ClpA AAA+ chaperone detected by cryo-electron microscopy: functional connotations. Structure. 2010; 18:553–62. [PubMed: 20462489]
106. Zeth K, Ravelli RB, Paal K, Cusack S, Bukau B, Dougan DA. Structural analysis of the adaptor protein ClpS in complex with the N-terminal domain of ClpA. Nat. Struct. Biol. 2002; 9:906–11. [PubMed: 12426582]
107. Ditzel M, Meier P. Ubiquitylation in apoptosis: DIAP1's (N-)en(d)igma. Cell Death Differ. 2005; 12:1208–12. [PubMed: 16094399]
108. Orme M, Meier P. Inhibitor of apoptosis proteins in *Drosophila*: gatekeepers of death. Apoptosis. 2009; 14:950–60. [PubMed: 19495985]
109. Yoo SJ, Huh JR, Muro I, Yu H, Wang L, et al. Hid, Rpr and Grim negatively regulate DIAP1 levels through distinct mechanisms. Nat. Cell Biol. 2002; 4:416–24. [PubMed: 12021767]
110. Vaux DL, Silke J. IAPs, RINGs and ubiquitylation. Nat. Rev. Mol. Cell Biol. 2005; 6:287–97. [PubMed: 15803136]
111. Choi YE, Butterworth M, Malladi S, Duckett CS, Cohen GM, Bratton SB. The E3 ubiquitin ligase cIAP1 binds and ubiquitinates caspase-3 and -7 via unique mechanisms at distinct steps in their processing. J. Biol. Chem. 2009; 284:12772–82. [PubMed: 19258326]
112. Wu JW, Cocina AE, Chai J, Hay BA, Shi Y. Structural analysis of a functional DIAP1 fragment bound to Grim and Hid peptides. Mol. Cell. 2001; 8:95–104. [PubMed: 11511363]
113. McGrath JP, Jentsch S, Varshavsky A. *UBA 1*: an essential yeast gene encoding ubiquitin-activating enzyme. EMBO J. 1991; 10:227–36. [PubMed: 1989885]
114. Chiu YH, Sun Q, Chen ZJ. E1-L2 activates both ubiquitin and FAT10. Mol. Cell. 2007; 27:1014–23. [PubMed: 17889673]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

115. Jin J, Li X, Gygi SP, Harper JW. Dual E1 activation systems for ubiquitin differentially regulate E2 enzyme charging. Nature. 2007; 447:1135–38. [PubMed: 17597759]
116. Lee PC, Sowa ME, Gygi SP, Harper JW. Alternative ubiquitin activation/conjugation cascades interact with N-end rule ubiquitin ligases to control degradation of RGS proteins. Mol. Cell. 2011; 43:392–405. [PubMed: 21816346]
117. Hu RG, Wang H, Xia Z, Varshavsky A. The N-end rule pathway is a sensor of heme. Proc. Natl. Acad. Sci. USA. 2008; 105:76–81. [PubMed: 18162538]
118. Hwang CS, Varshavsky A. Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 2008; 105:19188–93. [PubMed: 19033468]
119. Xia Z, Turner GC, Hwang CS, Byrd C, Varshavsky A. Amino acids induce peptide uptake via accelerated degradation of CUP9, the transcriptional repressor of the PTR2 peptide transporter. J. Biol. Chem. 2008; 283:28958–68. [PubMed: 18708352]
120. Ljungdahl PO. Amino-acid-induced signalling via the SPS-sensing pathway in yeast. Biochem. Soc. Trans. 2009; 37:242–47. [PubMed: 19143640]
121. Yang F, Xia X, Lei HY, Wang ED. Hemin binds to human cytoplasmic arginyl-tRNA synthetase and inhibits its catalytic activity. J. Biol. Chem. 2010; 285:39437–46. [PubMed: 20923763]
122. Lévy F, Johnsson N, Rümenapf T, Varshavsky A. Using ubiquitin to follow the metabolic fate of a protein. Proc. Natl. Acad. Sci. USA. 1996; 93:4907–12. [PubMed: 8643502]
123. Park EC, Finley D, Szostak JW. A strategy for the generation of conditional mutations by protein destabilization. Proc. Natl. Acad. Sci. USA. 1992; 89:1249–52. [PubMed: 1311090]
124. Ghislain M, Dohmen RJ, Lévy F, Varshavsky A. Cdc48p interacts with Ufd3p, a WD repeat protein required for ubiquitin-mediated proteolysis in *Saccharomyces cerevisiae*. EMBO J. 1996; 15:4884–99. [PubMed: 8890162]
125. Taxis C, Stier G, Spadaccini R, Knop M. Efficient protein depletion by genetically controlled deprotection of a dormant N-degron. Mol. Syst. Biol. 2009; 5:267. [PubMed: 19401679]
126. Dohmen RJ, Varshavsky A. Heat-inducible degron and the making of conditional mutants. Methods Enzymol. 2005; 399:799–822. [PubMed: 16338396]
127. Dohmen RJ, Wu P, Varshavsky A. Heat-inducible degron: a method for constructing temperature-sensitive mutants. Science. 1994; 263:1273–76. [PubMed: 8122109]
128. Lévy F, Johnston JA, Varshavsky A. Analysis of a conditional degradation signal in yeast and mammalian cells. Eur. J. Biochem. 1999; 259:244–52. [PubMed: 9914499]
129. Bernal JA, Venkitaraman AR. A vertebrate N-end rule degron reveals that Orc6 is required in mitosis for daughter cell abscission. J. Cell Biol. 2011; 192:969–78. [PubMed: 21422227]
130. Li X, Zhao X, Fang Y, Jiang X, Duong T, et al. Generation of destabilized green fluorescent protein as a transcription reporter. J. Biol. Chem. 1998; 273:34970–75. [PubMed: 9857028]
131. Dantuma NP, Lindsten K, Glas R, Jellne M, Masucci MG. Short-lived green fluorescent proteins for quantifying ubiquitin/proteasome-dependent proteolysis in living cells. Nat. Biotechnol. 2000; 18:538–43. [PubMed: 10802622]
132. Menéndez-Benito V, Heessen S, Dantuma NP. Monitoring of ubiquitin-dependent proteolysis with green fluorescent protein substrates. Methods Enzymol. 2005; 399:490–511. [PubMed: 16338378]
133. Yi H, Friedman JL, Ferreira PA. The cyclophilin-like domain of Ran-binding protein-2 modulates selectively the activity of the ubiquitin-proteasome system and protein biogenesis. J. Biol. Chem. 2007; 282:34770–78. [PubMed: 17911097]
134. Apel W, Schulze WX, Bock R. Identification of protein stability determinants in chloroplasts. Plant J. 2010; 63:636–50. [PubMed: 20545891]
135. Tresse E, Salomons FA, Vesa J, Bott LC, Kimonis V, et al. VCP/p97 is essential for maturation of ubiquitin-containing autophagosomes and this function is impaired by mutations that cause IBMPFD. Autophagy. 2010; 6:217–27. [PubMed: 20104022]
136. Lee P, Beem E, Segal MS. Marker for real-time analysis of caspase activity in intact cells. BioTechniques. 2002; 33:1284–87. [PubMed: 12503313]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

137. Hackett EA, Esch RK, Maleri S, Errede B. A family of destabilized cyan fluorescent proteins as transcriptional reporters in *S. cerevisiae*. Yeast. 2006; 23:333–49. [PubMed: 16598699]

138. de Groot RJ, Rümenapf T, Kuhn RJ, Strauss JH. Sindbis virus RNA polymerase is degraded by the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 1991; 88:8967–71. [PubMed: 1924357]

139. Kwon YT, Lévy F, Varshavsky A. Bivalent inhibitor of the N-end rule pathway. J. Biol. Chem. 1999; 274:18135–39. [PubMed: 10364269]

140. Lee MJ, Pal K, Tasaki T, Roy S, Jiang Y, et al. Synthetic heterovalent inhibitors targeting recognition E3 components of the N-end rule pathway. Proc. Natl. Acad. Sci. USA. 2008; 105:100–5. [PubMed: 18162545]

141. Sriram SM, Banerjee R, Kane RS, Kwon YT. Multivalency-assisted control of intracellular signaling pathways: application for ubiquitin-dependent N-end rule pathway. Chem. Biol. 2009; 16:121–31. [PubMed: 19246002]

142. Dyall SD, Brown MT, Johnson PJ. Ancient invasions: from endosymbionts to organelles. Science. 2004; 304:253–57. [PubMed: 15073369]

143. Tasaki T, Sohr R, Xia Z, Hellweg R, Hörtnagl H, et al. Biochemical and genetic studies of UBR3, a ubiquitin ligase with a function in olfactory and other sensory systems. J. Biol. Chem. 2007; 282:18510–20. [PubMed: 17462990]

144. Meisenberg C, Tait PS, Dianova II, Wright K, Edelmann MJ, et al. Ubiquitin ligase UBR3 regulates cellular levels of the essential DNA repair protein APE1 and is required for genome stability. Nucleic Acids Res. 2012; 40:701–11. [PubMed: 21933813]

145. Takeda K, Yanagida M. Regulation of nuclear proteasome by Rhp6/Ubc2 through ubiquitination and destruction of the sensor and anchor Cut8. Cell. 2005; 122:393–405. [PubMed: 16096059]

146. Wang L, Mao X, Ju D, Xie Y. Rpn4 is a physiological substrate of the Ubr2 ubiquitin ligase. J. Biol. Chem. 2004; 279:55218–23. [PubMed: 15504724]

147. Yamaguchi N, Suzuki M, Fukaki H, Morita-Terao M, Tasaka M, Komeda Y. CRM1/BIG-mediated auxin action regulates *Arabidopsis* inflorescence development. Plant Cell. Physiol. 2007; 48:1275–90. [PubMed: 17652113]

148. Abida WM, Nikolaev A, Zhao W, Zhang W, Gu W. FBXO11 promotes the neddylation of p53 and inhibits its transcriptional activity. J. Biol. Chem. 2007; 282:1797–804. [PubMed: 17098746]

149. Rye MS, Bhutta MF, Cheeseman MT, Burgner D, Blackwell JM, et al. Unraveling the genetics of otitis media: from mouse to human and back again. Mamm. Genome. 2011; 22:66–82. [PubMed: 21107580]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

SUMMARY POINTS

1. The N-end rule defines the destabilizing activity of a given N-terminal residue and its posttranslational modification. N-recognins recognize destabilizing N-terminal residues as essential elements of N-degrons and mediate protein degradation.
2. Pro-N-degrons can be converted to N-degrons through conjugation of destabilizing amino acids by aminoacyl-tRNA transferases, which transfer amino acids from aminoacyl-tRNA to the N termini. Eukaryotic R-transferases catalyze arginylation, generating the degron Arg, whereas prokaryotic leucyl/pheylalanyl-tRNA-protein (L/F)-transferases and L-transferases catalyze leucylation and phenylalanylation, generating the degrons Leu and Phe.
3. The tertiary destabilizing residues Asn and Gln are deamidated into the secondary destabilizing residues Asp and Glu, which are subsequently arginylated by R-transferase, generating the primary destabilizing residue Arg.
4. Known mammalian N-recognins bind to type 1 and type 2 substrates through the UBR box and the N domain. Positively charged type 1 residues bind to the negatively charged surface of the UBR box. The N domain is homologous to bacterial ClpS, which binds bulky hydrophobic type 2 residues through a deep hydrophobic pocket.
5. The N-end rule pathway can be regulated by small molecules, such as short peptides and hemin. Short peptides bind the UBR box and the N domain of yeast Ubr1 and synergistically induce the proteolysis of substrates carrying internal degrons, such as Cup9, as a feedback mechanism to maintain the homeostasis of peptide transport.

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

FUTURE ISSUES

1. Although a number of physiological substrates of the N-end rule pathway have been discovered, many more are likely to remain unknown. A reproducible assay system is urgently needed to screen substrates of the N-end rule pathway.
2. What are the mechanisms that regulate activation of pro-N-degrons, generation of N-degrons, and activity of N-recognins?
3. How are physiological processes in mammals regulated by selective proteolysis of substrates bearing N-degrons? How do mammalian N-recognins cooperate in targeting N-degrons?
4. The structure of the N domain is unknown. It is anticipated that the binding of small-molecule ligands to the UBR box and the N domain of canonical N-recognins synergistically induces a conformational change to expose a site that recognizes internal degrons of substrates.
5. The bulk of eukaryotic proteins (50%–80%) can be acetylated at the N termini. What is the physiological meaning of the acetylation-based N-end rule pathway?

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

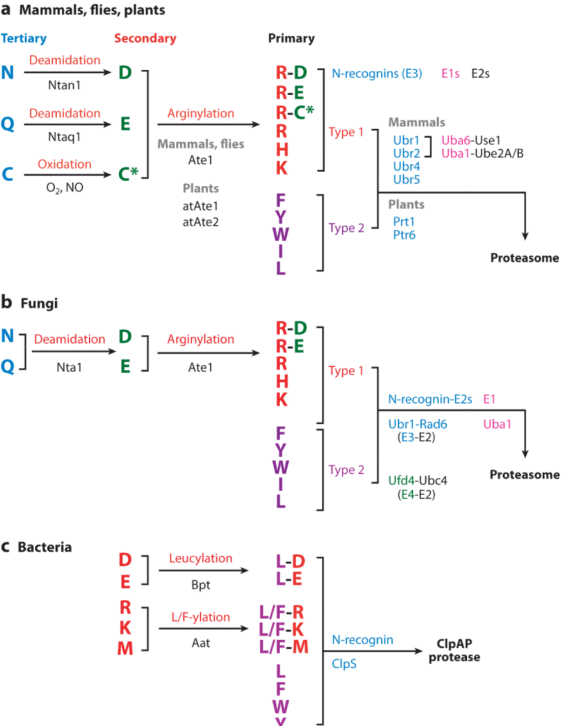

Figure 1.

The classical N-end rule pathway in various eukaryotes and prokaryotes. (*a*) The N-end rule pathway in mammals, flies, and plants. In mammals and other multicellular eukaryotes, the tertiary destabilizing residues Asn and Gln are, respectively, deamidated into the secondary destabilizing residues Asp and Glu by NTAN1 Nt^N-amidase and NTAQ1 Nt^Q-amidase, which in turn are arginylated by *ATE1*-encoded arginyl (R)-transferase isoforms generating the degron Arg. In mammals, N-terminal Cys is converted to a substrate of arginylation through its oxidation into $CysO_2(H)$ or $CysO_3(H)$ prior to arginylation. N-terminal Arg, together with other type 1 and type 2 residues, are recognized and bound by the N-recognin family members, which mediate ubiquitylation and proteasomal degradation, characterized by the UBR box in mammals. Although the components of the plant *Arabidopsis* and fly *Drosophila* N-end rule pathways are not fully characterized, their hierarchical structures appear to be more similar to the mammalian pathway compared to the yeast pathway. In contrast to mammals, the plant *Arabidopsis* genome expresses two distinct R-transferases, AtATE1 and AtATE2, from separate genes. To date, two plant N-recognins, PRT1 and PRT6, have been identified. (*b*) The *S. cerevisiae* N-end rule pathway. A single N-terminal amidohydrolase, Nta1 ($Nt^{N,Q}$-amidase), mediates deamidation of N-terminal Asn and Gln into Asp and Glu, which in turn are arginylated by a single Ate1 R-transferase, generating the degron Arg. N-terminal Arg and other primary degrons are recognized by a single N-recognin Ubr1. (*c*) The bacterial N-end rule pathway. The secondary destabilizing residues Arg, Lys, and Met are conjugated with Leu or Phe by the Aat leucyl/pheylalanyl-tRNA-protein (L/F)-transferase, generating the degrons Leu and Phe. The secondary destabilizing residues Asp and Glu can also be conjugated with Leu by Bpt leucyl-tRNA-protein (L)-transferase, generating the degron Leu. The primary destabilizing residues (Leu, Phe, Trp, and Tyr) are recognized and bound by the N-recognin ClpS, which delivers substrates to the ClpAP protease complex without involving Ub or Ub-like molecules. Abbreviations: Aat, aminoacyl transferase; Bpt, bacterial protein transferase; C*, oxidized Cys; E1, Ub activating enzyme; E2, Ub conjugating enzyme; E3, Ub protein ligase; E4, Ub conjugation

factor; L/F-ylation, leucylation/phenylalanylation; N, asparagine; Q, glutamine; Uba1 and Uba6, Ub activating enzymes 1 and 6; Ubc4, Ub conjugating enzyme 4; Ube2A/B, Ub conjugating enzyme E2 A/B; UBR box, Ub ligase N-recognin box; Ubr1, -2, -4, -5, Ub ligase N-recognin1, -2, -4, -5; Ufd4, Ub fusion degradation 4; Use1, Uba6-specific E2 1.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Figure 2.

Aminoacyl transferases of the N-end rule pathway and structurally related proteins. Eukaryotic and prokaryotic aminoacyl-tRNA (Aa)-transferases can be categorized into arginyl (R)-transferases, leucyl/pheylalanyl-tRNA-protein (L/F)-transferases, and other Aa-transferases on the basis of their enzymatic properties (16, 19, 23). ATE R-transferase and Aat L/F-transferase families can mediate the conjugation of destabilizing amino acids to the N termini of N-end rule substrates, whereas FemX and FemA of the FemABX family mediate the conjugation of amino acids to peptidoglycan pentapeptides, whose residues are shown as the capital letter X. Large (*bright blue*) and small (*orange*) ovals are the conserved ATE-N and AAT-N domains, respectively. Beige ovals represent the protein bodies (substrates) carrying arginylation-permissive N-terminal residues. Hexagons show the evolutionarily conserved GCN5-related N-acetyltransferase (GNAT) fold domains in ATE1 homologs, Aat homologs, and the FemABX family. Structural studies of *Escherichia coli* Aat L/F-transferase and *Weissella viridescens* FemX suggest that this GNAT fold domain is important for recognition of the donor aminoacyl-tRNA and for the enzymatic activity of the transferases (41, 45). Abbreviations: Cys*, the oxidized Cys residue of the acceptor substrate of R-transferase; *S. aureus*, *Staphylococcus aureus*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

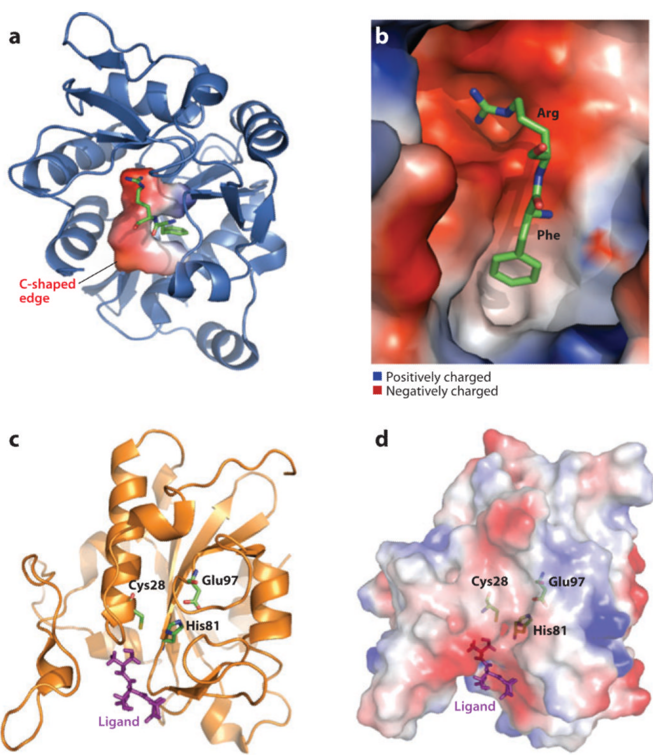

Figure 3.

The structures and binding sites of the *E. coli* leucyl/pheylalanyl-tRNA-protein (L/F)-transferase and human NTAQ1. (*a*) Crystal structure of the *E. coli* L/F-transferase complex [Protein Data Bank (PDB) code 2Z3L] with a product peptide. Only the Phe-Arg residues of the bound FRYLG peptide are shown here in a green-colored stick model. The C-shaped structure on the edge of the L/F-transferase hydrophobic pocket is also shown. (*b*) Electrostatic surface potential of L/F-transferase. Blue and red represent positively and negatively charged areas, respectively. The Phe and Arg residues of the FRYLG product peptide are shown in a green-colored stick model. Note that the side chain of Phe resides in the hydrophobic pocket, and the side chain of Arg resides in the negatively charged pocket. (*c*) Crystal structure of human NTAQ1 (PDB code 3C9Q) bound with a peptide-like ligand. The ligand is shown in a purple-colored stick model. NTAQ1 is a monomeric globular protein with a sandwich architecture. The catalytic triad (Cys28, His81, and Glu97) resembles an active site, which is conserved in NTAQ1 proteins. (*d*) Electrostatic surface potential of human NTAQ1. Blue and red colored surface depict positively and negatively charged areas, respectively. Note that the catalytic triad residues are in proximity to the binding pocket, bound by the ligand for deamidation.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

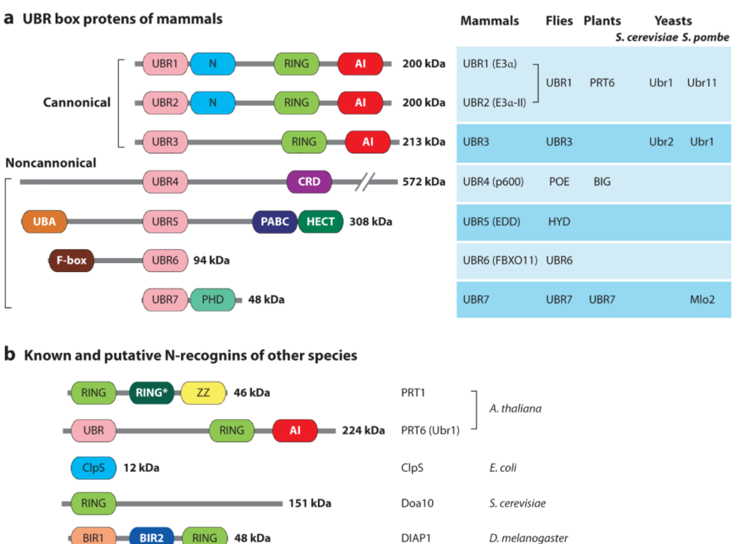

Figure 4.

The UBR box protein family and N-recognins. (*a*) The mammalian UBR box protein family. UBR1 and UBR2 (200 kDa) are functionally overlapping canonical UBR box N-recognins. UBR3 (213 kDa) is a canonical UBR box protein but does not show affinity to N-end rule peptides (65, 66, 143). Knockout of *UBR3* in mice resulted in neonatal death associated with female-specific anosmia, a finding consistent with its unique expression in neural tissues of the so-called five senses (olfaction, hearing, vision, touch, and taste). UBR3 mediates degradation of the DNA repair protein APE1 and is required for genomic stability (144). UBR3 is sequelogous and thought to be a functional homolog to *S. cerevisiae* Ubr2 and *Schizosaccharomyces pombe* Ubr1. The homologs in yeasts are involved in transcriptional regulation of the proteasome (through degradation of the transcriptional activator Rpn4), sexual differentiation, nuclear enrichment of the proteasome (through degradation of the nuclear envelope protein Cut8), and the oxidative stress response (145, 146 and references therein; reviewed in References 3, 4, and 8). UBR4 is a sequelog of the *Arabidopsis* BIG, which plays a role in auxin transport, root hair elongation, hormone and light responses, and the regulation of sulfur deficiency-responsive genes (147 and references therein; reviewed in Reference 4). The *Drosophila* homolog of UBR4, POE/PUSH/CALO, has been implicated as an interactor of calmodulin in the retina and the polar granule molecules Vasa and Tudor in germ plasm from early embryos; synaptic transmission; perineurial glial growth; male sterility; and meiotic chromosome pairing, recombination, and segregation in females (reviewed in Reference 4). UBR5 is a sequelog of the *Drosophila* HYD, whose mutations result in imaginal disc hyperplasia associated with uncontrolled cell proliferation through the independent activation of *hedgehog* and *decapentaplegic* (reviewed in Reference 4). UBR6/FBXO11 (94 kDa), a component of a SCF E3 complex, has been implicated in the neddylation of p53 (148) and the human diseases vitiligo (a skin disorder) and otitis media (a common childhood disease characterized by middle ear inflammation following infection) (149). UBR7 (48 kDa) and its sequelogs in multicellular organisms have a PHD domain, which resembles the RING domain. *S. pombe* mlo2, a putative UBR7 homolog, is implicated in chromosome transmission fidelity in mitosis (reviewed in Reference 4). (*b*) Known and putative N-recognins of other species. Abbreviations: AI, autoinhibitory domain; BIR; baculoviral inhibition of apoptosis protein repeat; ClpS, ATP-dependent Clp protease adaptor protein; CRD, cystein-rich domain; Doa10, the ER-localized ubiquitin ligase Doa10; DIAP1, *Drosophila* inhibitor of apoptosis 1; PHD, plant homeodomain finger; HECT, homologous to the E6-AP C terminus; N, N domain; PABC, poly(A)-binding protein

C-terminal domain; RING*, composite domain containing RING and CCCH-type Zn fingers; RING, RING finger; UBA domain, Ub association domain; UBR, UBR box; ZZ, a specific zinc finger domain that binds to two zinc ions.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

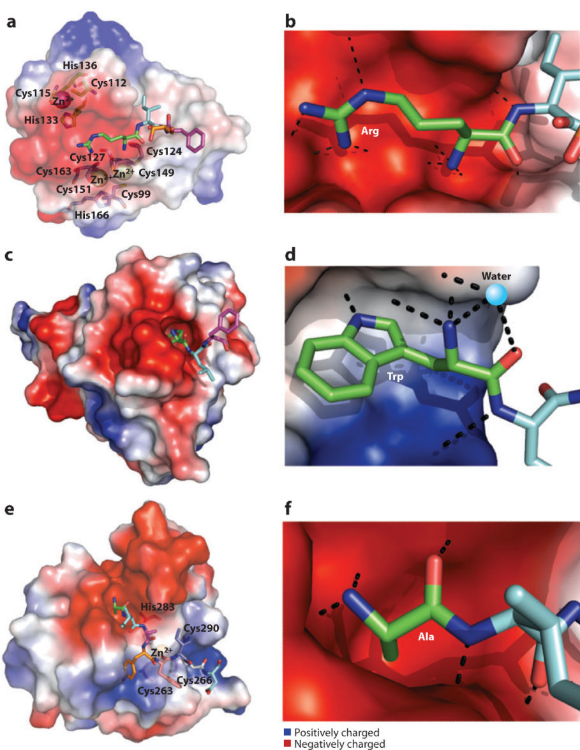

Figure 5.

Comparison of structures and binding sites of the mammalian UBR box, the bacterial ClpS, and the *Drosophila* BIR2 domain. (*a*) Electrostatic potential of human UBR2 UBR box bound with the peptide Arg-Ile-Phe-Ser [Protein Data Bank (PDB) code 3NY3]. The N-terminal Arg binds to a negatively charged, shallow binding groove, the second residue binds to a hydrophobic pocket, and the other following residues are turned away from the UBR box surface. The white surface color indicates a hydrophobic surface. Note the mononucleate (Cys_2His_2) zinc finger of the UBR box on top of the peptide and the unique binucleate zinc finger directly behind the UBR box contact surface with the peptide. (*b*) Key N-terminal recognition hydrogen bonds (*dotted lines*) of UBR box residues with Arg. Arg atoms are coded by color as follows: green, carbon; red, oxygen; and blue, nitrogen. Note the characteristic interactions, including the electrostatic interaction of the basic Arg side chain with the acidic UBR box surface and the H-bonding interactions from the α-amino group, side chain, and the peptide bond. (*c*) Electrostatic potential of *C.aulobacter crescentus* ClpS bound with the peptide Trp-Leu-Phe (PDB code 3GQ1). The hydrophobic N-terminal residue binds to a deep hydrophobic pocket, and the interaction from the second residue occurs outside of the pocket. (*d*) Key hydrogen bonds (*dotted lines*) of ClpS residues with N-terminal Trp. Trp atoms are coded by color as indicated above. The blue sphere is a water molecule. Note the characteristic interactions, including the hydrophobic interaction of Trp side chain with the hydrophobic ClpS pocket and the H-bonding interactions from the α-amino group, side chain, and the peptide bond. (*e*) Electrostatic potential of the *Drosophila melanogaster* DIAP1 BIR2 domain with Grim peptide (PDB code 1JD5). The eight-residue peptide (AIAYFIPD) binds the BIR2 surface groove in an extended conformation, and this binding induces a major conformational switch in DIAP1, although the BIR2 binding grove itself is fairly rigid. Note the DIAP1 zinc finger (Cys_3His). (*f*) Key hydrogen bonds (*dotted lines*) of BIR2 residues with the N-terminal Ala of the Grim peptide. Ala atoms are coded by color as indicated above. Note that the N-terminal Ala is positioned in a highly negatively charged BIR2 environment, with the main recognition specificity provided by the H-

bonding from the α-amino group and the first peptide bond. This mode of interaction of the N-terminal Ala in Grim peptide to the BIR2 domain of DIAP1 is similar to that of the UBR box interaction with N-terminal Arg, indicating a putative N-recognin behavior of DIAP1.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript