BIOCHEMICAL AND GENETIC STUDIES OF THE N-END RULE PATHWAY IN YEAST AND MAMMALS

Thesis by

Brandon Wadas

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2016

(Defended May 31, 2016)

ACKNOWLEDGEMENTS

I would like to first thank my advisor, Alexander Varshavsky, without whom this work would not have been possible. Alex's creativity and enthusiasm for science played significant roles in my graduate research. His support was instrumental not only in the completion of these experiments, but also in my development as a scientist.

I would also like to thank each of the current and former members of the Varshavsky. I had many terrific collaborators during my time in the Varshavsky Lab, in particular, Cheol-Sang Hwang, whose earlier discoveries in the lab led to many of the experiments performed during my graduate work. Konstantin Piatkov and Chris Brower were valuable resources for the work on Ate1 arginyltransferase, and other work in mammalian systems. Tri Vu, as my bench mate, was often my sounding board whether experiments were working, failing, or I just needed an opinion on a new idea. Our lab manager, Elena Udartseva, kept the lab running efficiently and ensured we could always focus on our research.

I also want to thank my committee, Raymond Deshaies, William Dunphy, and Shu-ou Shan for their support during my time at Caltech.

Lastly, I want to thank my fiancée and best friend, Taryn Aguiar. Her support through all of the late nights and weekends in the lab was an important part of my journey at Caltech.

I would like to dedicate this thesis to my parents, Robert and Jennifer Wadas, and my grandparents, Donna Wadas, and Richard and Audrey O'Lone, who have been indefatigable supporters of myself and my sisters.

ABSTRACT

Regulation of the *in vivo* half-lives of intracellular proteins is an important cellular process. Many intracellular proteins are short-lived, owing to their regulated and processive degradation by the Ubiquitin (Ub)-Proteasome System (UPS). In eukaryotes, the N-end rule pathway is one specific pathway within the UPS. The N-end rule pathway relates the identity of the N-terminal residue of a protein, or a protein fragment, to its *in vivo* half-life. Substrates of the N-end rule pathway are recognized by the presence of degradation signals, termed N-degrons. Recognition components of the N-end rule pathway are E3 ubiquitin ligases that are capable of binding to N-degrons. The N-end rule pathway consists of two distinct branches: the Arg/N-end rule pathway and the Ac/N-end rule pathway.

In the present studies, we demonstrate a complementary targeting of the rat serotonin N-acetyltransferase (AANAT), an important mediator of circadian physiology, by both branches of the N-end rule pathway. The co-targeting results from incomplete N-terminal (Nt-) acetylation of a Met-$ motif at the N-terminus of AANAT *in vivo*. In the same study, we demonstrate that human AANAT is substantially longer-lived than its rat counterpart, owing to differences in their N-terminal sequences. This molecular genetic investigation of the degradation of a physiological N-end rule substrate followed an analogous earlier study, in which we reported that a clinically-relevant (blood pressure-increasing) Q2L mutant of human RGS2 (termed ML-RGS2), a regulator of G proteins, could likewise be co-targeted by both branches of the N-end rule pathway. Together, AANAT and RGS2 are the first identified and characterized physiological substrates of the Ac/N-end rule pathway in mammals.

We also report on the development and use of *in vitro* N-terminal arginylation (Nt-arginylation) assays using CelluSpots peptide arrays, in conjunction with pulse-chase assays in rabbit reticulocyte extract, for the systematic investigation of the effects of N-terminus-proximal sequence context on the Nt-arginylation activity of the Ate1 arginyltransferase, a component of the Arg/N-end rule pathway. These experiments help to define the sequence requirements for efficient Nt-arginylation by Ate1. Finally, we demonstrate that Rec8, a subunit of the cohesin protein complex during meiosis, is a natural short-lived substrate of the mammalian Arg/N-end rule pathway.

PUBLISHED CONTENT AND CONTRIBUTIONS

Brower, C. S., Rosen, C. E., Jones, R. H., Wadas, B. C., Piatkov, K. I., and Varshavsky, A. (2014) Liat1, an arginyltransferase-binding protein whose evolution among primates involved changes in the numbers of its 10-residue repeats. Proc. Natl. Acad. Sci. USA 111, E4936-4945 doi: 10.1186/s12967-015-0614-x

B.C.W. contributed to the design of experiments, contributed new reagents/analytical tools, and helped write the paper.

Liu, Y. J., Liu, C., Chang, Z., Wadas, B., Brower, C. S., Song, Z. H., Xu, Z. L., Shang, Y. L., Liu, W. X., Wang, L. N., Dong, W., Varshavsky, A., Hu, R. G., and Li, W. (2016) Degradation of the separase-cleaved Rec8, a meiotic cohesin subunit, by the N-end rule pathway. J. Biol. Chem. 291(14), 7426-7438 doi: 10.1074/jbc.M116.714964

B.W. contributed to the design of the experiments and performed reticulocyte-based pulse-chase experiments.

Park, S. E., Kim, J. M., Seok, O. H., Cho, H., Wadas, B., Kim, S. Y., Varshavsky, A., and Hwang, C. S. (2015) Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. Science 347(6227), 1249-1252 doi: 10.1126/science.aaa3844

B.W. contributed to the design of experiments and writing of the paper.

Wadas, B., Borjigin, J., Oh, J.-H., Hwang, C.S., and Varshavsky, A. (2016) Degradation of AANAT, a Circadian Regulator, by the N-end Rule Pathway. (Submitted)

B.W. designed the experiments, performed the experiments, and wrote the paper.

TABLE OF CONTENTS

TABLE OF FIGURES

CHAPTER 1:

INTRODUCTION TO THE N-END RULE PATHWAY

INTRODUCTION

The Ubiquitin-Proteasome System

The regulated and processive degradation of intracellular proteins is carried out largely by the ubiquitin-proteasome system, in conjunction with molecular chaperones, autophagy, and lysosomal proteolysis. The half-lives of proteins in a cell range from as brief as a few seconds to many days. The functions of protein regulation through proteolysis include the destruction of regulatory proteins whose concentrations must vary with time and changes in the state of a cell, the generation of protein fragments that can act as hormones, antigens, or other effectors, the maintenance of amino acid pools in cells, the elimination of abnormal proteins, and the control of subunit stoichiometries within protein complexes (1-3).

In eukaryotes, proteins are targeted for degradation by conjugation of the 76-residue ubiquitin (Ub) protein to a protein substrate that contains a degradation signal, termed a degron (4). Following ubiquitylation, proteins are targeted to and recognized by the 26S proteasome, a processive, ATP-dependent protease (5). In the canonical mechanism of substrate ubiquitylation, Ub is conjugated to a specific Lys residue in a substrate protein, and then polyubiquitin chains are assembled on the substrate through linkage of the C-terminal Gly^{76} of one Ub to the Lys48 residue of the preceding Ub, thus forming a Lys48-linked polyUb Chain (6). Conjugation of Ub to another protein requires three sequential reactions. In the first step, the C-terminal Gly^{76} residue of Ub is joined via a thioester linkage to a Cys residue of an E1 Ub-activating enzyme, in an ATP-dependent reaction. In the second step, the “activated” Ub is transferred to a Cys residue in an E2 Ub-conjugating enzyme. In the third step, the Ub is transferred by an E3

ubiquitin ligase to (most commonly) a Lys residue in the target substrate, where it is linked to an μ-amino group of the Lys via an isopeptide bond (7) (Fig. 1.1).

A polyubiquitylated protein is targeted to the 26S proteasome, a multisubunit, ATP-dependent protease that recognizes, unfolds, deubiquitylates, and processively degrades the ubiquitylated substrate. The 26S proteasome is comprised of a 20S core particle (CP) and a 19S particle (also known as the regulatory particle (RP)). Recognition of substrates by the proteasome is mediated through binding of the 19S RP to the substrate's poly-Ub chain. After binding of the substrate to the 19S RP, its polyubiquitin linkage is disassembled by deubiquitylase activity, the protein is at least partially unfolded, and it is then translocated into the 20S CP (5,8). The hollow 20S CP chamber contains proteolytic activities that degrade substrates to short peptides, which can then be further degraded to amino acids by cytosolic peptidases, following the release of short peptides from the proteasome (9,10).

In addition to the canonical Lys48-linked polyubiquitin chains, Ub can also be conjugated to a substrate as a mono-Ub moiety, or polyubiquitin chains assembled through the use of other (non-Lys48) Lys residues in Ub. In addition, polyubiquitin chains can also be linked to specific Thr, Ser, or Cys residues in the substrate protein, or through the conjugation of Ub to the N-terminus of a substrate protein (11). Although polyubiquitin chains formed through nearly any of the seven Lys residues in Ub can target proteins for degradation, these (non-Lys-48) substrate-liked polyubiquitin chains also have diverse non-proteolytic functions (11-14).

The N-end Rule Pathway

In 1986, the N-end rule pathway became the first specific pathway of the Ub-proteasome system to be discovered (15,16). The N-end rule pathway relates the identity of a protein's N-terminal residue to its *in vivo* half-life. N-terminal degradation signals of the N-end rule pathway are called N-degrons, and recognition components of the N-end rule pathway are called N-recognins (1). The primary determinant of an N-degron is a destabilizing N-terminal residue of a protein. In eukaryotes, E3 Ub ligases that bind to specific N-degrons function as N-recognins, whereas in prokaryotes the 12-kDa protein ClpS functions as an N-recognin (17). The eukaryotic N-end rule pathway consists of two branches, the Arg/N-end rule pathway (discovered in 1986), and the Ac/N-end rule pathway (discovered in 2010).

The Arg/N-end Rule Pathway

The Arg/N-end rule pathway targets proteins for degradation through the recognition of specific unacetylated N-terminal residues by E3 N-recognins, termed Arg/N-recognins. Unacetylated N-terminal Arg, Lys, His, Leu, Ile, Phe, Trp, Tyr, Asp, Glu, Asn, Gln, and Cys comprise the main determinants of N-degrons in the Arg/N-end rule pathway. These residues become N-terminal through either the cotranslational removal of Met by Met-aminopeptidase (MetAPs), or following the posttranslational and conditional cleavage by nonprocessive proteases, such as, for example, calpains, caspases, and separases. Additionally, it was recently shown that the unacetylated N-terminal Met residue, when followed by a bulky hydrophobic ($) residue, can also act as a primary destabilizing residue (18).

E3 ligases that function as Arg/N-recognins contain highly spalogous (spatially similar (19)) ~80-residue regions called UBR domains or Type-1 binding sites (20-23). A UBR domain binds to the Type-1 primary destabilizing N-terminal residues Arg, Lys, or His. A second, usually adjacent, region of Arg/N-recognins, called the Type-2 binding site, recognizes N-terminal Leu, Ile, Phe, Trp, and Tyr. Together, the directly recognized N-terminal basic (Arg, Lys, His) and bulky hydrophobic (Leu, Ile, Phe, Trp, and Tyr) residues are denoted as "primary" destabilizing residues.

In contrast to the primary destabilizing residues, N-terminal Asp, Asn, Glu, Gln, and Cys function as destabilizing residues following their preliminary modifications, and are thus denoted as "secondary" or "tertiary" destabilizing residues, depending on the number of modifications required before targeting by N-recognins. One of the preliminary modifications is N-terminal arginylation (Nt-arginylation), in which the Arg-tRNA-protein transferase (R-transferase) conjugates Arg to N-terminal Asp, Glu, or oxidized Cys of proteins or peptides. R-transferases, which are apparently confined to eukaryotes, are encoded by the *Ate1* gene and utilize Arg-tRNA as the cosubstrate and donor of Arg (24-27). Although R-transferases cannot directly arginylate N-terminal Asn or Gln, the Arg/N-end rule pathway contains specific N-terminal amidases (Nt-amidases) that enzymatically convert N-terminal Asn and Gln (tertiary destabilizing residues) to Asp and Glu, respectively, which can then be Nt-arginylated (28-32) (Fig. 4.4).

An unmodified N-terminal Cys residue of a protein is yet another tertiary destabilizing residue, as it must undergo two modifications (oxidation and then arginylation) prior to recognition and binding of the substrate by an N-recognin. If the N-terminal Cys can be oxidized, through either enzymatic reactions mediated by N-terminal

Cys oxidases or non-enzymatic reactions that require nitric oxide (NO) and oxygen, the resulting Cys-sulfinate or Cys-sulfonate can then be Nt-arginylated by the Ate1 R-transferase (33,34). The requirement for NO- and oxygen-dependent oxidation of N-terminal Cys-containing substrates makes the Arg/N-end rule pathway a mechanistically distinct sensor of both NO and oxygen (34,35).

Early studies of the *S. cerevisiae* Arg/N-end rule pathway indicated that polyubiquitylation of substrates was mediated by a dimer comprised of the 225 kDa Ubr1 E3 ubiquitin ligase and the Ubr1-bound 20 kDa Rad6 E2 ubiquitin-conjugating enzyme (1,36-38). More recently, the targeting apparatus was discovered to be more elaborate, comprising a physical complex of the RING-type Ubr1 E3 N-recognin and the HECT-type Ufd4 E3, in association with their cognate E2s Rad6 and Ubc4/Ubc5, respectively (39,40). Ufd4 is the E3 of the Ub-fusion degradation (UFD) pathway that recognizes a "non-removable" N-terminal Ub moiety of a Ub fusion as a primary degron, and polyubiquitylates the moiety, leading to the fusion's proteasome-dependent degradation (41,42). The UFD4 pathway is present in both yeast and mammals, suggesting that the double-E3 organization (Ubr1-Ufd4) of the *S. cerevisiae* Arg/N-end rule pathway is universal among eukaryotes (39).

In contrast to *S. cerevisiae,* in which Ubr1 is the sole N-recognin of the Arg/N-end rule pathway, the genomes of multicellular eukaryotes, such as mammals, encode at least four UBR domain-containing E3 N-recognins, termed Ubr1, Ubr2, Ubr4, and Ubr5 (also known as Edd). Mammalian Ubr1 and Ubr2 are highly sequelogous N-recognins, whereas the sequelogy between Ubr1 and Ubr4 or Ubr5 is confined largely to their ~80 residue UBR domains. Mammalian Ubr1, Ubr2, and Ubr4 N-recognins contain both

Type-1 and Type-2 binding sites, whereas Ubr5/Edd N-recognin lacks the Type-2 binding site.

Regulated degradation of proteins or their fragments by the N-end rule pathway mediates a strikingly broad range of functions, including the sensing of heme, NO, oxygen, and short peptides; the selective elimination of misfolded proteins; the regulation of DNA repair through the degradation of the Mgt1 DNA repair protein; the cohesion/segregation of chromosomes through the degradation of a subunit of cohesin; signaling by transmembrane receptors through the degradation of the G-protein regulators Rgs2, Rgs4, Rgs5, and Rgs16; the control of peptide import through the degradation of Cup9; the regulation of apoptosis, meiosis, viral and bacterial infections, fat metabolism, cell migration, actin filaments, cardiovascular development, spermatogenesis, neurogenesis, and memory; the function of adult organs, including the brain, muscle, testis, and pancreas; and the regulation of leaf and shoot development, leaf senescence, and seed germination in plants (3,35,43-60).

In Chapter 2 of this thesis we further expand the functions of the Arg/N-end rule pathway to include the degradation of the AANAT enzyme, an important regulator of circadian physiology, and in Chapter 5 we expand these functions to include the degradation of the separase-cleaved mouse REC8, a subunit of cohesin complexes during meiosis. Mutational inactivation of human UBR1 causes Johanson-Blizzard syndrome, a set of severe birth defects (JBS). Ubr1-lacking mice exhibit phenotypes that partially overlap with those of human JBS patients (52). Ubr2-lacking mice have a variety of defects, including genomic instability and impaired spermatogenesis (as well as infertility) in males (61,62).

The Ac/N-end Rule Pathway

The second branch of the N-end rule pathway, termed the Ac/N-end rule pathway, was discovered by our laboratory in 2010, 24 years after discovery of the Arg/N-end rule pathway (63). This branch involves the cotranslational $N^{\pm}$-acetylation (Nt-acetylation) of nascent proteins whose N-termini bear either Met or the small, uncharged residues Ala, Cys, Ser, Thr, or Val. These residues become N-terminal after the cotranslational removal of the N-terminal Met residue by Met-aminopeptidases (MetAPs), and can be cotranslationally Nt-acetylated by specific ribosome Nt-acetylases, thereby forming degradation signals that can be recognized by the Ac/N-end rule pathway (64-68).

The Nt-acetylated Met, Ala, Cys, Ser, Thr, and Val residues comprise a class of N-degrons termed Ac/N-degrons (63). The cotranslational Nt-acetylation of nascent proteins is both mechanistically and functionally distinct from the largely posttranslational acetylation of internal Lys residues in many proteins (65,69-79). Furthermore, in contrast to acetylation of internal lysines, Nt-acetylation is apparently irreversible (there are, apparently, no N-terminus-specific deacetylases). Previous studies of Nt-acetylation have characterized the specificity and activity of Nt-acetyltransferases that catalyze this enzymatic reaction, but the biological function of Nt-acetylation remained largely unknown. The discovery of the Ac/N-end rule pathway revealed a major physiological role of both Nt-acetylases and MetAPs (2,63).

In *S. cerevisiae*, the Ac/N-end rule pathway is mediated by, at least, the RING-type Doa10 and Not4 E3 ligases. Doa10, an endoplasmic reticulum (ER)-embedded E3, functions together with the Ubc6/Ubc7 E2 enzymes, and is capable of targeting both

"soluble" (nuclear and cytosolic) and transmembrane proteins (80). In mammalian cells, this pathway is mediated by, at least, Teb4 (see Chapter 3), which is sequelogous to the yeast Doa10 E3 Ub ligase (81).

Together, the Arg/N-end rule pathway and the Ac/N-end rule pathway are capable of targeting a majority of cellular proteins for regulated degradation. More than 80% of human proteins are cotranslationally Nt-acetylated by ribosome-associated Nt-acetylases. Consequently, most proteins are likely to harbor an Ac/N-degron from the moment of their birth. (This statement is still inferential, awaiting more extensive, proteome-scale analyses that would determine how many of Nt-acetylated proteins actually bear an at least transiently active Ac/N-degron. The current assumption, based on already existing lines of evidence (refs. of 2010-2015), is that many, possibly most, Nt-acetylated proteins contain Ac/N-degrons.)

Interestingly, the Arg/N-end rule and Ac/N-end rule pathways are not mutually exclusive for many substrates, as many proteins that can be Nt-acetylated are only partially Nt-acetylated *in vivo*, even under normal (unstressful) conditions. As a result (and for a subset of N-terminal residues), a partial Nt-acetylation of a protein can still allow its complete destruction through its complementary targeting by both Arg/N-recognins and Ac/N-recognins (Fig 1.5).

One example of an N-terminal residue susceptible to "complementary" targeting by both branches of the N-end rule is N-terminal Cys, which can be Nt-acetylated (in proteins that contain the initially present N-terminal Met-Cys sequence) after the cotranslational removal of N-terminal Met by Met-aminopeptidases. However, some sequence contexts, for example, a basic residue at position 2 (following the removal of

Nt-Met), inhibit the Nt-acetylation of N-terminal Cys. The mammalian proteins Rgs4, Rgs5, and Rgs16 are one example of this inhibition. In each of these proteins, the N-terminal Cys residue is followed by a basic residue, which inhibits Nt-acetylation and thereby facilitates for Ate1 R-transferase-dependent Nt-arginylation of these proteins, following the conditional oxidation of their N-terminal Cys), and subsequent recognition of the N-terminal Arg residue by the Ubr1 E3 ligase.

Despite the current understanding of the sequence specificity of the Ate1 R-transferase (Fig. 1.2), a more complete understanding remained to be attained. For example, it remained unclear exactly how the Nt-arginylation of activity of Ate1 is influenced by specific sequence and conformational features of proteins bearing N-terminal Asp, Gly, and Cys residues, the "canonical" sequence determinants recognized by Ate1. Moreover, over the last decade there were several, still extant, lines of (non-definitive) evidence in the published literature that claimed the ability of Ate1 to Nt-arginylate "non-canonical" Nt-sequences as well. In Chapter 4, we utilize an *in vitro* arginylation assay and Celluspots peptide arrays, in combination with rabbit reticulocyte-based pulse-chase degradation assays, to systematically investigate the effects of sequence context on the Nt-arginylation of peptides containing N-terminal Asp, Glu, and Cys.

A second example of N-terminal residues that are susceptible to complementary targeting by both branches of the N-end rule pathway is N-terminal Met when it is followed by a hydrophobic residue ($). As mentioned above, the N-terminal Met residue is permissive for Nt-acetylation, thus making it possible for proteins that retain N-terminal Met and become Nt-acetylated to be targeted for degradation by the Ac/N-end

rule pathway. However, in cases where Nt-acetylation of N-terminal Met is incomplete, the unmodified N-terminal Met can be recognized by the Ubr1 Arg/N-recognin of the Arg/N-end rule pathway, if Nt-Met if followed by a bulky hydrophobic residue (18).

In Chapter 2 we show that the rat serotonin N-acetyltransferase (AANAT), an important mediator of circadian physiology, is targeted for degradation via complementary recognition of its partially-acetylated N-terminal M $ (Met-Leu-) sequence by both branches of the N-end rule pathway. In Chapter 3, we demonstrate that a clinically-relevant Q2L mutant of human RGS2 (termed ML-RGS2) is also co-targeted for degradation by both the Arg/N-end rule and Ac/N-end rule pathways, owing to its incomplete Nt-acetylation (55,82). Together, these studies advance our knowledge of how both branches of the N-end rule pathway can work together to ensure efficient overall degradation of at least a subset of incompletely Nt-acetylated protein substrates.

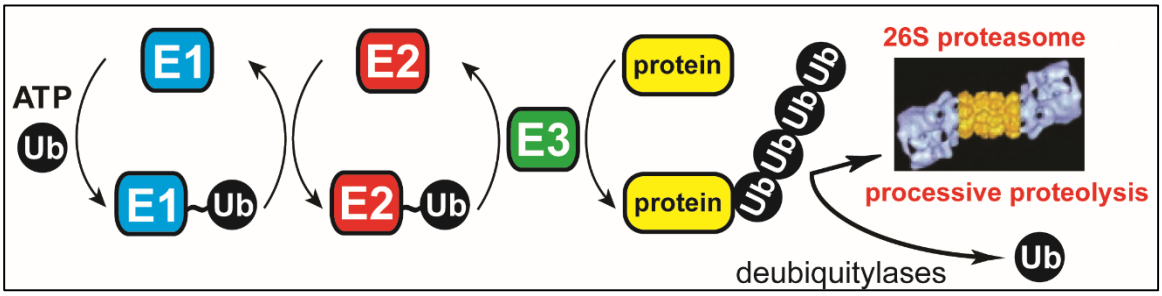


Fig. 1.1. The ubiquitin-proteasome system (UPS). The ubiquitin-proteasome system (Ub system). The conjugation of Ub to other proteins involves a preliminary ATP-dependent step in which the last residue of Ub (Gly^{76}) is joined, via a thioester bond, to a Cys residue of the E1 (Ub-activating) enzyme. The "activated" Ub moiety is transferred to a Cys residue in one of several Ub-conjugating (E2) enzymes, and from there, through an isopeptide bond, to a Lys residue of an ultimate acceptor, denoted as "protein". E2 enzymes function as subunits of E2-E3 Ub ligase complexes that can produce substrate-linked poly-Ub chains. Such chains have specific Ub-Ub topologies, depending on the identity of a Lys residue of Ub (which contains several lysines) that forms an isopeptide bond with C-terminal Gly^{76} of the adjacent Ub moiety in a chain. Specific poly-Ub chains can confer the degradation of a substrate by the 26S proteasome or other metabolic fates. Monoubiquitylation of some protein substrates can also occur, and has specific functions. One role of E3 is the recognition of a substrate's degradation signal (degron). Individual mammalian genomes encode at least 1,000 distinct E3 Ub ligases.

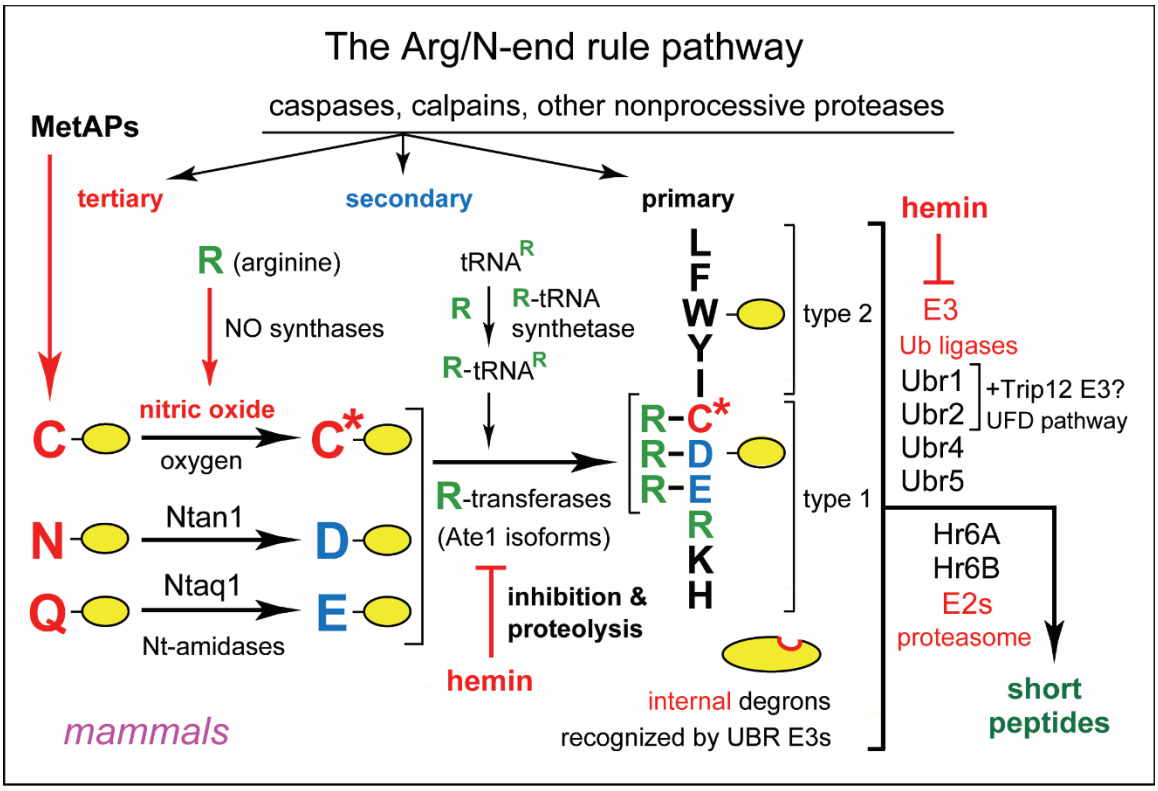


Fig. 1.2. The mammalian Arg/N-end rule pathway. N-terminal residues are indicated by single-letter abbreviations for amino acids. Yellow ovals denote the rest of a protein substrate. “Primary”, “secondary”, and “tertiary” denote mechanistically distinct subsets of destabilizing N-terminal residues. C* denotes oxidized N-terminal Cys, either Cys-sulfinate or Cys-sulfonate, produced *in vivo* through reactions that require both nitric oxide (NO) and oxygen. The mammalian *N*-recognins Ubr1, Ubr2, Ubr4, and Ubr5 (Edd) have multiple substrate binding sites that also recognize internal (non-N-terminal) degrons in other substrates of the Arg/N-end rule pathway, the ones that lack *N*-degrons. A question mark after Trip12 (which mediates the mammalian UFD pathway) and is a sequelog of the *S. cerevisiae* Ufd4 E3) denotes the untested possibility that mammalian Ubr1 and/or Ubr2 form complexes with Trip12, by analogy with the Ubr1–Ufd4 complex in *S. cerevisiae.*

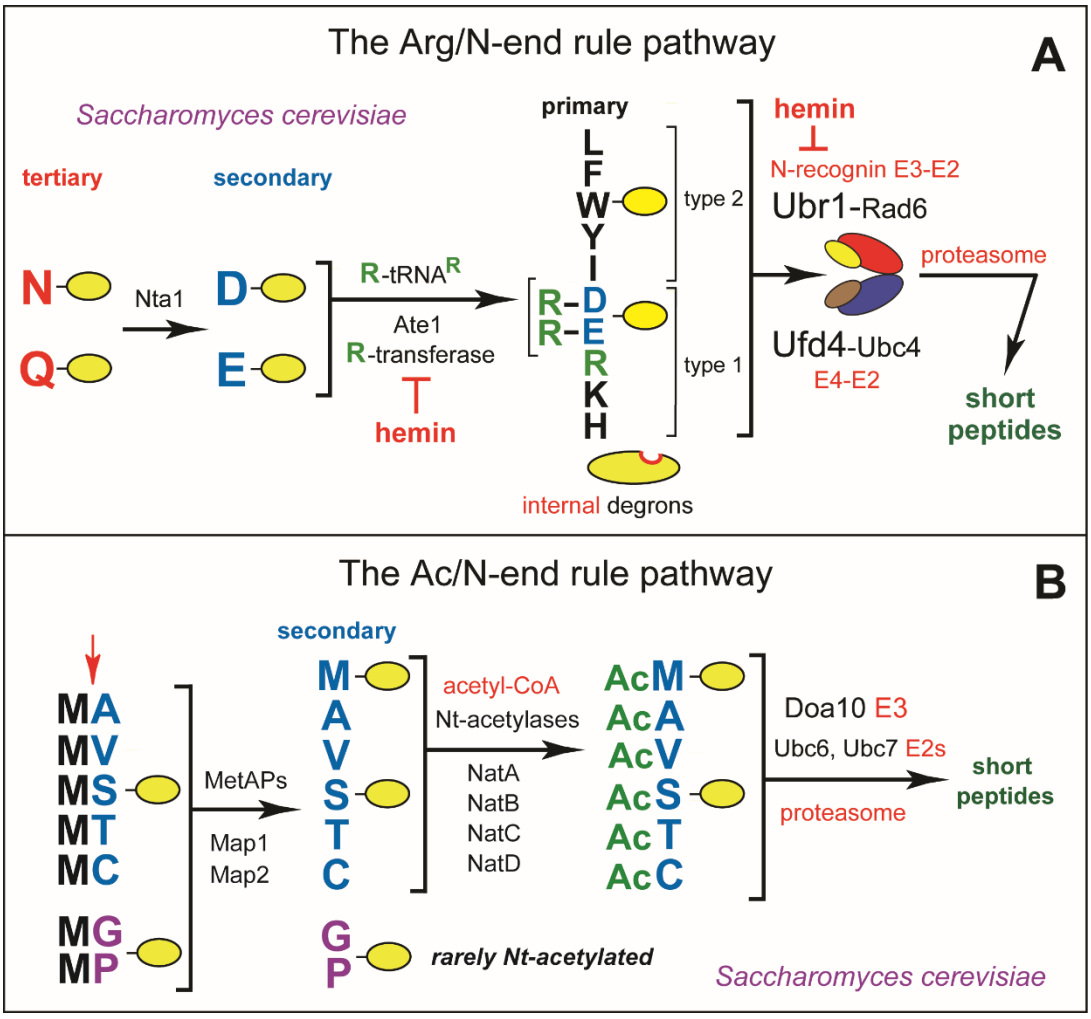


Fig. 1.3. The N-end rule pathway in *Saccharomyces cerevisiae*. A. The Arg/N-end rule pathway. Yellow ovals denote the rest of a protein substrate. "Primary", "secondary" and "tertiary" denote mechanistically distinct subsets of destabilizing N-terminal residues. The physically associated Ubr1 (*N*-recognin) and Ufd4 E3s have substrate-binding sites that recognize internal (non-N-terminal) degrons in substrates of the Arg/N-end rule pathway that lack *N*-degrons. Ubr1 (but not Ufd4) recognizes *N*-degrons as well. B. The Ac/N-end rule pathway. Red arrow on the left indicates the removal of N-terminal Met by Met-aminopeptidases (MetAPs). This Met residue is retained if a residue at Position 2 is nonpermissive (too large) for Met-aminopeptidases. If the (retained) N-terminal Met or N-terminal Ala, Val, Ser, Thr and Cys are followed by residues that allow Nt-acetylation (see the main text), these N-terminal residues are usually Nt-acetylated. The resulting *N*-degrons are called ^{Ac}N-degrons. The term "secondary" refers to the necessity of modification (Nt-acetylation) of a destabilizing N-terminal residue before a protein can be recognized by a cognate Ub ligase. Proteins containing ^{Ac}N-degrons are targeted for ubiquitylation and proteasome-mediated degradation by the Doa10 E3 *N*-recognin, in conjunction with the Ubc6 and Ubc7 E2 enzymes. Although Gly and Pro can be made N-terminal by MetAPs, and although Doa10 can recognize Nt-acetylated Gly and Pro, few proteins with N-terminal Gly or Pro are Nt-acetylated.

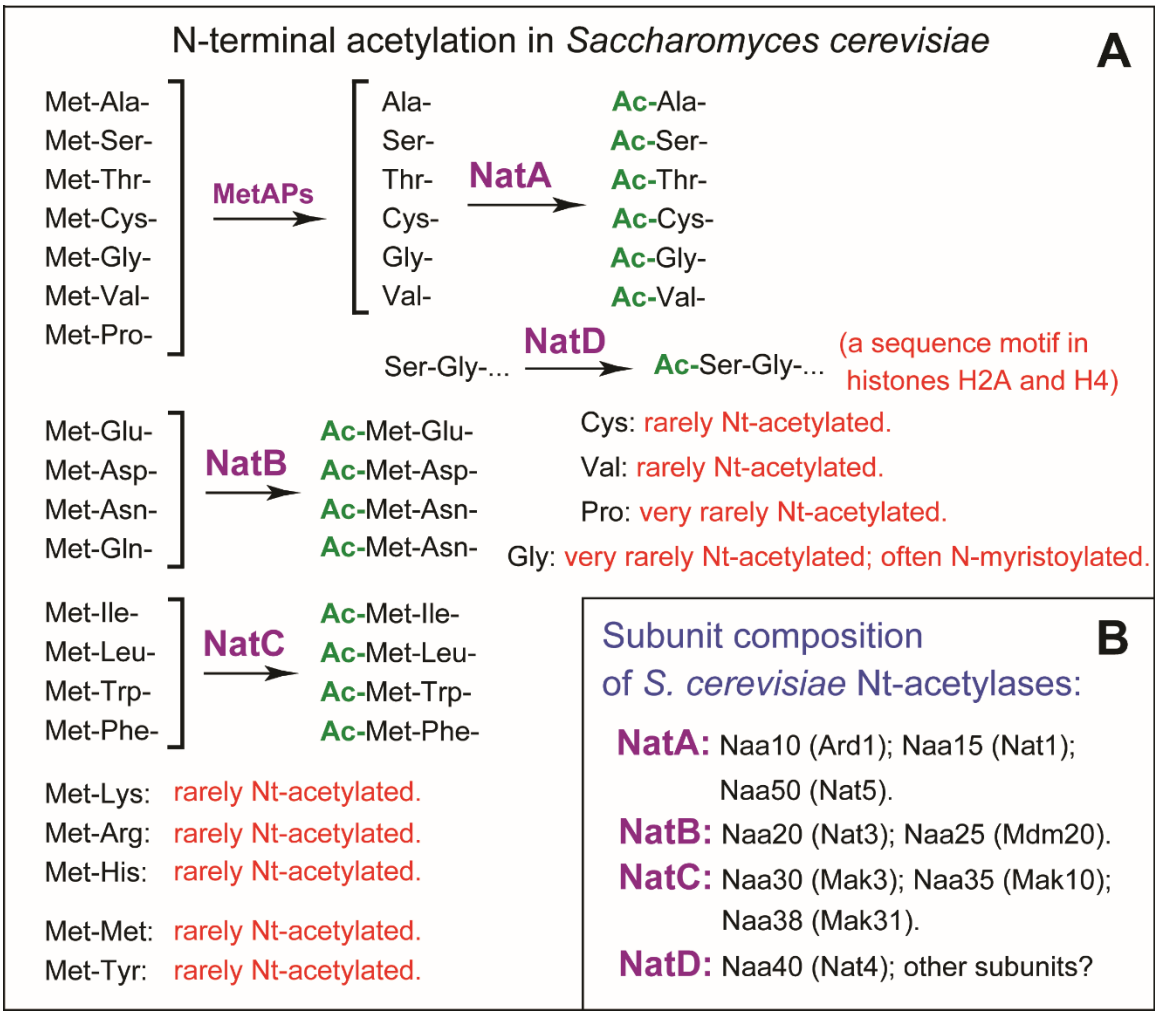


Fig. 1.4. Specificities and subunit compositions of N± -terminal acetylases (Nt-acetylases). A. Substrate specificities of S. cerevisiae Nt-acetylases. •Ac• denotes the N± -terminal acetyl moiety. The bulk of Nt-acetylases are associated with ribosomes. The specificities of mammalian Nt-acetylases are similar to those of their yeast counterparts, but an individual mammalian genome encodes more than ten Nt-acetylases, in contrast to four in S. cerevisiae. Some mammalian Nt-acetylases, such as NatF (its catalytic subunit is called Naa60, can Nt-acetylate N-terminal motifs that include Met-Lys or Met-Arg, which are rarely if ever Nt-acetylated in S. cerevisiae. This compilation of Nt-acetylases and their specificities is derived from data in the literature. (B) Subunits of S. cerevisiae Nt-acetylases.

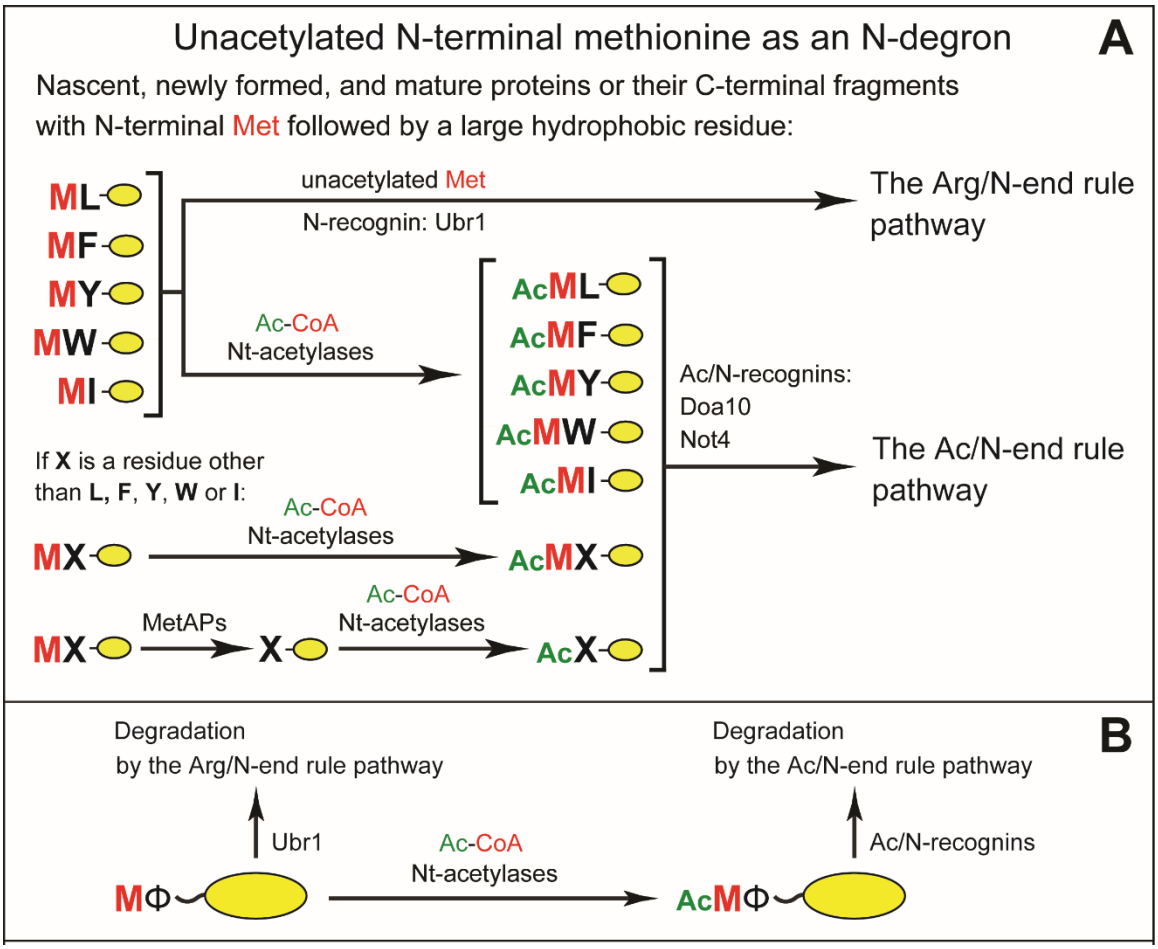


Fig. 1.5. Functional complementarity between the Arg/N-end rule and Ac/N-end rule pathways. This complementarity stems from the recognition of unacetylated Met¦ /N-degrons in Met-¦ proteins vs. the recognition of the Nt-acetylated AcMet¦ /N-degrons in AcMet-¦ proteins. AcMet¦ /N-degrons are a subset of Ac/N-degrons in Nt-acetylated cellular proteins. Met-¦ proteins are defined, in this study, as those that bear N-terminal Met followed by a large hydrophobic (¦) non-Met residue.

REFERENCES

1. Varshavsky, A. (1996) The N-end rule: functions, mysteries, uses. Proc. Natl. Acad. Sci. USA 93, 12142-12149
2. Shemorry, A., Hwang, C. S., and Varshavsky, A. (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50, 540-551
3. Varshavsky, A. (2011) The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20, 1298-1345
4. Varshavsky, A. (1991) Naming a targeting signal. Cell 64, 13-15
5. Finley, D. (2009) Recognition and processing of ubiquitin-protein conjugates by the proteasome. Annu. Rev. Biochem. 78, 477-513
6. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243, 1576-1583
7. Hershko, A., and Ciechanover, A. (1998) The ubiquitin system. Annu. Rev. Biochem. 76, 425-479
8. Finley, D., Ulrich, H. D., Sommer, T., and Kaiser, P. (2012) The ubiquitin-proteasome system of Saccharomyces cerevisiae. Genetics 192, 319-360
9. Lecker, S. H., Goldberg, A. L., and Mitch, W. E. (2006) Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. J Am Soc Nephrol 17, 1807-1819
10. Kisselev, A. F., Akopian, T. N., Woo, K. M., and Goldberg, A. L. (1999) The sizes of peptides generated from protein by mammalian 26 and 20 S proteasomes.

Implications for understanding the degradative mechanism and antigen presentation. J Biol Chem 274, 3363-3371

11. Kravtsova-Ivantsiv, Y., and Ciechanover, A. (2012) Non-canonical ubiquitin-based signals for proteasomal degradation. J Cell Sci 125, 539-548
12. Chen, Z. J., and Sun, L. J. (2009) Nonproteolytic functions of ubiquitin in cell signaling. Mol. Cell 33, 275-286
13. Behrends, C., and Harper, J. W. (2011) Constructing and decoding unconventional ubiquitin chains. Nat Struct Mol Biol 18, 520-528
14. Kulathu, Y., and Komander, D. (2012) Atypical ubiquitylation - the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. Nat Rev Mol Cell Biol 13, 508-523
15. Bachmair, A., Finley, D., and Varshavsky, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234, 179-186
16. Bachmair, A., and Varshavsky, A. (1989) The degradation signal in a short-lived protein. Cell 56, 1019-1032
17. Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A., and Bukau, B. (2006) ClpS is an essential component of the N-end rule pathway in Escherichia coli. Nature 439, 753-756
18. Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A., and Hwang, C. S. (2014) The N-terminal methionine of cellular proteins as a degradation signal. Cell 156, 158-169
19. Varshavsky, A. (2004) Spalog and sequelog: neutral terms for spatial and sequence similarity. Curr. Biol. 14, R181-R183

20. Choi, W. S., Jeong, B.-C., Joo, Y. J., Lee, M.-R., Kim, J., Eck, M. J., and Song, H. K. (2010) Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat. Struct. Mol. Biol. 17, 1175-1181

21. Matta-Camacho, E., Kozlov, G., Li, F. F., and Gehring, K. (2010) Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat. Struct. Mol. Biol. 17, 1182-1188

22. Sriram, S. M., and Kwon, Y. T. (2010) The structural basis of N-end rule recognition. Nat. Struct. Mol. Biol. 17, 1164-1165

23. Sriram, S. M., Kim, B. Y., and Kwon, Y. T. (2011) The N-end rule pathway: emerging functions and molecular principles of substrate recognition. Nat. Rev. Mol. Cell Biol. 12, 735-747

24. Hu, R.-G., Brower, C. S., Wang, H., Davydov, I. V., Sheng, J., Zhou, J., Kwon, Y. T., and Varshavsky, A. (2006) Arginyl-transferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem. 281, 32559-32573

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

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

27. Wang, J., Han, X., Saha, S., Xu, T., Rai, R., Zhang, F., Wolf, Y. I., Wolfson, A., Yates, J. R., and Kashina, A. (2011) Arginyltransferase is an ATP-independent self-regulating enzyme that forms distinct functional complexes in vivo. Chem. Biol. 18, 121-130

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

29. Wang, H., Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell 34, 686-695

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

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

32. Cantor, J. R., Stone, E. M., and Georgiou, G. (2011) Expression and biochemical characterization of the human enzyme N-terminal asparagine amidohydrolase. Biochemistry 50, 3025-3033

33. Weits, D. A., Giuntoli, B., Kosmacz, M., Parlanti, S., Hubberten, H. M., Riegler, H., Hoefgen, R., Perata, P., van Dongen, J. T., and Licausi, F. (2014) Plant

cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nat Commun. 5, 3425

34. Hu, R.-G., Sheng, J., Xin, Q., Xu, Z., Takahashi, T. T., and Varshavsky, A. (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437, 981-986

35. Lee, M. J., Tasaki, T., Moroi, K., An, J. Y., Kimura, S., Davydov, I. V., and Kwon, Y. T. (2005) RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 102, 15030-15035

36. Varshavsky, A. (2008) Discovery of cellular regulation by protein degradation ("Reflections" article). J. Biol. Chem. 283, 34469-34489

37. Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M., and Varshavsky, A. (2008) Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. J. Biol. Chem. 283, 24011-24028

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

39. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 12, 1177-1185

40. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2009) Two proteolytic pathways regulate DNA repair by cotargeting the Mgt1 alkylguanine transferase. Proc. Natl. Acad. Sci. USA 106, 2142-2147

41. Xie, Y., and Varshavsky, A. (2002) UFD4 lacking the proteasome-binding region catalyses ubiquitination but is impaired in proteolysis. Nature Cell Biol. 4, 1003-1007

42. Ju, D., Wang, X., Xu, H., and Xie, Y. (2007) The armadillo repeats of the Ufd4 ubiquitin ligase recognize ubiquitin-fusion proteins. FEBS Lett. 581, 265-270

43. Hwang, C. S., and Varshavsky, A. (2008) Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 105, 19188-19193

44. Hwang, C.-S., Shemorry, A., and Varshavsky, A. (2009) Two proteolytic pathways regulate DNA repair by co-targeting the Mgt1 alkylguanine transferase. Proc. Natl. Acad. Sci. USA 106, 2142-2147

45. Kwon, Y. T., Kashina, A. S., Davydov, I. V., Hu, R.-G., An, J. Y., Seo, J. W., Du, F., and Varshavsky, A. (2002) An essential role of N-terminal arginylation in cardiovascular development. Science 297, 96-99

46. Liu, Y. J., Liu, C., Chang, Z., Wadas, B., Brower, C. S., Song, Z. H., Xu, Z. L., Shang, Y. L., Liu, W. X., Wang, L. N., Dong, W., Varshavsky, A., Hu, R. G., and Li, W. (2016) Degradation of the separase-cleaved Rec8, a meiotic cohesin subunit, by the N-end rule pathway. J. Biol. Chem. (in press)

47. Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2012) The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. Proc. Natl. Acad. Sci. USA 109, E1839-E1847

48. Piatkov, K. I., Colnaghi, L., Bekes, M., Varshavsky, A., and Huang, T. T. (2012) The auto-generated fragment of the Usp1 deubiquitylase is a physiological substrate of the N-end rule pathway. Mol. Cell 48, 926-933

49. Piatkov, K. I., Oh, J.-H., Liu, Y., and Varshavsky, A. (2014) Calpain-generated natural protein pragments as short-lived substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 111, E817-E826

50. Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001) Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature 410, 955-960

51. Xia, Z., Turner, G. C., Hwang, C.-S., Byrd, C., and Varshavsky, A. (2008) Amino acids induce peptide uptake via accelerated degradation of CUP9, the transcriptional repressor of the PTR2 peptide transporter. J. Biol. Chem. 283, 28958-28968

52. Zenker, M., Mayerle, J., Lerch, M. M., Tagariello, A., Zerres, K., Durie, P. R., Beier, M., Hülskamp, G., Guzman, C., Rehder, H., Beemer, F. A., Hamel, B., Vanlieferinghen, P., Gershoni-Baruch, R., Vieira, M. W., Dumic, M., Auslender, R., Gil-da-Silva-Lopes, V. L., Steinlicht, S., Rauh, R., Shalev, S. A., Thiel, C., Winterpacht, A., Kwon, Y. T., Varshavsky, A., and Reis, A. (2005) Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). Nat. Genet. 37, 1345-1350

53. An, J. Y., Seo, J. W., Tasaki, T., Lee, M. J., Varshavsky, A., and Kwon, Y. T. (2006) 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 103, 6212-6217

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

55. Park, S. E., Kim, J. M., Seok, O. H., Cho, H., Wadas, B., Kim, S. Y., Varshavsky, A., and Hwang, C. S. (2015) Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. Science 347, 1249-1252

56. Bodenstein, J., Sunahara, R. K., and Neubig, R. R. (2007) N-terminal residues control proteasomal degradation of Rgs2, Rgs4, and Rgs5 in human embryonic kidney 293 cells. Mol. Pharmacol. 71, 1040-1050

57. Hu, R.-G., Wang, H., Xia, Z., and Varshavsky, A. (2008) The N-end rule pathway is a sensor of heme. Proc. Natl. Acad. Sci. USA 105, 76-81

58. Gibbs, D. J., Lee, S. C., Isa, N. M., Gramuglia, S., Fukao, T., Bassel, G. W., Correia, C. S., Corbineau, F., Theodoulou, F. L., Bailey-Serres, J., and Holdsworth, M. J. (2011) Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature 479, 415-418

59. Graciet, E., and Wellmer, F. (2010) The plant N-end rule pathway: structure and functions. Trends Plant Sci. 15, 447-453

60. Licausi, F., Kosmacz, M., Weits, D. A., Giuntoli, B., Giorgi, F. M., Voesenek, L. A. C. J., Perata, P., and van Dongen, J. T. (2011) Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature 479, 419-422

61. Kwon, Y. T., Xia, Z. X., An, J. Y., Tasaki, T., Davydov, I. V., Seo, J. W., Xie, Y., and Varshavsky, A. (2003) Female lethality and apoptosis of spermatocytes in mice lacking the UBR2 ubiquitin ligase of the N-end rule pathway. Mol. Cell. Biol. 23, 8255-8271

62. Ouyang, Y., Kwon, Y. T., An, J. Y., Eller, D., Tsai, S.-C., Diaz-Perez, S., Trke, J., Teitell, M. A., and Marahrens, Y. (2006) Loss of Ubr2, an E3 ubiquitin ligase, leads to chromosome fragility and impaired homologous recombinational repair. Mut. Res. 596, 64-75

63. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327, 973-977

64. Addlagatta, A., Hu, X., Liu, J. O., and Matthews, B. W. (2005) Structural basis for the functional differences between Type I and Type II human methionine aminopeptidases. Biochemistry 44, 14741-14749

65. Moerschell, R. P., Hosokawa, Y., Tsunasawa, S., and Sherman, F. (1990) The specificities of yeast methionine aminopeptidase and acetylation of amino-terminal methionine in vivo. Processing of altered iso-1-cytochromes created by oligonucleotide transformation. J. Biol. Chem. 265, 19638-19643

66. Li, X., and Chang, Y.-H. (1995) Amino-terminal protein processing in Saccharomyces cerevisiae is an essential function that requires two distinct methionine aminopeptidases. Proc. Natl. Acad. Sci. USA 92, 12357-12361

67. Bradshaw, R. A., Brickey, W. W., and Walker, K. W. (1998) N-terminal processing: the methionine aminopeptidase and N-alpha-acetyl transferase families. Trends Biochem. Sci. 23, 263-267

68. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage. Mol. Cell. Proteomics 5, 2336-2349

69. Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325, 834-840

70. Mullen, J. R., Kayne, P. S., Moerschell, R. P., Tsunasawa, S., Gribskov, M., Colavito-Shepanski, M., Grunstein, M., Sherman, F., and Sternglanz, R. (1989) Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast. EMBO J. 8, 2067-2075

71. Park, E. C., and Szostak, J. W. (1992) ARD1 and NAT1 proteins form a complex that has N-terminal acetyltransferase activity. EMBO J. 11, 2087-2093

72. Arnesen, T. (2011) Toward a functional understanding of protein N-terminal acetylation. PLoS Biol. 9, e1001074

73. Arnesen, T., Van Damme, P., Polevoda, B., Helsens, K., Evjenth, R., Colaert, N., Varhaug, J. E., Vandekerckhove, J., Lillehaug, J. R., Sherman, F., and Gevaert, K. (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast to humans. Proc. Natl. Acad. Sci. USA 106, 8157-8162

74. Starheim, K. K., Gromyko, D., Velde, R., Varhaug, J. E., and Arnesen, T. (2009) Composition and biological significance of the human N-alpha terminal acetyltransferases. BMC Proceedings 3 (Suppl 6), S3

75. Van Damme, P., Evjenth, R., Foyn, H., Demeyer, K., De Bock, P. J., Lillehaug, J. R., Vandekerckhove, J., Arnesen, T., and Gevaert, K. (2011) Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. Mol. Cell. Proteomics 10, M110.004580

76. Van Damme, P., Lasac, M., Polevoda, B., Gazquezc, C., Elosegui-Artolae, A., Kim, D. S., De Juan-Pardoe, E., Demeyera, K., Holef, K., Larreac, E., Timmermana, E., Prietoc, J., Arnesenf, T., Sherman, F., Gevaert, K., and Aldabec, R. (2012) N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proc. Natl. Acad. Sci. USA 109, 12449-12454

77. Polevoda, B., Brown, S., Cardillo, T. S., Rigby, S., and Sherman, F. (2008) Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. J. Cell. Biochem. 103, 492-508

78. Polevoda, B., Norbeck, J., Takakura, H., Blomberg, A., and Sherman, F. (1999) Identification and specificities of N-terminal acetyltransferases from Saccharomyces cerevisiae. EMBO J. 18, 6155-6168

79. Polevoda, B., and Sherman, F. (2003) N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. 325, 595-622

80. Ravid, T., Kreft, S. G., and Hochstrasser, M. (2006) Membrane and soluble substrates of the Doa10 ubiquitin ligase are degraded by distinct pathways. EMBO J. 25, 533-543

81. Kreft, S. G., Wang, L. H.-C., and Hochstrasser, M. (2006) Membrane topology of the yeast endoplasmic reticulum-localized ubiquitin ligase Doa10 and comparison with its human ortholog TEB4 (MARCH-VI). J. Biol. Chem. 281, 4646-4653

82. Yang, J., Kamide, K., Kokubo, Y., Takiuchi, S., Tanaka, C., Banno, M., Miwa, Y., Yoshii, M., Horio, T., Okayama, A., Tomoike, H., Kawano, Y., and Miyata, T. (2005) Genetic variations of regulator of G-protein signaling 2 in hypertensive patients and in the general population. J. Hypertens. 8, 1497-1505

CHAPTER 2:

DEGRADATION OF AANAT, A CIRCADIAN REGULATOR, BY THE N-END RULE PATHWAY

From Wadas, B., Borjigin, J., Oh, J.-H., Hwang, C.S., and Varshavsky, A. (2016)

submitted

INTRODUCTION

Arylalkylamine N-acetyltransferase (AANAT) converts the neurotransmitter serotonin to *N*-acetylserotonin (NAS) (1-11). Regulatory functions of NAS include its activity as an agonist of TrkB, the receptor for the brain-derived neurotrophic factor (BDNF) (12-14). NAS is the immediate precursor of melatonin, a circulating hormone that regulates sleep and other circadian processes in vertebrates. Melatonin is also present in invertebrates, including insects, as well as in animals that lack recognizable neurons (1,15-21). The functions of melatonin in mammals include the modulation of circadian rhythms in response to light-dark cycles. Melatonin also contributes to the control of seasonal physiology (20). These mechanisms are a part of a broader range of processes that involve distinct biological oscillators in all organisms (22-24). In the course of daily light-dark cycles, the activity of AANAT and the level of the ~23-kDa AANAT protein in the brain's pineal gland are high during nocturnal periods and rapidly decrease following exposure to light (1,20,25).

In mammals, changes of AANAT levels in the brain's pineal gland are controlled by the circadian oscillator in the suprachiasmatic nucleus (SCN) of the hypothalamus. At night, axons of rodent superior cervical ganglion neurons release norepinephrine (NE) in the pineal gland, in response to circadian signals from the SCN, thereby activating adrenergic receptors to increase intracellular Ca^{2+} and cAMP. The resulting phosphorylation of CREB, the cAMP response element (CRE)-binding protein, upregulates the *Aanat* transcriptional promoter, which contains CRE sequence elements (1,4,25-27). Increases in cAMP also stimulate site-specific phosphorylation of AANAT by the protein kinase A (PKA), thereby causing the binding of AANAT to 14-3-3

proteins. These transitions augment the enzymatic activity of AANAT and inhibit its degradation (1,6,7,28). While AANAT is predominantly expressed in the pineal gland and the retina, this enzyme (and, consequently, the AANAT-dependent synthesis of NAS and melatonin) is also present in cells of the gastrointestinal tract, and apparently at other sites as well, including the hippocampus, the olfactory bulb, the cerebellum, and the spinal cord (12,29,30).

The activity of AANAT determines the rate of NAS synthesis. Melatonin is produced from NAS by hydroxyindole-*O*-methyltransferase (HIOMT). The levels of NAS, synthesized by AANAT, can constrain the rate of melatonin synthesis from NAS during light parts of daily cycles (when NAS levels are low) but appear to be not the rate-limiting step during dark parts of these cycles, when both AANAT and NAS levels are high (21). In rodents such as rats the activity of AANAT starts to increase 3 to 4 h after the onset of darkness and begins to decrease before the onset of light in the morning (21,31). Therefore the importance of regulating the expression and enzymatic activity of AANAT, as these parameters determine the levels of both NAS and melatonin.

Earlier studies have shown that the concentration of AANAT, a largely cytosolic enzyme (32), is controlled through both transcriptional and posttranslational mechanisms, and that vertebrate species can differ in their mode of AANAT regulation (1,11,31,33-35). In rodents, the onset of darkness leads to an increase in melatonin (requiring a preceding increase of NAS) after a lag period, whereas in sheep and primates increases in melatonin occur rapidly upon the onset of darkness. The levels of *Aanat* mRNA in rodents can vary by >100-fold during light-dark cycles (10), whereas in non-rodent mammals, such as sheep for example, these differences can be as low as ~2-fold,

suggesting a major involvement of translational and/or posttranslational AANAT regulation in latter cases (11,36). In addition to variability in regulation at the level of AANAT-encoding mRNAs, AANAT is also unusual in the extent of variability of its N-terminal amino acid sequences during vertebrate evolution (Fig. 1C). In 2010, one of our laboratories showed that N-terminal residues of rat AANAT play a role in the control of its degradation, suggesting that at least rodent AANATs may be targeted for degradation by the N-end rule pathway (37).

The N-end rule pathway is a set of cytosolic/nuclear proteolytic systems whose unifying feature is the ability to recognize and polyubiquitylate proteins containing N-terminal degradation signals (degrons) called N-degrons, thereby causing degradation of these proteins by the proteasome (Fig. 1A, B) (38-46). Ubiquitin (Ub) ligases of the N-end rule pathway, called N-recognins, can recognize (bind to) not only N-degrons but also specific internal (non-N-terminal) degradation signals (47-50). The main determinant of an N-degron is either an unmodified or chemically modified destabilizing N-terminal residue of a protein. The identity of the next residue, at position 2, is often important as well. Another determinant of an N-degron is a protein's internal Lys residue(s). It functions as the site of a protein's polyubiquitylation and tends to be located in a conformationally disordered region (41,51,52). Bacteria also contain the N-end rule pathway, but Ub-independent versions of it (53-58).

Regulated degradation of proteins and their natural fragments by the N-end rule pathway has been shown to mediate a strikingly broad range of biological functions, including the sensing of heme, nitric oxide (NO), oxygen, and short peptides; the control, through subunit-selective degradation, of the input stoichiometries of subunits in

oligomeric protein complexes; the elimination of misfolded or otherwise abnormal proteins; the degradation of specific proteins after their retrotranslocation to the cytosol from mitochondria or other membrane-enclosed compartments; the regulation of apoptosis and repression of neurodegeneration; the regulation of DNA repair, transcription, replication, and chromosome cohesion/segregation; the regulation of G proteins, cytoskeletal proteins such as actin and myosin, autophagy, peptide import, meiosis, immunity, fat metabolism, cell migration, cardiovascular development, spermatogenesis, and neurogenesis; the functioning of adult organs, including the brain, muscle, testis, and pancreas; and the regulation of leaf and shoot development, leaf senescence, and many other processes in plants (Figs. 1A, B) ((41-46) and references therein).

In eukaryotes, the N-end rule pathway consists of two branches. One branch, called the Ac/N-end rule pathway, targets proteins for degradation through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues (Fig. 1B) (39,40,46,59-62). Degradation signals and E3 Ub ligases of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively. Nt-acetylation of cellular proteins is apparently irreversible, in contrast to acetylation-deacetylation of proteins' internal Lys residues. About 90% of human proteins are cotranslationally Nt-acetylated by ribosome-associated Nt-acetylases (63). Posttranslational Nt-acetylation is known to occur as well. Ac/N-degrons are present in many, possibly most, Nt-acetylated proteins (Fig. 1B). Natural Ac/N-degrons are regulated through their reversible shielding in cognate protein complexes (59).

The pathway's other branch, called the Arg/N-end rule pathway, targets specific unacetylated N-terminal residues (Fig. 1A) (40,64-68). The "primary" destabilizing N-terminal residues Arg, Lys, His, Leu, Phe, Tyr, Trp, and Ile are directly recognized by N-recognins. The unacetylated N-terminal Met, if it is followed by a bulky hydrophobic ($) residue, also acts as a primary destabilizing residue (Fig. 1A) (40). In contrast, the unacetylated N-terminal Asn, Gln, Asp, and Glu (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications, which include N-terminal deamidation (Nt-deamidation) of Asn and Gln and Nt-arginylation of Asp, Glu and oxidized Cys (Fig. 1A) (41-43,69). In the yeast *Saccharomyces cerevisiae*, the Arg/N-end rule pathway is mediated by the Ubr1 N-recognin, a 225 kDa RING-type E3 Ub ligase and a part of the targeting complex comprising the Ubr1-Rad6 and Ufd4-Ubc4/5 E2-E3 holoenzymes (41,70). In multicellular eukaryotes, several E3 Ub ligases, including Ubr1, function as N-recognins of the Arg/N-end rule pathway (Fig. 1A).

In *S. cerevisiae*, the Ac/N-end rule pathway is mediated by (at least) the cytosolic/nuclear E3 Ub ligase Not4 and by Doa10, an endoplasmic reticulum (ER) membrane-embedded E3 (59). In mammalian cells, this pathway is mediated by (at least) the Teb4 E3, which is sequelogous (similar in sequence (71)) to yeast Doa10 (60). Human RGS2, a regulator of specific G proteins, has been shown to be a short-lived substrate of the Ac/N-end rule pathway both in mammalian cells and in the heterologous setting of *S. cerevisiae* (60,72,73). In addition, the naturally occurring (blood pressure-elevating) human $RGS2^{Q2L}$ mutant (in which Gln at position 2 is replaced by Leu) is targeted for degradation by both the Ac/N-end rule pathway and the Arg/N-end rule

pathway in either yeast or mammalian cells (60). The Ac/N-end rule pathway recognizes the Nt-acetylated Ac-RGS2^{Q2L} (specifically, its N-terminal Ac-Met residue), whereas the non-Nt-acetylated RGS2^{Q2L} is targeted (through its N-terminal Met-Leu sequence) by the Arg/N-end rule pathway (60).

In the present work, we analyzed the proteasome-mediated degradation of rat AANAT (37), whose N-terminal sequence Met-Leu (Fig. 1C) is identical to that of the previously examined and otherwise unrelated human RGS2^{Q2L} protein (60). We also characterized human AANAT, which is highly sequelogous (71) (84% identical) to rat AANAT but bears a different N-terminal sequence (Fig. 1C). Our analyses employed, among other things, the previously helpful approach of dissecting degradation of a mammalian protein of interest not only in a homologous (mammalian) setting but also in *S. cerevisiae* (60,66), thereby making possible the use of yeast genetics.

We show here that two alternative versions of rat AANAT, its Nt-acetylated and non-Nt-acetylated forms, are targeted for degradation by the Ac/N-end rule pathway and the Arg/N-end rule pathway, respectively. In contrast, human AANAT, whose N-terminal sequence differs from that of rodent AANATs, is significantly longer-lived than its rat counterpart, and appears to be largely refractory to degradation by the N-end rule pathway. Together, these and related results indicate both a major involvement of the N-end rule pathway in the control of rodent AANATs and substantial differences in the regulation of rodent and human AANATs that stem from differences in their N-terminal sequences.

EXPERIMENTAL PROCEDURES

Yeast Strains, Media, and Genetic Techniques – Standard yeast genetic techniques were used (74-76). *S. cerevisiae* strain BWY29 was constructed by transforming the strain CHY345 (*ubr1•::LEU2* in the strain background of BY4742) with a PCR-amplified DNA fragment that encoded the selection marker KanMX6 (making cells resistant to kanamycin) and was targeted to the 5' and 3' flanking regions of the *NAA10* gene, thereby replacing the open reading frame (ORF) of *NAA10* with *KanMX6* (Table S1). *S. cerevisiae* strain JOY487 was made by transforming BY4742 with a PCR-amplified DNA fragment that encoded the natNT2 marker (making cells resistant to nourseothricin) and was targeted to the 5' and 3' flanking regions of the *PDR5* gene, thereby replacing the ORF of *PDR5* with *natNT2* (Table S1). The resulting *pdr5•* cells were used in experiments that involved the MG132 proteasome inhibitor (Fig. A1.S1A, B). Other *S. cerevisiae* strains used in this study were constructed previously and are cited in Table S1.

S. cerevisiae were transformed using the LiAc/PEG method (77). *S. cerevisiae* media included YPD medium (1% yeast extract, 2% peptone, 2% glucose; only the most relevant components are cited); synthetic complete (SC) medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose); and synthetic drop-out (SD) medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose, plus a drop-out mixture of amino acids as required by a given auxotrophic strain) (75,76).

Plasmids, cDNAs, and Primers – NEB Turbo *Escherichia coli* (New England Biolabs (NEB), Ipswich, MA) was used for cloning and maintaining plasmids. Phusion High-Fidelity DNA polymerase (NEB) was used for carrying out PCR. Nucleotide

sequences of all plasmids were verified by DNA sequencing. The plasmids and PCR primers used in this study are described in Tables S2 and S3, respectively.

The low copy pBW105, pBW106, pBW107, and pBW209 plasmids expressed respectively, the wild-type rat ^{MLS}rAANAT$_{3f}$ (wild-type rat AANAT bearing the N-terminal sequence Met-Leu-Ser and C-terminally tagged with a triple-FLAG epitope) and its N-terminal mutants $^{(M)SI}$rAANAT$_{3f}$, $^{(M)SMLS}$rAANAT$_{3f}$, and $^{(M)PLS}$rAANAT$_{3f}$ from the P_{GAL1} promoter. To construct these plasmids, the rat *Aanat* ORF was amplified from pCISII-rAANAT (78), using primers BW2394, BW240, BW241, BW412, BW243, and BW244 (Table S3). The resulting PCR products were digested with EcoRI and HindIII, and cloned in EcoRI/HindIII-cut pRS416$_{GAL1}$. See items i-iv at the beginning of Results for information about the above rat AANAT notations.

The low copy pBW135 and pBW394 plasmids expressed, respectively, the wild-type rat ^{MLS}rAANAT C-terminally tagged with the triple-HA epitope (^{MLS}rAANAT$_{3ha}$) and its Lys-8-to-Arg-8 mutant ^{MLS}rAANAAT$_{3ha}^{K8R}$ from the P_{CUP1} promoter. To construct these plasmids, the rat *Aanat* ORF was amplified from pCISII-rAANAT (78) using primers BW339, BW704, BW338, BW339, BW340 (Table S3). The resulting PCR-amplified DNA fragments were digested with EcoRI and XhoI, and cloned in EcoRI/XhoI-cut pRS313$_{CUP1}$ (59,79). See items v and vi at the beginning of Results for information about the above rat AANAT notations.

The low copy pBW395 and pBW419 plasmids expressed, respectively, the wild-type human $^{(M)ST}$hAANAT C-terminally tagged with a triple-FLAG epitope ($^{(M)ST}$hAANAT$_{3f}$) and its N-terminal mutant $^{(M)PT}$hAANAT$_{3f}$ from the P_{GAL1} promoter. To construct these plasmids, the human *AANAT* ORF from Genscript Clone #OHu55705

(Genscript, Piscataway, NJ) was amplified using primers BW705, BW711, BW712, and BW244, or primers BW771, BW711, BW712, and BW244, respectively (Table S3). The PCR-amplified DNA fragments were digested with EcoRI and HindIII, and were ligated into EcoRI/HindIII-cut pRS416$_{GAL1}$ (79). See items vi and vii at the beginning of Results for information about the above human AANAT notations.

For expression of the above AANAT-encoding ORFs in human HEK293T cells, a wild-type *Aanat* ORF and its derivatives were cloned into a modified expression vector derived from pcDNA3. That (modified) plasmid, termed pcDNA3$_{CMVt1}$, contained a truncated (weakened) version of the P$_{CMV}$ promoter, termed P$_{CMVt1}$. The pcDNA3$_{CMVt1}$ plasmid was constructed by amplifying, using PCR, the 276-base pairs (bp) fragment that encompassed 104 bp at the 3' end of the P$_{CMV}$ promoter region in pcDNA3, its T7 promoter site, and its multiple cloning site, using primers BW375 and BW376 (Table S3). The amplified DNA fragment was digested with NruI and XbaI and ligated into NruI/XbaI-cut pcDNA3, yielding the plasmid pBW173 (Tables S2). Its truncated P$_{CMVt1}$ promoter retained several proximal promoter elements, including a consensus NFkB site, a CREB-binding site, and two SP1/SP3-binding sites, but lacked more distal enhancer elements. This resulted in a ~70% lower activity of the P$_{CMVt1}$ promoter in HEK293T cells, in comparison to P$_{CMV}$ (Fig. A1.S3A-C).

The pcDNA3$_{CMVt1}$-based plasmids pBW206, pBW211, pBW212, and pBW213 expressed, in HEK293T cells, either the wild-type rat ^{MLS}rAANAT$_{3f}$ or its N-terminal mutants $^{(M)SI}$rAANAT$_{3f}$, $^{(M)SMLS}$rAANAT$_{3f}$, and $^{(M)PLS}$rAANAT$_{3f}$ from the P$_{CMVt1}$ promoter. To construct these plasmids, the *Aanat* ORF was amplified from the plasmids pBW105, pBW106, pBW107, and pBW209 using primers BW352, BW353, BW354,

BW413, and BW355, respectively (Table S3). The amplified DNA fragments were digested with EcoRI and XhoI, and ligated in EcoRI/XhoI-cut pBW173 (Table S2). All pcDNA3-based constructs designed for expression in HEK293T cells contained the Kozak sequence 5'-ACC immediately upstream of the start codon.

pBW477, which expressed the ^{MLS}rAANAT$_{3ha}$ from the P$_{CMVt1}$ promoter, was constructed by using pBW135 as a PCR template and the primers BW352/BW340. The resulting PCR-amplified DNA fragment was digested with EcoRI and XhoI, and ligated into EcoRI/XhoI-cut pBW173 (Table S2). pBW478 expressed ^{MLS}rAANAAT$_{3ha}^{Kzero}$, i.e., the lysine-lacking derivative of ^{MLS}rAANAT$_{3ha}$. It was constructed by using pBW135 as a PCR template to generate four separate DNA fragments containing lysine-to-arginine mutations. The primer pairs were 704/784, 785/786, 787/788, and 789/340 (Table S3). These fragments were then assembled into a contiguous ORF using PCR. The resulting DNA fragment, encoding lysine-to-arginine mutations at each of the 4 lysines contained in the wild-type rat ^{MLS}rAANAT$_{3f}$, was digested with EcoRI and XhoI, and ligated into EcoRI/XhoI-cut pBW173.

pBW466 and pBW467, the pcDNA3-based counterparts of pBW395 and pBW419 (Table S2) were generated by PCR using pBW395 and pBW419 as templates, and the primers BW866/BW355 and BW868/BW355, respectively (Table S3). PCR-amplified DNA fragments were digested with EcoRI and XhoI and ligated into EcoRI/XhoI-cut pBW173 (Table S2).

Cell Culture – The HEK293T cell line (derived from human embryonic kidney cells) was obtained from American Type Culture Collection (ATCC, Manassas, VA), and was grown at 37°C in 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) medium

supplemented with 10% fetal bovine serum (FBS) (Gemini Bio-Products, West Sacramento, CA) and penicillin/streptomycin (100 units/ml; Hyclone). Cells were transfected using Lipofectamine-2000 (Invitrogen) according to the manufacturer's protocol.

Cycloheximide Chase Assay in S. cerevisiae – Cycloheximide (CHX) chase assays in *S. cerevisiae* were performed largely as described (39,59). Briefly, *S. cerevisiae* cells were grown in overnight cultures, and expression of epitope-tagged test proteins was induced by the addition of 2% galactose (for plasmids containing the P_{GAL1} promoter) or 0.1 mM $CuSO_4$ (for plasmids containing the P_{CUP1} promoter (80)). 3 h post-induction, CHX was added to the final concentration of 0.1 mg/ml. Samples were collected at the indicated time points, centrifuged at 11,200*g* for 2 min to pellet the cells, and snap-frozen in liquid nitrogen. Proteins were extracted by resuspending the pellets in 1 ml of 0.2 M NaOH and incubating on ice for 20 min (81). Cells were then pelleted and resuspended in 50 µl of SUMEB loading buffer (1% SDS, 8 M urea, 10 mM EDTA, 0.01% bromophenol blue, 10 mM MOPS, pH 6.8), followed by heating at 95°C for 10 min. Samples were then centrifuged at 11,200g for 5 min, followed by the SDS-PAGE (with 10 µl loaded per well) using 4-12% NuPAGE gels (Invitrogen). Fractionated proteins were transferred onto nitrocellulose membranes for immunoblotting analyses (see below).

Cycloheximide Chase Assay in Mammalian Cell Lines – HEK293T cells were transfected with 4 µg of plasmid DNA per well (10 cm^2 surface area) in 6-well plates using Lipofectamine-2000 (Invitrogen) according to the manufacturer's protocol. 24 h after transfection cells were treated with cycloheximide (CHX; 0.1 mg/ml) to initiate the

CHX chase. All cells from one well were collected at each of indicated time-points by washing cells quickly in ice-cold phosphate-buffered saline (PBS), then rapidly scraping cells into 1.5 ml tubes, pelleting them by centrifugation at 4°C (at 10,000g for 1 min), and snap-freezing the pellets in liquid nitrogen. Cells in these samples were later lysed by the addition of 0.2 ml of "mammalian lysis buffer" (1% NP40, 0.15 M NaCl, 50 mM Tris-HCl, pH 7.5) containing 1x Roche Complete Protease Inhibitor Cocktail, and brief sonication, followed by centrifugation at 11,200*g* for 10 min. Total protein concentration in the supernatants was measured by bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, Waltham, MA). 30 µg of total protein in a thus prepared extract in lithium dodecyl sulfate (LDS)-sample buffer (in the volume of 45 µl) were heated at 70°C for 10 min, followed by LDS-PAGE on a 4-12% Bis-Tris NuPAGE gel (Invitrogen), and subsequent transfer onto nitrocellulose membranes for immunoblotting.

Immunoblotting – Following electrophoresis, proteins separated by LDS-PAGE or SDS-PAGE, as indicated above, were electroblotted on nitrocellulose membranes by iBlot (Invitrogen, Carlsbad, CA; Program 3; 7 min transfer). Membranes were blocked and incubated with the appropriate primary antibody, followed by LI-COR IRDye-conjugated secondary antibodies. IRDye fluorescence was detected using LI-COR Odyssey 9120 (LI-COR, Lincoln, NE), facilitating quantification of immunoblots. All quantification was done on LI-COR Odyssey software. Antibodies used include anti-FLAG mouse monoclonal antibody (clone M2; 1:2,000 dilution; Sigma-Aldrich, St. Louis, MO), anti-±-tubulin mouse monoclonal antibody (clone B-5-1-2; 1:10,000 dilution; Sigma-Aldrich), anti-HA mouse monoclonal antibody (clone HA-7; 1:2000 dilution; Sigma-Aldrich), anti-² -actin mouse monoclonal antibody Ab8224 (Abcam,

1:4,000 dilution; Cambridge, UK), and anti-14-3-3¶ rabbit monoclonal antibody EPR6379 (1:100,000 dilution; Abcam).

Immunoprecipitation of Polyubiquitylated AANAT – HEK293T cells were transfected with 4 µg of plasmid DNA per well (10 cm^2 surface area) in 6-well plates using Lipofectamine-2000 (Invitrogen) according to the manufacturer's protocol. 24 h after transfection, cells were treated with the proteasome inhibitor MG132 (final concentration of 10 µM; AG Scientific, San Diego, CA), or with an equivalent volume of dimethylsulfoxide (DMSO) in which the 10 mM stock solution of MG132 was made, as indicated. 6 h post-treatment cells were collected by washing them quickly on a plate in ice-cold phosphate-buffered saline (PBS), then rapidly scraping cells into 1.5 ml tubes and pelleting them by centrifugation at 4°C at 10,000*g* for 1 min). Pellets were resuspended in 0.1 ml of buffer containing 1% SDS and 50 mM Tris (pH 7.5), and heated at 95°C for 10 min. After boiling, samples were diluted with 1 ml of buffer TNN (0.5% NP40, 0.25 M NaCl, 5 mM EDTA, 50 mM Tris, pH 7.5) containing 1x Roche Complete Protease Inhibitor Cocktail, 20 mM *N*-ethylmaleimide (NEM, Sigma-Aldrich) and 50 µM PR-619 (LifeSensors, Malvern, PA) as inhibitors of deubiquitylation. 25 µl of anti-HA magnetic beads were then added to each sample, which were incubated at 4°C overnight, followed by 3 washes in TNN, 1 wash in 10 mM Tris-HCl, pH 8.5, and elution with 50 µl 0.1 M glycine, pH 2.0 by incubation at room temperature for 10 min. After elution, the samples were neutralized, then heated at 70°C in LDS-PAGE sample buffer, followed by LDS-PAGE on a 4-12% NuPAGE gel. Fractionated proteins were then electroblotted onto nitrocellulose membranes. The resulting membranes (containing transferred proteins) were autoclaved as previously described (82) prior to "downstream"

immunoblotting procedures with either anti-HA or anti-ubiquitin antibodies to increase the sensitivity of detection of polyubiquitin chains.

RESULTS

Wild-type Rat and Human AANATs and their Mutants – The start (AUG) codon-encoded N-terminal Met residue of nascent proteins is cotranslationally cleaved off by ribosome-associated Met-aminopeptidases (MetAPs) if and only if a residue at position 2 (it becomes N-terminal upon the removal of Met) is not larger than Val (41,83). Thus, for example, the Met residue of the N-terminal Met-Leu-Ser sequence of wild-type rat AANAT (denoted as ^{MLS}rAANAT) (Fig. 1C) is retained in mature ^{MLS}rAANAT, inasmuch as Leu is larger than Val. Given the previously demonstrated usefulness of employing genetic tractability of *S. cerevisiae* for understanding the targeting of mammalian N-end rule substrates (60,66), we carried out this study by expressing AANAT test proteins not only in human HEK293T cells but also in *S. cerevisiae*. Cited below are specific wild-type and mutant AANAT proteins that have been examined in this study.

(i) ^{MLS}rAANAT$_{3f}$, the wild-type rat ^{MLS}rAANAT C-terminally tagged with a triple-FLAG epitope (Figs. 1C and 2A).

(ii) $^{(M)SMLS}$rAANAT$_{3f}$, in which the sequence (Met)-Ser was placed before the wild-type Met-Leu-Ser sequence of rat ^{MLS}rAANAT$_{3f}$, yielding the N-terminal sequence (Met)-Ser-Met-Leu-Ser. Its N-terminal Met residue would be cotranslationally cleaved off by MetAPs (Fig. 2A). The (Met)-Ser sequence was added to wild-type ^{MLS}rAANAT$_{3f}$

in order to mimic the (Met)-Ser N-terminal sequence of primate (including human) AANATs (Fig. 1C).

(iii) $^{(M)SI}$rAANAT$_{3f}$, in which Leu at position 2 of the wild-type rat ^{MLS}rAANAT$_{3f}$ (Fig. 1C) was deleted, thereby making the N-terminal Met of the resulting $^{(M)SI}$rAANAT$_{3f}$ removable by MetAPs (Figs. 1C and 2A).

(iv) $^{(M)PLS}$rAANAT$_{3f}$, in which the Pro residue was inserted between N-terminal Met and second-position Leu of wild-type rat ^{MLS}rAANAT$_{3f}$. This alteration aimed to address the relevance of Nt-acetylation to the degradation of wild-type rat ^{MLS}rAANAT$_{3f}$. Specifically, the N-terminal sequence Met-Leu-Ser is *a priori* likely to be cotranslationally Nt-acetylated *in vivo* at its (retained) N-terminal Met (59,62,63). In contrast, the N-terminal Met residue of the mutant N-terminal sequence (Met)-Pro-Leu-Ser of $^{(M)PLS}$rAANAT$_{3f}$ would be cotranslationally cleaved off by MetAPs. The resulting N-terminal Pro is not Nt-acetylated at least in *S. cerevisiae*, and is usually not Nt-acetylated in mammalian cells as well (59,63).

(v) ^{MLS}rAANAT$_{3ha}$, the wild-type rat ^{MLS}rAANAT (Fig. 1C) C-terminally tagged with a triple-HA epitope. It was used as a control to the ^{MLS}rAANAAT$_{3ha}^{K8R}$ mutant described in item viii.

(vi) $^{(M)ST}$hAANAT$_{3f}$, the wild-type human $^{(M)ST}$hAANAT (Fig. 1C) C-terminally tagged with a triple-FLAG epitope.

(vii) $^{(M)PT}$hAANAT$_{3f}$, in which the second-position Ser of the wild-type human $^{(M)ST}$hAANAT$_{3f}$ (Fig. 1C) was replaced by the Pro residue, yielding $^{(M)PT}$hAANAT$_{3f}$. The N-terminal Met residue of either the wild-type human $^{(M)ST}$hAANAT$_{3f}$ or the mutant $^{(M)PT}$hAANAT$_{3f}$ would be cotranslationally removed by MetAPs (41,83). However, in

contrast to the resulting N-terminal Ser of wild-type SThAANAT$_{3f}$, which would be expected to be cotranslationally Nt-acetylated, the N-terminal Pro of the mutant ^{PT}hAANAT$_{3f}$ is not Nt-acetylated at least in *S. cerevisiae*, and is usually not Nt-acetylated in mammalian cells as well (59,63).

(viii) ^{MLS}rAANAAT$_{3ha}^{K8R}$, the otherwise wild-type triple-HA-tagged rat ^{MLS}rAANAT in which the Lys-8 residue was replaced by Arg.

(ix) ^{MLS}rAANAAT$_{3ha}^{Kzero}$, the otherwise wild-type triple-HA-tagged rat ^{MLS}rAANAT in which each of its four Lys residues were replaced by Arg (Fig. 5B, D).

(x) $^{(M)ST}$hAANAAT$_{3ha}^{Kzero}$, the otherwise wild-type triple-HA-tagged human $^{(M)ST}$hAANAT in which both its Lys residues were replaced by Arg (Fig. 5B, D).

Degradation of Rat AANAT by the S. cerevisiae N-end Rule Pathway – In cycloheximide (CHX) assays, a protein of interest is analyzed by immunoblotting as a function of time after the inhibition of translation by CHX (39,59,60). Both the wild-type rat ^{MLS}rAANAT$_{3f}$ (Fig. 1C) and its $^{(M)SMLS}$rAANAT$_{3f}$ mutant (see item ii above) were rapidly degraded in wild-type *S. cerevisiae*, with half-lives ($t_{1/2}$) < 5 min (Fig. 2B, lanes 3-10, and Fig. 2D). In striking contrast, $^{(M)SI}$rAANAT$_{3f}$ (see item iii above), produced from wild-type ^{MLS}rAANAT$_{3f}$ by deleting the Leu residue at position 2, was virtually completely stable under the same conditions, with a $t_{1/2}$ > 1 h (Fig. 2B, lanes 11-14 and Fig. 2D). When present, parentheses around a superscript's N-terminal Met residue in the notations of AANAT test proteins denote the fact that this Met is cotranslationally cleaved off by MetAPs. See items ii and iv above.

Since the N-terminal Met residue in the Met-Leu-Ser sequence of the wild-type rat ^{MLS}rAANAT$_{3f}$ would be likely to be Nt-acetylated *in vivo* (59,62,63), the rapid

degradation of ^{MLS}rAANAT$_{3f}$ ($t_{1/2}$ < 5 min) (Fig. 2B, D) suggested the involvement of the Ac/N-end rule pathway (Fig. 1A). To address this possibility, we constructed the non-Nt-acetylatable $^{(M)PLS}$rAANAT$_{3f}$ mutant (see item iv above). (This mutant would also not be expected to be targetable by the Arg/N-end rule pathway, as the latter does not recognize N-terminal Pro (40,41)). Remarkably, $^{(M)PLS}$rAANAT$_{3f}$ was long-lived in *S. cerevisiae*, in comparison to the short-lived wild-type ^{MLS}rAANAT$_{3f}$ ($t_{1/2}$ H1h versus $t_{1/2}$ < 5 min, respectively), suggesting (but not proving) a role for the Ac/N-end rule pathway in the degradation of the wild-type rat ^{MLS}rAANAT$_{3f}$ (Fig. 2B, lanes 15-18, and Fig. 2D; see also below).

To further address the involvement of the Ac/N-end rule pathway and the Arg/N-end rule pathway (Fig. 1A, B) in the degradation of wild-type ^{MLS}rAANAT$_{3f}$, we performed CHX-chases in *S. cerevisiae* mutants that lacked essential components of either one or both of these pathways. The normally short-lived rat ^{MLS}rAANAT$_{3f}$ ($t_{1/2}$ < 5 min; Fig 2C, lanes 2-5, and Fig. 2E) was partially stabilized in both *naa30*" and *ubr1*" mutants (Fig. 2C, lanes 6-9, lanes 14-17, and Fig. 2E). *S. cerevisiae naa30*" cells lack the cognate NatC Nt-acetylase whose substrates include proteins bearing the N-terminal Met-Leu sequence (59,62,63,84,85). *S. cerevisiae ubr1*" cells lack Ubr1, the E3 Ub ligase (N-recognin) that is essential for the proteolytic activity of the Arg/N-end rule pathway (Fig. 1A) (40,41,59).

Crucially, the relative stabilization of ^{MLS}rAANAT$_{3f}$ was dramatically higher in the double mutant *naa30*" *ubr1*" than in either one of the single mutants (Fig 2C, lanes 18-21, and Fig.2E). These findings should be considered together with the conceptually independent result that the non-Nt-acetylatable (and also not targetable by the Arg/N-end

rule pathway) $^{(M)PLS}$rAANAT$_{3f}$ mutant (see the preceding paragraph) was strikingly longer-lived in wild-type *S. cerevisiae* than its wild-type ^{MLS}rAANAT$_{3f}$ counterpart (Fig. 2B, lanes 15-18 vs. lanes 3-6, and Fig. 2D). Together, these sets of findings indicated that the wild-type rat ^{MLS}rAANAT$_{3f}$ (its two alternative forms) was targeted for degradation by both the Ac/N-end rule pathway and the Arg/N-end rule pathway. Specifically, the Nt-acetylated Ac-^{MLS}rAANAT$_{3f}$ is destroyed through the recognition of its Nt-acetylated N-terminal Met residue by the Ac/N-end rule pathway (Fig. 1B), whereas the non-Nt-acetylated ^{MLS}rAANAT$_{3f}$ is targeted by the "complementary" Arg/N-end rule pathway, which recognizes the unacetylated N-terminal Met-Leu sequence of ^{MLS}rAANAT$_{3f}$ as the Met-$ motif (N-terminal Met followed by a bulky hydrophobic residue) (Fig 1A).

This conclusion was in agreement with the finding that the rapid degradation of ^{MLS}rAANAT$_{3f}$ in wild-type *S. cerevisiae* ($t_{1/2}$ < 5 min) remained unchanged in *naa10*" cells, which lacked the non-cognate NatA Nt-acetylase (which does not Nt-acetylate N-terminal Met) (59,85). Furthermore and also in agreement with the above conclusion, the degradation of ^{MLS}rAANAT$_{3f}$ in double-mutant *naa10*" *ubr1*" cells proceeded at a rate similar to that in the single *ubr1*" mutant (Fig. 2C, lanes 22-25, and Fig. 2E). Thus, a double-mutant background that nearly completely stabilizes the wild-type rat ^{MLS}rAANAT$_{3f}$ must be *naa30*" *ubr1*" in that it must also lack the activity of the cognate NatC Nt-acetylase (Fig. 2C, lanes 18-21, and Fig. 2E). In agreement with this conclusion, the ablation of *NAA10*, encoding the non-cognate NatA Nt-acetylase, did not have a synergistic effect on the rate of ^{MLS}rAANAT$_{3f}$ degradation in double-mutant *naa10*" *ubr1*" cells (Fig. 2C, lanes 22-25, and Fig. 2E).

Previous work has identified the ER membrane-embedded E3 Ub ligase Doa10 as one Ac/N-recognin of the *S. cerevisiae* Ac/N-end rule pathway (39,59). We found that the degradation of the wild-type rat ^{MLS}rAANAT$_{3f}$ ($t_{1/2}$ < 5 min in wild-type *S. cerevisiae*) was not significantly impaired in *doa10*" cells, in comparison to wild-type cells (Fig. 3D, F). We did not examine, so far, the other yeast Ac/N-recognin, Not4 (see the Introduction), for its possible role in targeting the Nt-acetylated rat Ac-^{MLS}rAANAT$_{3f}$ in *S. cerevisiae*. We also don't know, thus far, whether Teb4, the mammalian counterpart of the yeast Doa10 N-recognin (60), is involved in the degradation of rat ^{MLS}rAANAT$_{3f}$ in a homologous (mammalian) setting.

Degradation of Rat AANAT Is Proteasome-Dependent and Does Not Require Lys-8 – The demonstrated degradation of two forms of the wild-type rat ^{MLS}rAANAT$_{3f}$ (Ac-^{MLS}rAANAT$_{3f}$ and its non-Nt-acetylated counterpart) by the two branches of the N-end rule pathway (Fig. 2) already implied the proteasome dependency of this degradation, given the known organization of the N-end rule pathway (Fig. 1A, B). To verify this in *S. cerevisiae* by independent means, CHX-chases with rat ^{MLS}rAANAT$_{3f}$ were carried out in wild-type vs. *pdr5*" cells in either the presence or absence of the MG132 proteasome inhibitor. (Cells lacking the transmembrane transporter PDR5 are more sensitive to MG132 (86,87).)

As expected, given the proteasome dependence of the N-end rule pathway (Fig. 1A, B), the normally rapid degradation of the wild-type rat ^{MLS}rAANAT$_{3f}$ was substantially inhibited in wild-type *S. cerevisiae* in the presence of 50 μM MG132 (Fig. A1.S1A, B). This degradation was inhibited even more strongly in mutant *pdr5*" cells

under the same conditions, confirming the proteasome dependence of at least the bulk of ^{MLS}rAANAT$_{3f}$ degradation (Fig. A1.S1A, B).

N-terminal regions of mammalian AANATs contain a highly conserved Lys residue (e.g., Lys-8 in rat AANAT and Lys-10 in human AANAT) that has been suggested as a potential ubiquitylation site and thus a determinant of AANAT's degron (33,88-90). To address this possibility, CHX-chases were performed in *S. cerevisiae* with wild-type ^{MLS}rAANAT$_{3ha}$ and a mutant containing a Lys-8-Arg mutation (^{MLS}rAANAAT$^{K8R}_{3ha}$). Wild-type ^{MLS}rAANAT$_{3ha}$ and the mutant ^{MLS}rAANAAT$^{K8R}_{3ha}$ were degraded at similar rates, indicating that the conserved lysine at this position (Lys-8 in rat ^{MLS}rAANAT) is not, by itself, a significant determinant of ^{MLS}rAANAT stability (Fig. A1.S1C). However, in contrast to the Lys-8-Arg mutation, the conversion of all four lysines of rat ^{MLS}rAANAT to Arg residues did confer metabolic stability on ^{MLS}rAANAT in HEK293T cells (see below and Fig. 4B).

Degradation of human AANAT in S. cerevisiae – Fig. 3A, lanes 6-9, and Fig. 3B show the results of a CHX-chase, in wild-type yeast, with the 207-residue (not counting the tag) $^{(M)ST}$hAANAT$_{3f}$, the wild-type human AANAT C-terminally tagged with a triple-FLAG epitope. Parentheses around the superscript's N-terminal Met residue denote the fact that this Met is cotranslationally cleaved off by MetAPs. While human $^{(M)ST}$hAANAT$_{3f}$ was relatively unstable in *S. cerevisiae* ($t_{1/2}$ H30 min), it was much longer-lived (including its higher zero-time, pre-chase level) than the wild-type rat ^{MLS}rAANAT$_{3f}$ under the same conditions (Fig. 3A, B).

As was also done with the wild-type rat ^{MLS}rAANAT$_{3f}$ (Fig. 2C), we characterized the degradation of wild-type human $^{(M)ST}$hAANAT$_{3f}$ in *S. cerevisiae* strains

deficient in specific components of the N-end rule pathway. The N-terminal Ser-Thr sequence of wild-type human $^{(M)ST}$hAANAT$_{3f}$ (after the cotranslational removal of the initially present N-terminal Met) made it likely that $^{(M)ST}$hAANAT$_{3f}$ was cotranslationally Nt-acetylated *in vivo* (59,63). This modification of human $^{(M)ST}$hAANAT$_{3f}$ would make it a potential target for degradation by the Ac/N-end rule pathway (Fig. 1B). In contrast, the non-Nt-acetylated human $^{(M)ST}$hAANAT$_{3f}$ would not be expected to be recognized by either the Ac/N-end rule or the Arg/N-end rule pathway (Fig. 1A), in the latter case owing to the absence of both a destabilizing N-terminal residue and a bulky hydrophobic ($) second residue. (A Met-$ N-terminal motif, containing a second-position $ residue, can act as an N-degron recognized by the Arg/N-end rule pathway (60,91).)

Indeed, the rate of degradation of human $^{(M)ST}$hAANAT$_{3f}$ in *ubr1*" *S. cerevisiae*, which lacked the Ubr1 N-recognin and therefore lacked the Arg/N-end rule pathway (Fig. 1A), was similar to the rate of $^{(M)ST}$hAANAT$_{3f}$ degradation in wild-type cells (Fig. 3, lanes 1-4 vs. lanes 9-12). The rate of $^{(M)ST}$hAANAT$_{3f}$ degradation characteristic of wild-type *S. cerevisiae* was also not significantly changed in *doa10*" cells, which lacked one of two known Ac/N-recognins of the Ac/N-end rule pathway (Fig. 1B and Fig. 3D, lanes 10-13 vs. lanes 2-5). Surprisingly, however, $^{(M)ST}$hAANAT$_{3f}$ was strongly destabilized in *naa10*" cells, which lacked the cognate Nt-acetylase for the N-terminal Ser residue (59,62); similar results were obtained with double-mutant *naa10*" *ubr1*" cells (Fig. 3C-F).

Given the much faster degradation of $^{(M)ST}$hAANAT$_{3f}$ in the absence of Naa10 (e.g., Fig. 3C, lanes 1-4 vs. lanes 5-8, and Fig. 3E), i.e., in cells that would be incapable of Nt-acetylating $^{(M)ST}$hAANAT$_{3f,}$ we also carried out CHX-chases, in wild-type

S. cerevisiae, with $^{(M)PT}$hAANAT$_{3f}$. In the latter mutant, the N-terminal Ser was replaced by the non-Nt-acetylatable Pro residue. In contrast to wild-type $^{(M)ST}$hAANAT$_{3f}$ in *naa10*" cells, in which the (non-Nt-acetylated) $^{(M)ST}$hAANAT became strikingly short-lived, the (non-Nt-acetylatable) $^{(M)PT}$hAANAT mutant was found to be longer-lived, in wild-type cells, than the Nt-acetylatable wild-type $^{(M)ST}$hAANAT$_{3f}$ (Fig. A1.S2). These results indicated that the observed accelerated degradation of the wild-type human $^{(M)ST}$hAANAT in *naa10*" *S. cerevisiae* was not caused by the failure to Nt-acetylate $^{(M)ST}$hAANAT in the absence of the NatA Nt-acetylase.

We do not understand the mechanistic cause of the (reproducibly observed) much faster degradation of human $^{(M)ST}$hAANAT in *naa10*" *S. cerevisiae* (Fig. 3C-F). One possibility, which remains to be examined, is the apparent inhibition of the yeast Hsp90 chaperone system (centered on the *S. cerevisiae* Hsc82/Hsp82 proteins) in *naa10*" cells (J.-H.O. and A.V., unpublished data). If human $^{(M)ST}$hAANAT expressed in wild-type *S. cerevisiae* is a protected (from degradation) client of the Hsc82/Hsp82 system, a failure of this protection in *naa10*" mutant cells would make $^{(M)ST}$hAANAT vulnerable to a currently unknown pathway of the Ub system that is distinct from the Arg/N-end rule pathway. (A proteolytic pathway in question would be distinct from the Arg/N-end rule pathway because $^{(M)ST}$hAANAT was equally short-lived in single-mutant *naa10*" and double-mutant *naa10*" *ubr1*" cells (Fig. 3C-F).) In sum, the above Hsp90-based mechanism (which remains to be addressed in the context of the $^{(M)ST}$hAANAT protein) might underlie the strongly accelerated degradation of $^{(M)ST}$hAANAT in *naa10*" *S. cerevisiae*.

Degradation of Rat AANAT in Human HEK293T Cells – To compare the relative rates of degradation of N-terminal mutants of rat AANAT in human HEK293T cells, the wild-type rat ^{MLS}rAANAT$_{3f}$ and its mutants $^{(M)SMLS}$rAANAT$_{3f}$, $^{(M)SI}$rAANAT$_{3f}$, and $^{(M)PLS}$rAANAT$_{3f}$ (see items i-iv at the beginning of Results) were expressed in these cells using transient transfection. Preliminary experiments with wild-type ^{MLS}rAANAT$_{3f}$ encoded by a pcDNA3-based plasmid and expressed from the full-strength P$_{CMV}$ promoter indicated that the levels of expression of ^{MLS}rAANAT$_{3f}$ were high enough to saturate or near-saturate pathways that targeted ^{MLS}rAANAT$_{3f}$ for degradation in HEK293T cells. Specifically, at those (high) levels of expression, wild-type rat ^{MLS}rAANAT$_{3f}$ was either stable or was degraded slowly (data not shown).

We addressed this problem by constructing a set of truncated derivatives of the original P$_{CMV}$ promoter, aiming to attenuate its activity. The resulting nested set of truncated promoters, denoted as P$_{CMVt1}$, P$_{CMVt2,}$ and P$_{CMVt3}$, comprised progressively shortened versions of the original P$_{CMV}$. These truncations (Fig. A1.S2A) were similar (though not identical) to a series of P$_{CMV}$ truncations described earlier by Promega Inc. (92). Each of the P$_{CMVt1}$, P$_{CMVt2,}$ and P$_{CMVt3}$ promoter variants was cloned into pcDNA3, replacing full-length P$_{CMV}$.

The wild-type rat ^{MLS}rAANAT$_{3f}$ was cloned into each of the above pcDNA3-based plasmids, and CHX chases were performed in HEK293T cells transiently-transfected with these constructs. Expression of ^{MLS}rAANAT$_{3f}$ from the P$_{CMVt1}$ and P$_{CMVt2}$ promoters reduced time-zero levels of ^{MLS}rAANAT$_{3f}$ (measured by immunoblotting at the beginning of CHX-chase) by ~70% and by ~97%, respectively, in comparison to the unmodified P$_{CMV}$ promoter (Fig. A1.S3). As to the most truncated

promoter, P_{CMVt3}, no ^{MLS}rAANAT$_{3f}$ could be detected at the same level of immunoblotting sensitivity (data not shown). The lower levels of ^{MLS}rAANAT$_{3f}$ expression that have been attained through the use of the P_{CMVt1} and P_{CMVt2} promoters resulted in a strongly increased rate of the posttranslational degradation of ^{MLS}rAANAT$_{3f}$ ($t_{1/2}$ < 2 h vs. $t_{1/2}$ > 8 h when expressed from the full-length P_{CMV} promoter; Fig. A1.S3). Treatment of HEK293T cells with the proteasome inhibitor MG132 during the chase increased the levels of ^{MLS}rAANAT$_{3f}$ expressed from these promoters, thereby confirming that in each case the bulk of degradation was proteasome-dependent (Fig. A1.S3). The pcDNA3$_{CMVt1}$ plasmid, which containing P_{CMVt1}, a weakened version of the P_{CMV} promoter, was chosen for experiments with HEK293T cells in this study.

In agreement with the relative metabolic stabilities of AANAT proteins expressed in *S. cerevisiae* (Fig. 1A, B), the mutant rat $^{(M)SI}$rAANAT$_{3f}$, bearing a deletion of the Leu residue at position 2 of wild-type ^{MLS}rAANAT$_{3f}$ (Fig. 2A), was the longest-lived protein in the set of rat AANAT mutants examined in HEK293T cells ($t_{1/2}$ > 4 h; Fig. 4A, lanes 12-16, and Fig. 4C). The wild-type rat ^{MLS}rAANAT$_{3f}$ and its $^{(M)SMLS}$rAANAT$_{3f}$ mutant (see items i and ii at the beginning of Results) were significantly shorter-lived in HEK293T cells than $^{(M)SI}$rAANAT$_{3f}$, with $t_{1/2}$ H1 h and $t_{1/2}$ < 1 h, respectively, versus $t_{1/2}$ > 4 h (Fig. 4A, lanes 2-11, and Fig. 4C).

Although the mutant $^{(M)PLS}$rAANAT$_{3f}$ was partially stabilized relative to wild-type ^{MLS}rAANAT$_{3f}$, the former protein was still significantly shorter-lived ($t_{1/2}$ H2 h) than the most stable $^{(M)SI}$rAANAT$_{3f}$ mutant (Fig. 4A, lanes 17-21, and Fig. 4C). While N-terminal Pro is not Nt-acetylated (59,63) (this makes N-terminal Pro-bearing proteins invulnerable to the Ac/N-end rule pathway), the N-terminal Pro residue might still be (weakly)

recognized by the Arg/N-end rule pathway in some sequence contexts. Specifically, our SPOT-type, peptide-based binding assays (93) suggested that the Ubr1 N-recognin of the *S. cerevisiae* Arg/N-end rule pathway (and, by inference, mammalian Ubr1 and Ubr2 as well) may weakly bind to N-terminal Pro when this residue is followed by a hydrophobic residue (J.-H. O., B.W. and AV., unpublished data). Thus, one interpretation of the "residual" instability of the rat $^{(M)PLS}$rAANAT$_{3f}$ mutant (in comparison to the longest-lived $^{(M)SI}$rAANAT$_{3f}$ mutant) is that the (non-Nt-acetylated) N-terminal sequence Pro-Leu of $^{(M)PLS}$rAANAT$_{3f}$ is weakly targeted by the Arg/N-end rule pathway, in contrast to the absence of targeting of the N-terminal Ser-Ile sequence of the $^{(M)SI}$rAANAT$_{3f}$ mutant (Fig. 4A, C; see also Fig. 2B, D).

The Bulk of Degradation of Rat AANAT Is Mediated by Polyubiquitylation of Its Lys residue(s) – An earlier study, in which the wild-type rat ^{MLS}rAANAT was apparently not stabilized by conversion of its Lys residues to Arg, was based on measurements of steady-state levels of ^{MLS}rAANAT in HEK293 cells that had been treated or left untreated with the proteasome inhibitor MG132 (37). To address this issue more directly, we generated a triple-HA-tagged mutant, denoted ^{MLS}rAANAAT$_{3ha}^{Kzero}$, of the wild-type rat ^{MLS}rAANAT in which all four of its Lys residues were converted to Arg. (A FLAG-based epitope tag could not be employed in these experiments, because FLAG contains a Lys residue that could potentially serve as a target for polyubiquitylation.)

CHX chases in HEK293T cells (performed as described above) with the wild-type ^{MLS}rAANAT$_{3ha}$ versus its lysine-lacking ^{MLS}rAANAAT$_{3ha}^{Kzero}$mutant indicated that ^{MLS}rAANAAT$_{3ha}^{Kzero}$was nearly completely stable during the chase ($t_{1/2}$ >> 4 h), whereas the wild-type ^{MLS}rAANAT$_{3ha}$ was degraded, with $t_{1/2}$ H2 h (Fig. 4B and 4D). A treatment

with MG132 during CHX-chase strongly stabilized the wild-type ^{MLS}rAANAT$_{3ha}$, confirming that its degradation is proteasome-dependent (Figure 4B, lanes 2 and 6, and Fig. 4D). In contrast, the MG132 treatment during CHX-chase of the lysine-lacking ^{MLS}rAANAAT$_{3ha}^{Kzero}$ mutant produced at most a marginal stabilizing effect, suggesting that even if there is a small (residual) amount of lysine-independent degradation of ^{MLS}rAANAAT$_{3ha}^{Kzero}$, such a degradation would still require the proteasome (Figure 4B, lanes 7 and 11, and Fig. 4D).

In a different approach to the same problem, HEK293T cells were transiently-transfected with the pcDNA3 vector or with plasmids expressing either ^{MLS}rAANAT$_{3ha}$ or ^{MLS}rAANAAT$_{3ha}^{Kzero}$. Cells were then treated for 6 h with either MG132 or an equivalent volume of DMSO (in which the stock solution of MG132 was made). Thereafter cell extracts were immunoprecipitated with anti-HA magnetic beads, and the immunoprecipitates were fractionated by SDS-PAGE, followed by immunoblotting with anti-HA and anti-ubiquitin antibodies. Polyubiquitin chains (presumably linked to ^{MLS}rAANAT$_{3ha}$) could be readily detected in fractionated immunoprecipitates of extracts from cells that expressed ^{MLS}rAANAT$_{3ha}$ (but, crucially, not from cells that expressed an empty vector), and the levels of polyubiquitin chains were significantly increased in the presence of MG132 (Fig. 4F, lanes 1-4). In contrast, significantly lower amounts of polyubiquitin chains were observed in fractionated immunoprecipitates of extracts from cells that expressed the lysine-lacking ^{MLS}rAANAAT$_{3ha}^{Kzero}$ (Fig. 4F, lane 5). Moreover, the presence of MG132 did not increase the low levels of detected polyubiquitin chains (Fig. 4F, lane 5 vs. lane 6). The latter result suggested that (residual) polyubiquitin chains observed in cells that expressed ^{MLS}rAANAAT$_{3ha}^{Kzero}$ may be of a kind (e.g., (94)) that do

not contribute to the proteasome-dependent degradation of ^{MLS}rAANAAT$_{3ha}^{Kzero}$. In agreement with this interpretation, the lysine-lacking ^{MLS}rAANAAT$_{3ha}^{Kzero}$ mutant was expressed at significantly higher steady-state levels than the wild-type rat ^{MLS}rAANAT$_{3ha}$ in the absence of MG132 (in DMSO-treated cells) (Fig. 4E, lower panel), suggesting that the absence of Lys residues stabilized ^{MLS}rAANAAT$_{3ha}^{Kzero}$ against degradation throughout life histories of ^{MLS}rAANAAT$_{3ha}^{Kzero}$ molecules, including, possibly, their degradation during or immediately after their synthesis.

Degradation of Human AANAT in Human HEK293T Cells – Degradation of the wild-type human $^{(M)ST}$hAANAT$_{3f}$ and its N-terminal Pro residue-bearing mutant $^{(M)PT}$hAANAT$_{3f}$ (see items vi and vii at the beginning of Results) were examined by CHX-chases in HEK293T cells. In agreement with the findings in *S. cerevisiae* (Figs. 2 and 3), the wild-type human $^{(M)ST}$hAANAT was longer-lived than the wild-type rat ^{MLS}rAANAT$_{3f}$ in HEK293T cells (compare Fig. 5A with Fig. 4A). Interestingly, the mutant human $^{(M)PT}$hAANAT$_{3f}$ was less stable than the wild-type human ^{MLS}rAANAT$_{3f}$ during the course of a 2 h chase in HEK293T cells (Fig. 5A, C).

We also asked whether the (relatively slow) degradation of the wild-type human $^{(M)ST}$hAANAT$_{3f}$ (Fig. 5A, C) required both its lysine-dependent ubiquitylation and the proteasome. CHX-chases were carried out with $^{(M)ST}$hAANAT$_{3ha}$ and its lysine-lacking mutant $^{(M)ST}$hAANAAT$_{3ha}^{Kzero}$. The latter protein was stable during the chase ($t_{1/2}$ >> 4 h), whereas wild-type $^{(M)ST}$hAANAT$_{3ha}$ exhibited a weak but observable instability (Fig. 5B, D). Treatment with the MG132 proteasome inhibitor significantly increased the level of the wild-type human $^{(M)ST}$hAANAT$_{3ha}$ during CHX-chase, confirming proteasome dependence of its (slow) degradation (Fig. 5B, lanes 3 and 7, and Fig. 5D).

DISCUSSION

The ~23-kDa arylalkylamine N-acetyltransferase (AANAT), an enzyme apparently universal among animals (it is present in plants as well), converts the neurotransmitter serotonin to *N*-acetylserotonin (NAS), a regulatory compound in its own right and the immediate precursor of melatonin, a circulating hormone that regulates sleep and other circadian processes in vertebrates. The levels of melatonin and NAS are modulated by oscillatory circadian circuits and at the same time impact those circuits, in addition to other effects of these compounds (see the Introduction), hence the importance of temporal control of the AANAT enzyme. The expression of AANAT is regulated through both transcriptional and posttranslational mechanisms. Specific details of these mechanisms differ among vertebrates. In addition, although AANAT enzymes from different species are highly sequelogous (similar in sequence (71)) throughout their ORFs, the sequences of the first ~10 amino acid residues of ANAATs tend to differ even among relatively closely related mammals, let alone other animals (Fig. 1C). This pattern of variation suggests a blend of adaptive (function-linked, natural selection-based) changes in N-terminal sequences of AANATs versus an unknown (extent-wise) but apparently strong flux of unselected, drift-mediated changes of these N-terminal sequences on evolutionary timescales.

In the present study, we approached the proteolysis-based regulation of rat and human AANATs (Fig. 1C) by expressing them and their mutants not only in mammalian cells but also in the yeast *S. cerevisiae*. This strategy made possible the use of yeast genetics to illuminate specific proteolytic pathways involved, in parallel with experiments in mammalian cells.

We found that the wild-type rat AANAT (^{MLS}rAANAT; the superscript indicates the N-terminal Met-Leu-Ser sequence of this AANAT) is targeted for degradation by two complementary branches of the N-end rule pathway. Specifically, we showed that the $N^{\pm}$-terminally acetylated (Nt-acetylated) Ac-^{MLS}rAANAT is destroyed through the recognition of its Nt-acetylated N-terminal Met residue by the Ac/N-end rule pathway, whereas the non-Nt-acetylated ^{MLS}rAANAT is targeted for degradation by the Arg/N-end rule pathway, which recognizes the unacetylated N-terminal Met-Leu sequence of ^{MLS}rAANAT. (See the Introduction and Fig. 1A, B for a description of mechanisms and functions of the N-end rule pathway.)

The N-terminal (Met)-Ser-Thr sequence of the wild-type human $^{(M)ST}$hAANAT is different from the Met-Leu-Ser sequence of rat ^{MLS}rAANAT (Fig. 1C). (The Met residue of $^{(M)ST}$hAANAT is in parentheses to indicate its cotranslational removal by MetAPs; see the beginning of Results.) We found that the wild-type human $^{(M)ST}$hAANAT is considerably longer-lived than its rat ^{MLS}rAANAT counterpart, and does not appear to be an efficacious N-end rule substrate. Together, these and related results (Figs. 2-5) indicated both a major involvement of the N-end rule pathway in the control of rodent AANATs and substantial differences in the regulation of rodent and human AANATs that stem from differences in their N-terminal sequences.

The observed co-targeting of the two “complementary” forms of the rat ^{MLS}rAANAT by the two “complementary” branches of the N-end rule pathway (Figs. 2 and 3) is in agreement with the earlier finding that the N-terminal region of ^{MLS}rAANAT, and specifically its Leu-2 residue, are an important determinant of the observed metabolic instability of this AANAT (37). The cited study also suggested that either the Lys

residues of ^{MLS}rAANAT or its surface-exposed Cys residues were not essential for its proteasome-mediated degradation (37). In the present work, we compared the degradation of wild-type ^{MLS}rAANAT and its lysine-lacking ^{MLS}rAANAAT$^{Kzero}_{3ha}$ mutant using cycloheximide (CHX) chases. Our results showed that the (C-terminally triple-HA-tagged) ^{MLS}rAANAAT$^{Kzero}_{3ha}$ mutant was completely stabilized in human HEK293T cells, in contrast to instability of wild-type ^{MLS}rAANAT, indicating that at least some Lys residues of ^{MLS}rAANAT were required for its degradation (Fig. 4B).

A possible explanation of the above discrepancy stems from different choices of C-terminal epitope tags in the earlier and the present study. Specifically, Huang et al. (37) employed ^{MLS}rAANAT C-terminally tagged with c-Myc epitope (EQKLISEEDL), which contains a Lys residue. In the present work, we used a triple-HA tag, which lacks Lys residues (the sequence of single HA is YPYDVPDYA). Previous analyses of N-degrons have shown that the targeting apparatus of the Arg/N-end rule pathway selects a substrate's internal lysine as a polyubiquitylation site through a process in which different (spatially competing, conformationally mobile) Lys residues can be chosen stochastically, depending on their ability to become transiently close to the substrate-bound E2-E3 complex (41,51,95). Thus, the presence of a Lys residue in the C-terminal c-Myc epitope of the earlier study (37) might have rendered the otherwise lysine-lacking ^{MLS}rAANAT mutant still targetable by the N-end rule pathway. Indeed and tellingly, our initial experiments with lysine-lacking mutants of ^{MLS}rAANAT have employed, inadvertently, a lysine-containing C-terminal FLAG tag, instead of the lysine-lacking HA tag. Remarkably, the otherwise lysine-lacking but FLAG-tagged ^{MLS}rAANAAT$^{Kzero}_{3f}$ mutant was found to be unstable and ubiquitylated (B.W. and A.V., unpublished data), in

contrast to the actually (completely) lysine-lacking, HA-tagged ^{MLS}rAANAAT$^{Kzero}_{3ha}$ mutant (Fig. 4), in agreement with the above tag-based explanation.

In addition to polyubiquitylation and degradation, AANAT is also regulated through a site-specific phosphorylation, a modification that stimulates the binding of AANAT to at least one of the 14-3-3 chaperone-like proteins (they are encoded by a family of several *14-3-3* genes in mammals). This interaction both protects (phosphorylated) AANAT from degradation and enhances its catalytic activity. Rat, human and other AANATs contain two previously mapped phosphorylation sites, Thr-29 and Ser-203 (numbers refer to the sequence of rat ^{MLS}rAANAT). Phosphorylation at Thr-29 is conserved in all examined mammalian AANATs (31). This phosphorylation, by protein kinase A (PKA) and/or by protein kinase C (PKC), augments the interaction between AANAT and 14-3-3 both *in vitro* and *in vivo* (1,5,37). By expressing AANATs not only in mammalian cells but also in the heterologous setting of *S. cerevisiae*, we could analyze the degradation of AANATs in the possible absence of phosphorylation at Thr-29 (or its equivalent) and also in the possible absence of AANAT interactions with 14-3-3 proteins. (While *S. cerevisiae* has two 14-3-3 proteins (96), there is no evidence, so far, that bears on their binding, *in vivo*, to a mammalian AANAT.)

Our attempts to observe a phosphorylation-induced *in vivo* binding of the wild-type rat ^{MLS}rAANAT either to an endogenous yeast 14-3-3 protein or to the (overexpressed) human 14-3-3¶ protein were unsuccessful so far, despite a variety of tried approaches (B.W. and A.V., unpublished data). Nevertheless, the likely conceptual benefit of achieving a phosphorylation-inducible interaction between a mammalian AANAT and a specific 14-3-3 protein, in a setting of greatly reduced complexity of 14-3-

3s (in comparison to the large set of their mammalian counterparts), justifies further work in this direction. 14-3-3 proteins are abundant, broadly expressed, multifunctional chaperone-like proteins that modulate catalytic activities of enzymes to which 14-3-3s bind, and also regulate protein-protein interactions and subcellular targeting of specific proteins (97-99). It is possible, indeed likely, that 14-3-3 proteins may bind, in a regulated manner, not only to, for example, rat ^{MLS}rAANAT but to other N-end rule substrates as well, thereby controlling the rates of degradation of such substrates in ways that remain to be explored.

Recently, the naturally occurring mutant human protein RGS2^{Q2L} was shown to be a conditionally short-lived substrate of both the Ac/N-end rule pathway and the Arg/N-end rule pathway (60). This clinically-relevant mutant of RGS2 (a negative regulator of specific G proteins) was identified in a cohort of hypertensive Japanese patients (100) (see also the Introduction). In the present study, we demonstrated that the otherwise unrelated rat ^{MLS}rAANAT protein, whose N-terminal Met-Leu sequence is identical to that of RGS2^{Q2L}, is also targeted for degradation by both branches of the N-end rule pathway (Figs. 1 and 2). These results further expanded the set of identified mammalian N-end rule substrates of the Met-¦ type (N-terminal Met followed by a bulky hydrophobic residue; Fig. 1A).

Interestingly, RGS2, an N-end rule substrate (60), has been identified as a physiologically relevant inhibitor of the AANAT-dependent melatonin production (101). Light-stimulated, norepinephrine-induced increases in the levels and activity of AANAT are mediated by G-protein-coupled adrenergic receptors, whose activation increases intracellular cAMP, and subsequently the levels of AANAT as well, at least in part

through an increased phosphorylation of AANAT and the resulting induction of its binding to protective 14-3-3 proteins (5,28). While acting to increase the levels and activity of AANAT (which is required for the synthesis of NAS and melatonin), cAMP also acts as a negative-feedback inhibitor of melatonin production through up-regulation of *Rgs2* transcription. The resulting increase in RGS2 inhibits the activity of G proteins that are coupled to the norepinephrine-responsive receptor, a negative-feedback circuit that acts to decrease cAMP levels (101). AANAT results of the present study, together with the earlier evidence that RGS2 is both a conditionally short-lived N-end rule substrate and an (indirect) down-regulator of AANAT (60,101), indicate that the N-end rule pathway is a pleiotropic controller of circuits that either directly or indirectly modulate AANAT.

Previous studies of AANAT demonstrated that its transcriptional regulation can be different in different species, but it has been conjectured that mechanisms and regulation of AANAT degradation would be more conserved in evolution (11,34-36). As shown in the present study, the rates of degradation of the human and rat AANAT proteins are quite different. Moreover, whereas rat ^{MLS}rAANAT is destroyed by the N-end rule pathway (Figs. 2 and 4), human $^{(M)ST}$hAANAT is substantially resistant to this pathway (Figs. 3 and 5). Nevertheless, if the identification of RGS2 as a feedback inhibitor of AANAT expression in the rat pineal gland (101) could also be shown to extend to human RGS2, the N-end rule pathway would be a regulator of AANAT in humans as well, albeit an indirect one.

An earlier study of human $^{(M)ST}$hAANAT described its apparent degradation by the proteasome, but did not detect its ubiquitylation (33). In that study, the instability of

human $^{(M)ST}$hAANAT was inferred from increases in its steady-state levels in the presence of a proteasome inhibitor. In the present work, we employed, in particular, CHX-chases to show that while human $^{(M)ST}$hAANAT is degraded by the proteasome, it is substantially longer-lived than rat ^{MLS}rAANAT under the same conditions, either in mammalian cells or in *S. cerevisiae* (Figs. 2-5).

In sum, our results indicate both a major involvement of the N-end rule pathway in the control of rodent AANATs and substantial differences in the regulation of rodent and human AANATs that stem from differences in their N-terminal regions (Fig. 1C). What is the cause of the remarkable variability of N-terminal sequences of AANATs during mammalian (and, more generally, animal) evolution? Given the level of extant variability (Fig. 1C), it is nearly certain that a substantial fraction of these changes resulted from a quasi-neutral genetic drift (102). At the same time, the wide and biologically relevant variation in detailed circadian rhythms among different animal species (owing to specific ecological and physiological adaptations) and the role of AANAT in these rhythms suggest that the seeming randomness of AANATs' N-terminal sequences (Fig. 1C) may be obscuring, thus far, a set of adaptive (selected) changes in these sequences over evolutionary timescales.

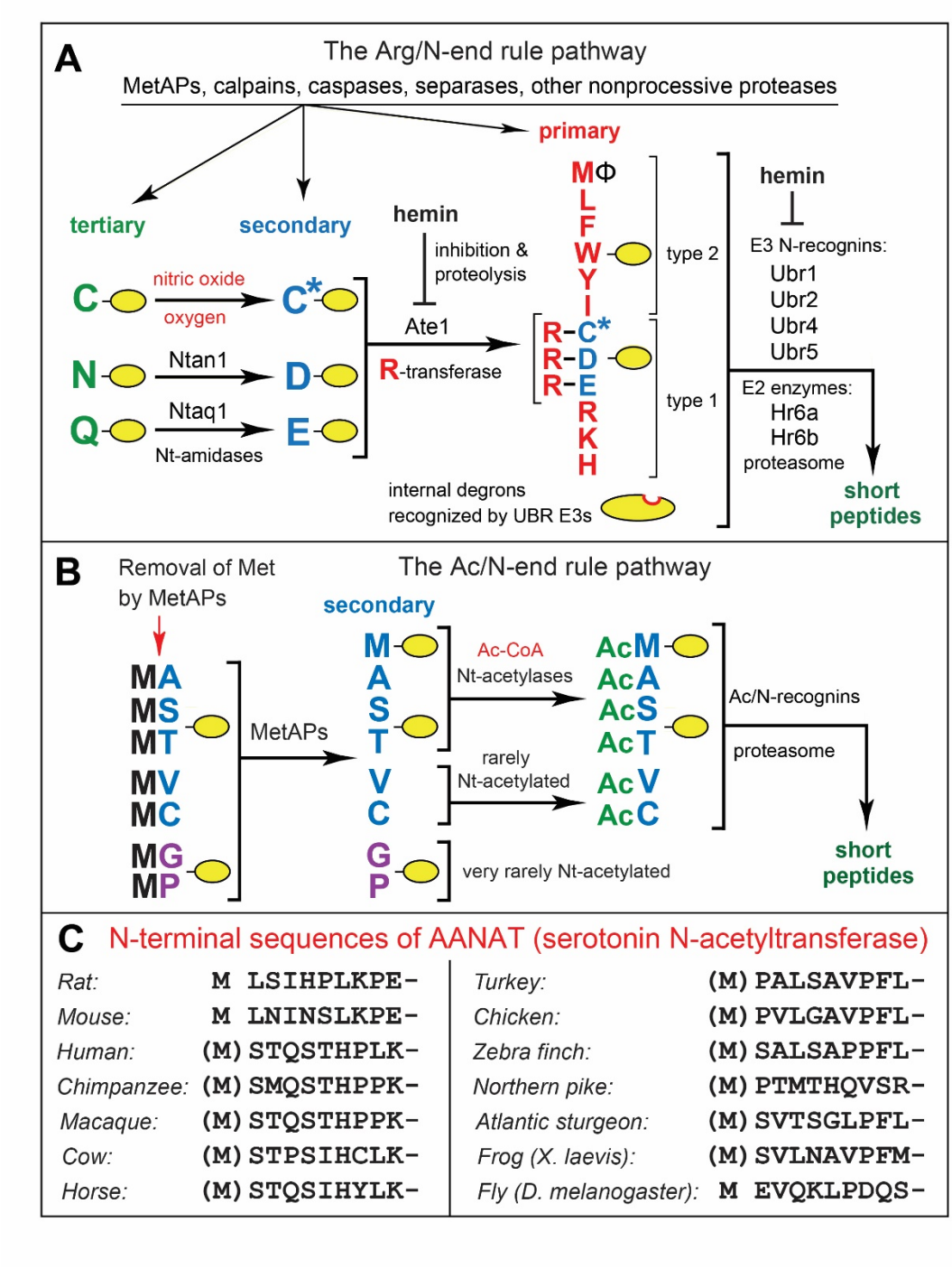


Fig. 2.1. The mammalian N-end rule pathway and N-terminal regions of AANAT enzymes. See the Introduction for references and descriptions of the N-end rule pathway's mechanistic aspects and biological functions. Amino acid residues are denoted by single-letter abbreviations. *A*, the mammalian Arg/N-end rule pathway. It targets proteins for degradation through their specific unacetylated N-terminal residues. A yellow oval denotes the rest of a protein substrate. "Primary", "secondary", and "tertiary" refer to mechanistically distinct classes of destabilizing N-terminal residues. Ntan1 and

Ntaq1 are N-terminal amidases (Nt-amidases) that convert, respectively, the tertiary destabilizing N-terminal residues Asn and Gln to Asp and Glu. The Ate1 arginyltransferase (R-transferase) conjugates Arg, a primary destabilizing residue, to N-terminal Asp, Glu and (oxidized) Cys. “Type 1” and “type 2” refer to two sets of primary destabilizing N-terminal residues, basic (Arg, Lys, His) and bulky hydrophobic (Leu, Phe, Trp, Tyr, Ile, and Met, if the latter is followed by a bulky hydrophobic residue ($)), respectively. These sets of N-terminal residues are recognized by two distinct substrate-binding sites of N-recognins, the pathway’s E3 ubiquitin ligases, whose (possibly incomplete) list includes Ubr1, Ubr2, Ubr4, and Ubr5. *B*, the mammalian Ac/N-end rule pathway. It targets proteins through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues. Red arrow on the left indicates the cotranslational removal of the N-terminal Met residue by Met-aminopeptidases (MetAPs). N-terminal Met is retained if a residue at position 2 is larger than Val. *C*, the first 10 amino acid residues of animal AANAT enzymes, in vertebrates and an invertebrate such as *Drosophila melanogaster*. See the Introduction for descriptions of AANATs.

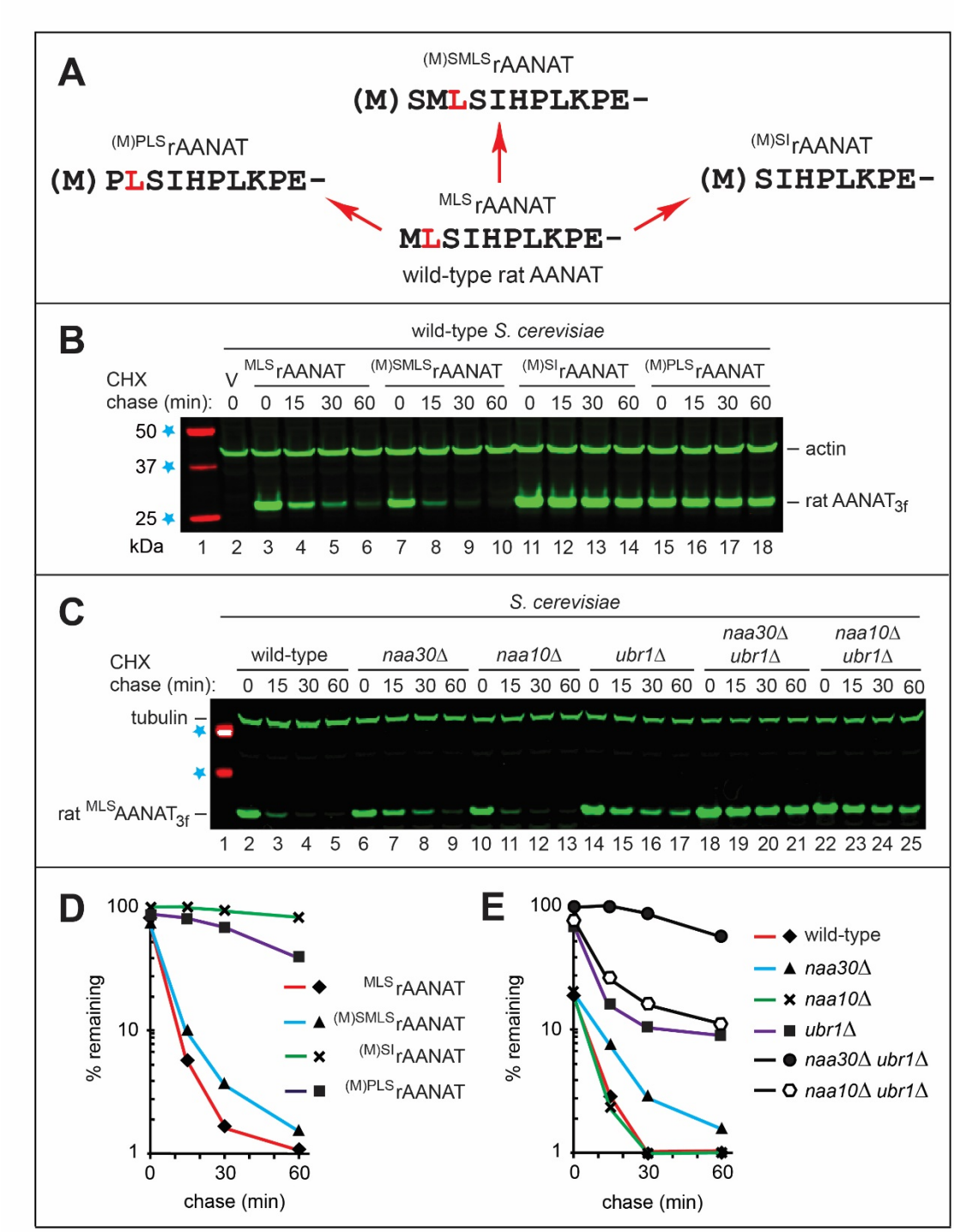


Fig 2.2. The wild-type rat ^{MLS}rAANAT as a substrate of both branches of the N-end rule pathway. *A*, the wild-type rat ^{MLS}rAANAT and its mutants that were constructed and analyzed in the present study. Amino acid residues are denoted by single-letter abbreviations. The Leu (L) residue, at position 2 in wild-type ^{MLS}rAANAT, is highlighted in red to make it easier to follow sequence alterations in specific mutants. *B*,

immunoblotting with antibody to actin was used as a loading control. Lane 1, fluorescently labeled molecular mass markers (Li-Cor), with their masses, in kDa, indicated on the left. Lane 2, wild-type *S. cerevisiae* were transformed with vector (V) alone (control). Cycloheximide (CHX)-chases were performed at 30°C for the indicated times in wild-type *S. cerevisiae* with the wild-type rat ^{MLS}rAANAT$_{3f}$ (lanes 3-6), and with its mutants $^{(M)SMLS}$rAANAT$_{3f}$ (lanes 7-10), $^{(M)SI}$rAANAT$_{3f}$ (lanes 11-14), and $^{(M)PLS}$rAANAT$_{3f}$ (lanes 15-18). The bands of AANAT$_{3f}$ and actin are indicated on the right. *C*, immunoblotting with antibody to tubulin was used as a loading control. Lane 1, 37 kDa and 50 kDa molecular mass markers (see lane 1 in *B*). CHX-chases with the wild-type rat ^{MLS}rAANAT$_{3f}$ in wild-type *S. cerevisiae* (lanes 2-5) and its mutants *naa30*" (lanes 6-9), *naa10*" (lanes 10-13), *ubr1*" (lanes 14-17), *naa30*" *ubr1*" (lanes 18-21), and *naa10*" *ubr1*" (lanes 22-25). The bands of ^{MLS}rAANAT$_{3f}$ and tubulin are indicated on the right. *D*, quantification of data in *B*. *E*, quantification of data in *C*.

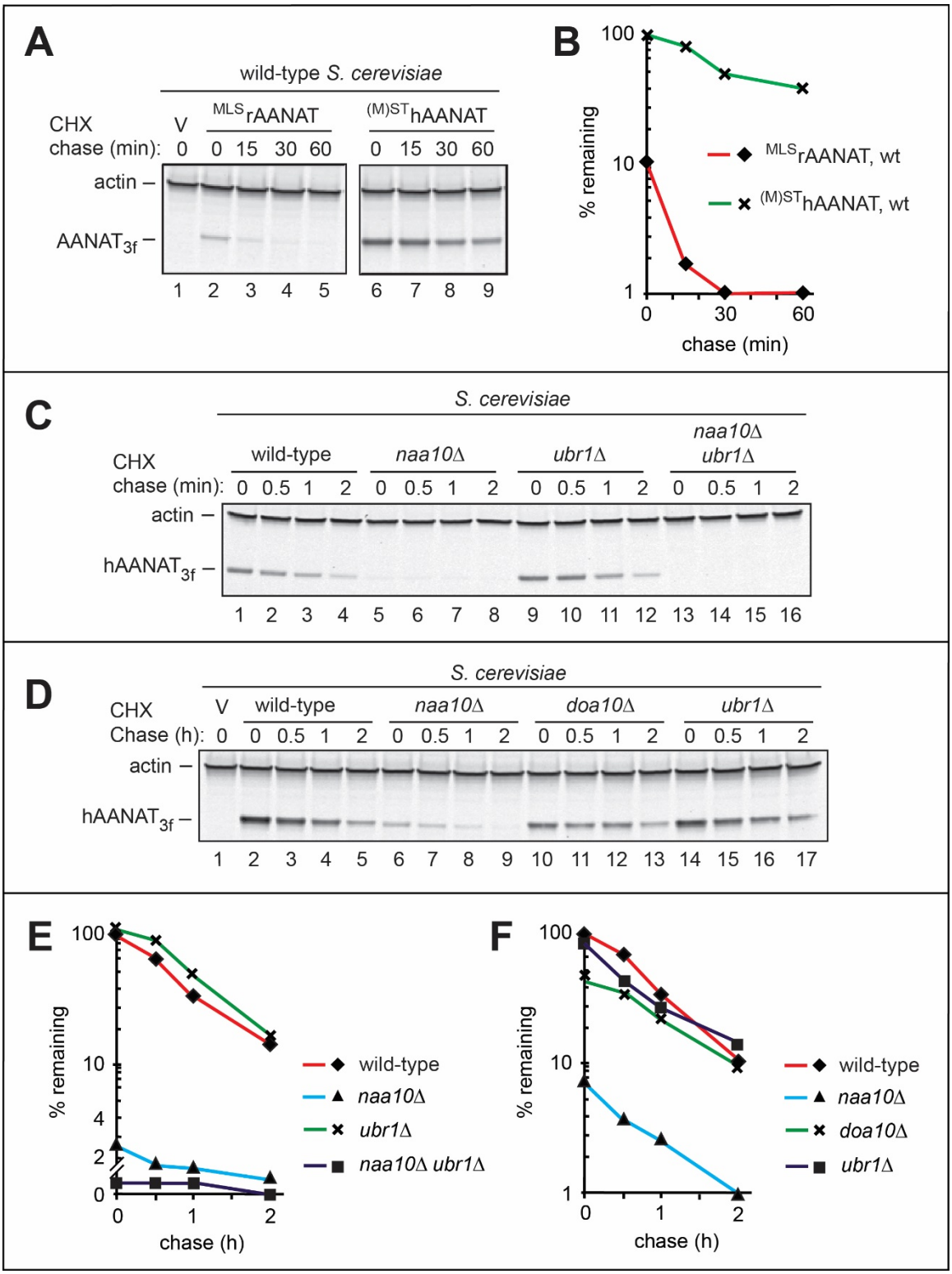


Fig. 2.3. Degradation assays with the wild-type human $^{(M)ST}$hAANAT in *S. cerevisiae*. *A*, Comparison of degradation rates of the wild-type rat ^{MLS}rAANAT$_{3f}$ and the wild-type human $^{(M)ST}$hAANAT$_{3f}$ in *S. cerevisiae*. Immunoblotting with antibody to

actin was used as a loading control. Lane 1, wild-type *S. cerevisiae* were transformed with vector (V) alone (control). CHX-chases were performed at 30°C for the indicated times in wild-type *S. cerevisiae* with the wild-type rat ^{MLS}rAANAT$_{3f}$ (lanes 2-5) and with the wild-type human $^{(M)ST}$hAANAT$_{3f}$ (lanes 6-9). The bands of AANAT and actin are indicated on the left. *B*, quantification of data in *A*. *C*, CHX-chases with the wild-type human $^{(M)ST}$hAANAT$_{3f}$ in wild-type *S. cerevisiae* (lanes 1-4) and its mutants *naa10*" (lanes 5-8), *ubr1*" (lanes 9-12), and *naa10*" *ubr1*" (lanes 13-16). The bands of human $^{(M)ST}$hAANAT$_{3f}$ and actin are indicated on the left. *D*, Lane 1, wild-type *S. cerevisiae* were transformed with vector (V) alone (control). CHX-chases with the wild-type human $^{(M)ST}$hAANAT$_{3f}$ in wild-type *S. cerevisiae* (lanes 2-5) and its mutants *naa10*" (lanes 6-9), *doa10*" (lanes 10-13), and *ubr1*" (lanes 14-17). The bands of human $^{(M)ST}$hAANAT$_{3f}$ and actin are indicated on the left. *E*, quantification of data in *C*. *F*, quantification of data in *D*.

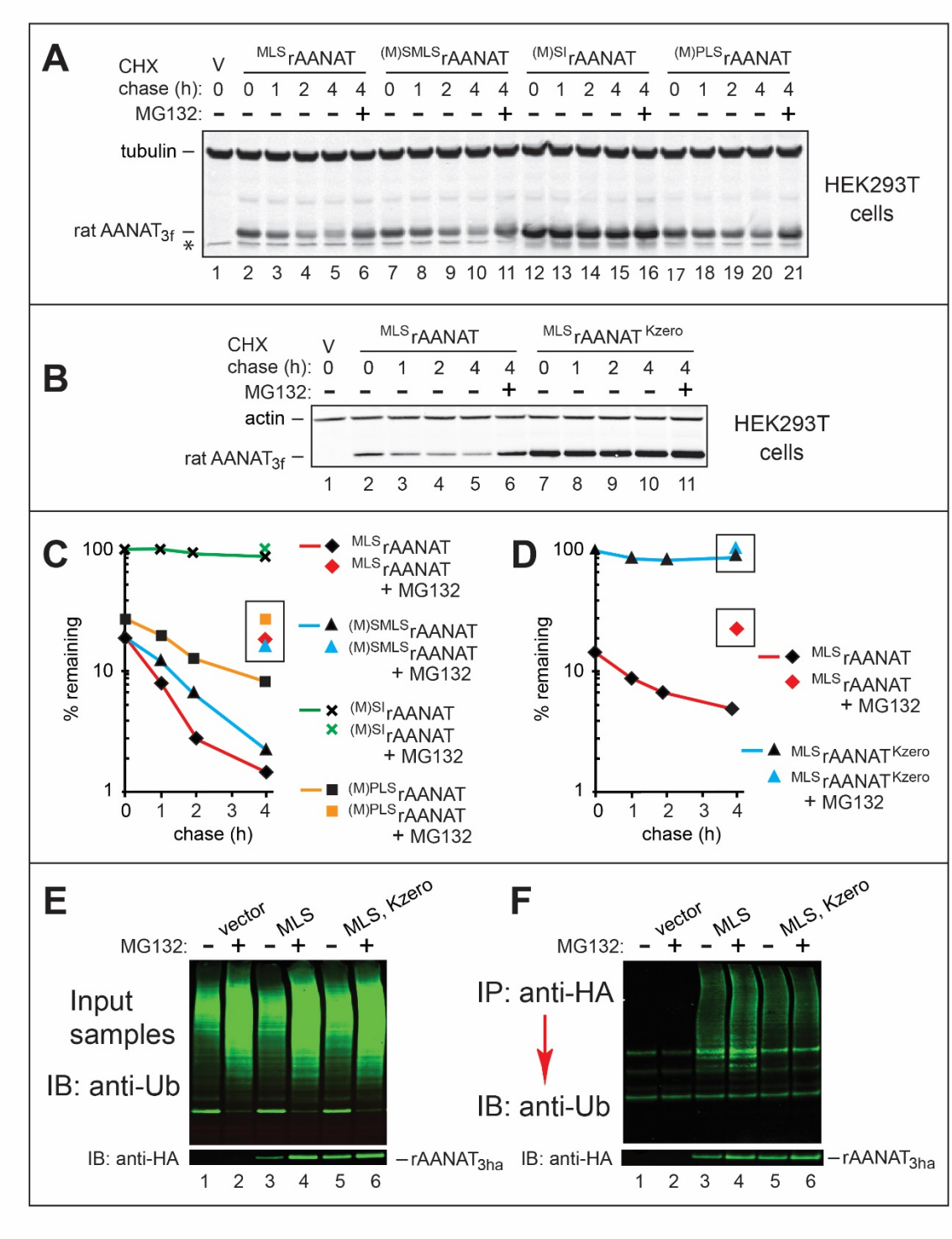


Fig. 2.4. Degradation and ubiquitylation assays with rat and human AANATs in human HEK293T cells. *A*, Lane 1, HEK293T cells were transformed with vector (V) alone (control) bearing the (weakened) P_{CMVt1} promoter (see Results and Fig. A1.S3). Immunoblotting with antibody to tubulin was used as a loading control. CHX-chases were performed at 37°C for the indicated times in HEK293T cells with the wild-type rat

^{MLS}rAANAT$_{3f}$ (lanes 2-5), and with its mutants $^{(M)SMLS}$rAANAT$_{3f}$ (lanes 7-10), $^{(M)SI}$rAANAT$_{3f}$ (lanes 12-15), and $^{(M)PLS}$rAANAT$_{3f}$ (lanes 17-20). Lanes 6, 11, 16, and 21 are the same as lanes 5, 10, 15, and 20 except that the MG132 proteasome inhibitor was present during a 4-h CHX-chase, in each case (see Experimental Procedures). The bands of AANAT$_{3f}$ and tubulin are indicated on the right. An asterisk on the left denotes a band of protein that crossreacted with anti-flag antibody (the band is also present in lane 1, the vector-only control). *B*, Lane 1, the same as the control in lane 1 of *A*. CHX-chases were performed at 37°C for the indicated times in HEK293T cells with the wild-type rat ^{MLS}rAANAT$_{3ha}$ (lanes 2-5) and with its lysine-lacking mutant ^{MLS}rAANAAT$_{3ha}^{Kzero}$(lanes 7-10). Lanes 6 and 11, same as lanes 5 and 10 except that MG132 was present during a 4-h CHX-chase, in each case. *C*, quantification of data in *A*. *D*, quantification of data in *B*. Symbols shown separately and framed in black squares or rectangles correspond to altered levels of test proteins in the presence of MG132. *E*, immunoblot analyses, using anti-Ub antibody, of total, input (before immunoprecipitation) extracts from HEK293T cells that had been transfected either with vector alone (lanes 1 and 2), or with plasmid expressing the wild-type rat ^{MLS}rAANAT$_{3ha}$ (lanes 3 and 4), or with plasmid expressing its lysine-lacking counterpart ^{MLS}rAANAAT$_{3ha}^{Kzero}$(lanes 5 and 6). The MG132 proteasome inhibitor was either omitted (lanes 1, 3, and 5) or added to HEK293T cells (to the final concentration of 10 μM) 6 h before preparation of extracts (lanes 2, 4, and 6). The lower panel shows the results of detecting non-ubiquitylated ^{MLS}rAANAT$_{3ha}$ (absent in the control lanes 1 and 2) with anti-ha antibody. *F*, same as in *E*, but immunoblot analyses (using anti-Ub antibody) of the input samples after their immunoprecipitation with anti-ha antibody (followed by SDS-PAGE), to isolate and detect ubiquitylated AANAT$_{3ha}$ species.

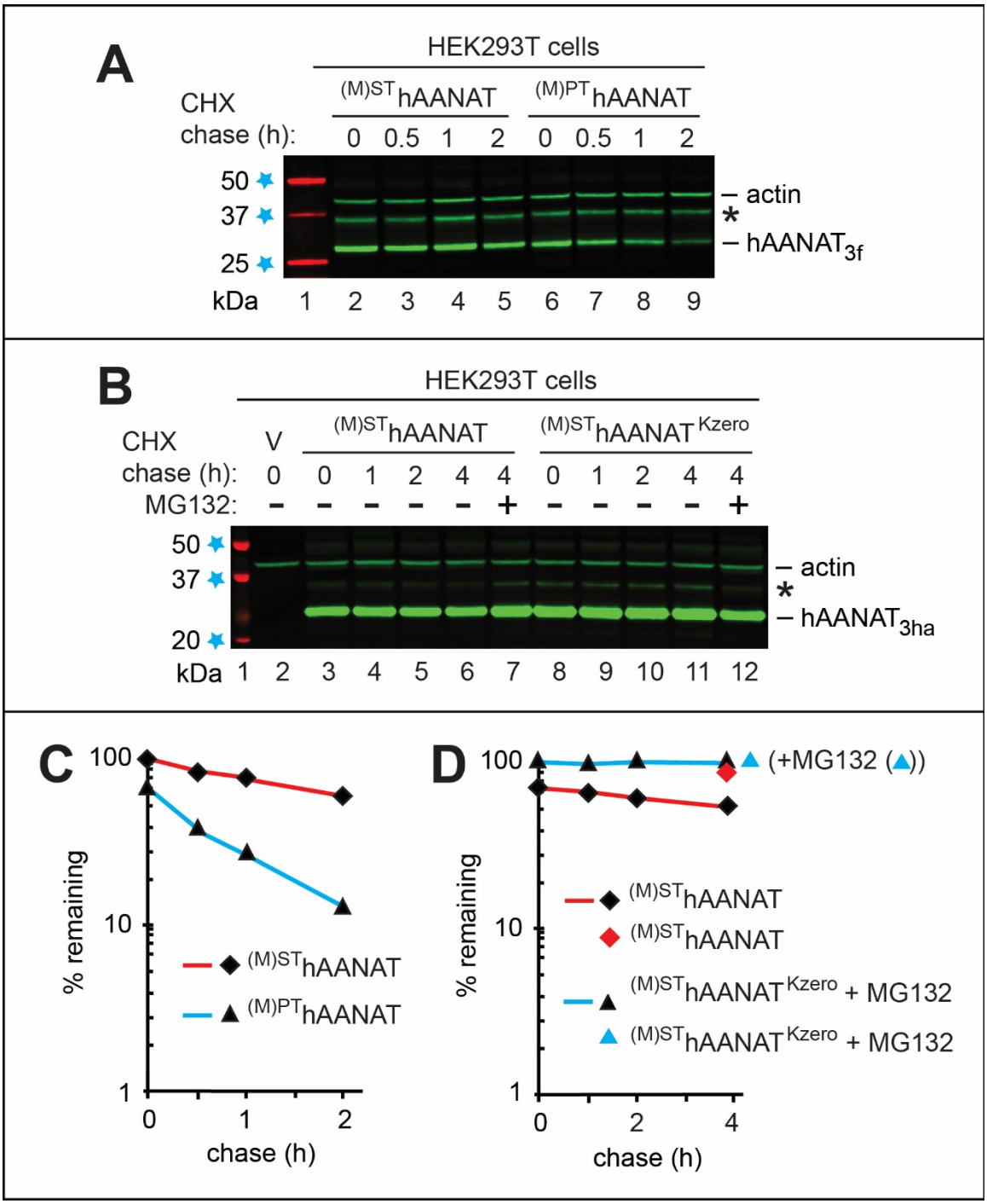


Fig. 2.5. Degradation of the wild-type human $^{(M)ST}$hAANAT and its mutants in human HEK293T cells. The test proteins were expressed in HEK293T cells using a plasmid bearing the (weakened) P_{CMVt1} promoter (see Results and Fig. A1.S3). *A*, Lane 1, fluorescently labeled molecular mass markers (Li-Cor), with their masses, in kDa, indicated on the left. CHX-chases were performed at 37°C for the indicated times in HEK293T cells with the wild-type human $^{(M)ST}$hAANAT$_{3f}$ (lanes 2-5) and with its mutant

$^{(M)PT}$hAANAT$_{3f}$ (lanes 6-9). *B*, Lane 1, fluorescently labeled molecular mass markers, with their masses, in kDa, indicated on the left. Lane 2, HEK293T cells were transformed with vector (V) alone (control). CHX-chases were performed at 37°C for the indicated times in HEK293T cells with the wild-type human $^{(M)ST}$hAANAT$_{3ha}$ (lanes 3-6) and with its lysine-lacking mutant $^{(M)ST}$hAANAAT$_{3ha}^{Kzero}$ (lanes 8-11). Lanes 7 and 12, same as lanes 6 and 11, except that the MG132 proteasome inhibitor was present during a 4-h CHX-chase, in each case (see Experimental Procedures). The bands of actin and AANAT are indicated on the right. Asterisks on the right in *A* and *B* denote a protein band (possibly the same protein and possibly a derivative of AANAT) that reacted with both anti-flag antibody in *A* and anti-ha antibody in *B*. *C*, quantification of data in *A*. *D*, quantification of data in *B*.

REFERENCES

1. Klein, D. C. (2006) Evolution of the vertebrate pineal gland: the AANAT hypothesis. *Chronobiol. Int.* **23**, 5-20
2. Coon, S. L., and Klein, D. C. (2006) Evolution of arylalkylamine N-acetyltransferase: emergence and divergence. *Mol. Cell. Endocrinol.* **252**, 2-10
3. Li, J., You, X., Bian, C., Yu, H., Coon, S. L., and Shi, Q. (2016) Molecular Evolution of Aralkylamine N-Acetyltransferase in Fish: A Genomic Survey. *Int. J. Mol. Sci.* **17**, E51
4. Borjigin, J., Zhang, L. S., and Calinescu, A. A. (2012) Circadian regulation of pineal gland rhythmicity. *Mol. Cell. Endocrinol.* **349**, 13-19
5. Choi, B. H., Chae, H. D., Park, T. J., Oh, J., Lim, J., Kang, S. S., Ha, H., and Kim, K. T. (2004) Protein kinase C regulates the activity and stability of serotonin N-acetyltransferase. *J. Neurochem.* **90**, 442-454
6. Szewczuk, L. M., Tarrant, M. K., Sample, V., Drury, W. J., 3rd, Zhang, J., and Cole, P. A. (2008) Analysis of serotonin N-acetyltransferase regulation in vitro and in live cells using protein semisynthesis. *Biochemistry* **47**, 10407-10419
7. Obsil, T., Ghirlando, R., Klein, D. C., Ganguly, S., and Dyda, F. (2001) Crystal structure of the 14-3-3zeta:serotonin N-acetyltransferase complex. a role for scaffolding in enzyme regulation. *Cell* **105**, 257-267
8. Hickman, A. B., Namboodiri, M. A., Klein, D. C., and Dyda, F. (1999) The structural basis of ordered substrate binding by serotonin N-acetyltransferase: enzyme complex at 1.8 A resolution with a bisubstrate analog. *Cell* **97**, 361-369

9. Hickman, A. B., Klein, D. C., and Dyda, F. (1999) Melatonin biosynthesis: the structure of serotonin N-acetyltransferase at 2.5 A resolution suggests a catalytic mechanism. *Mol. Cell* **3**, 23-32

10. Borjigin, J., Wang, M. M., and Snyder, S. H. (1995) Diurnal variation in mRNA encoding serotonin N-acetyltransferase in pineal gland. *Nature* **378**, 783-785

11. Coon, S. L., Roseboom, P. H., Baler, R., Weller, J. L., Namboodiri, M. A., Koonin, E. V., and Klein, D. C. (1995) Pineal serotonin N-acetyltransferase: expression cloning and molecular analysis. *Science* **270**, 1681-1683

12. Tosini, G., Ye, K., and Iuvone, P. M. (2012) N-acetylserotonin: neuroprotection, neurogenesis, and the sleepy brain. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry* **18**, 645-653

13. Jang, S. W., Liu, X., Pradoldej, S., Tosini, G., Chang, Q., Iuvone, P. M., and Ye, K. (2010) N-acetylserotonin activates TrkB receptor in a circadian rhythm. *Proc. Natl. Acad. Sci. USA* **107**, 3876-3881

14. Sompol, P., Liu, X., Baba, K., Paul, K. N., Tosini, G., Iuvone, P. M., and Ye, K. (2011) N-acetylserotonin promotes hippocampal neuroprogenitor cell proliferation in sleep-deprived mice. *Proc. Natl. Acad. Sci. USA* **108**, 8844-8849

15. Schippers, K. J., and Nichols, S. A. (2014) Deep, dark secrets of melatonin in animal evolution. *Cell* **159**, 9-10

16. Zhdanova, I. V., Wang, S. Y., Leclair, O. U., and Danilova, N. P. (2001) Melatonin promotes sleep-like state in zebrafish. *Brain Res.* **903**, 263-268

17. Gandhi, A. V., Mosser, E. A., Oikonomou, G., and Prober, D. A. (2015) Melatonin is required for the circadian regulation of sleep. *Neuron* **85**, 1193-1199

18. Gonzalez, S., Moreno-Delgado, D., Moreno, E., Perez-Capote, K., Franco, R., Mallol, J., Cortes, A., Casado, V., Lluis, C., Ortiz, J., Ferre, S., Canela, E., and McCormick, P. J. (2012) Circadian-related heteromerization of adrenergic and dopamine D(4) receptors modulates melatonin synthesis and release in the pineal gland. *PLoS Biol.* **10**, e1001347

19. Rawashdeh, O., and Maronde, E. (2012) The hormonal Zeitgeber melatonin: role as a circadian modulator in memory processing. *Front. Mol. Neurosci.* **5**, 27

20. Arendt, J. (2006) Melatonin and human rhythms. *Chronobiol. Int.* **23**, 21-37

21. Liu, T., and Borjigin, J. (2005) N-acetyltransferase is not the rate-limiting enzyme of melatonin synthesis at night. *J. Pineal. Res.* **39**, 91-96

22. Tu, B. P., and McKnight, S. L. (2006) Metabolic cycles as an underlying basis of biological oscillations. *Nat. Rev. Mol. Cell Biol.* **7**, 696-701

23. Masri, S., and Sassone-Corsi, P. (2013) The circadian clock: a framework linking metabolism, epigenetics and neuronal function. *Nat. Rev. Neurosci.* **14**, 69-75

24. Edgar, R. S., Green, E. W., Zhao, Y., van Ooijen, G., Olmedo, M., Qin, X., Xu, Y., Pan, M., Valekunja, U. K., Feeney, K. A., Maywood, E. S., Hastings, M. H., Baliga, N. S., Merrow, M., Millar, A. J., Johnson, C. H., Kyriacou, C. P., O'Neill, J. S., and Reddy, A. B. (2012) Peroxiredoxins are conserved markers of circadian rhythms. *Nature* **485**, 459-464

25. Borjigin, J., Li, X., and Snyder, S. H. (1999) The pineal gland and melatonin: molecular and pharmacologic regulation. *Annu. Rev. Pharmacol. Toxicol.* **39**, 53-65

26. Stehle, J. H., Saade, A., Rawashdeh, O., Ackermann, K., Jilg, A., Sebesteny, T., and Maronde, E. (2011) A survey of molecular details in the human pineal gland in the light of phylogeny, structure, function and chronobiological diseases. *J. Pineal Res.* **51**, 17-43

27. Mohawk, J. A., and Takahashi, J. S. (2011) Cell autonomy and synchrony of suprachiasmatic nucleus circadian oscillators. *Trends Neurosci.* **34**, 349-358

28. Ganguly, S., Gastel, J. A., Weller, J. L., Schwartz, C., Jaffe, H., Namboodiri, M. A., Coon, S. L., Hickman, A. B., Rollag, M., Obsil, T., Beauverger, P., Ferry, G., Boutin, J. A., and Klein, D. C. (2001) Role of a pineal cAMP-operated arylalkylamine N-acetyltransferase/14-3-3-binding switch in melatonin synthesis. *Proc. Natl. Acad. Sci. USA* **98**, 8083-8088

29. Mukherjee, S., and Maitra, S. K. (2015) Gut melatonin in vertebrates: chronobiology and physiology. *Front. Endocrinol.* **6**, 112

30. Hardeland, R. (2010) Melatonin metabolism in the central nervous system. *Curr. Neuropharm.* **8**, 168-181

31. Chattoraj, A., Liu, T., Zhang, L. S., Huang, Z., and Borjigin, J. (2009) Melatonin formation in mammals: in vivo perspectives. *Rev. Endocr. Metab. Disord.* **10**, 237-243

32. Schomerus, C., Laedtke, E., Olcese, J., Weller, J. L., Klein, D. C., and Korf, H. W. (2002) Signal transduction and regulation of melatonin synthesis in bovine pinealocytes: impact of adrenergic, peptidergic and cholinergic stimuli. *Cell Tissue Res.* **309**, 417-428

33. Schomerus, C., Korf, H. W., Laedtke, E., Weller, J. L., and Klein, D. C. (2000) Selective adrenergic/cyclic AMP-dependent switch-off of proteasomal proteolysis alone switches on neural signal transduction: an example from the pineal gland. *J. Neurochem.* **75**, 2123-2132

34. Bernard, M., Iuvone, P. M., Cassone, V. M., Roseboom, P. H., Coon, S. L., and Klein, D. C. (1997) Avian melatonin synthesis: photic and circadian regulation of serotonin N-acetyltransferase mRNA in the chicken pineal gland and retina. *J. Neurochem.* **68**, 213-224

35. Gastel, J. A., Roseboom, P. H., Rinaldi, P. A., Weller, J. L., and Klein, D. C. (1998) Melatonin production: proteasomal proteolysis in serotonin N-acetyltransferase regulation. *Science* **279**, 1358-1360

36. Roseboom, P. H., Coon, S. L., Baler, R., McCune, S. K., Weller, J. L., and Klein, D. C. (1996) Melatonin synthesis: analysis of the more than 150-fold nocturnal increase in serotonin N-acetyltransferase messenger ribonucleic acid in the rat pineal gland. *Endocrinology* **137**, 3033-3045

37. Huang, Z., Liu, T., and Borjigin, J. (2010) N-terminal residues regulate proteasomal degradation of AANAT. *J. Pineal Res.* **48**, 290-296

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

39. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. *Science* **327**, 973-977

40. Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A., and Hwang, C. S. (2014) The N-terminal methionine of cellular proteins as a degradation signal. *Cell* **156**, 158-169

41. Varshavsky, A. (2011) The N-end rule pathway and regulation by proteolysis. *Prot. Sci.* **20**, 1298-1345

42. Gibbs, D. J., Bacardit, J., Bachmair, A., and Holdsworth, M. J. (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. *Trends Cell Biol.* **24**, 603-611

43. Tasaki, T. S., Sriram, S. M., Park, K. S., and Kwon, Y. T. (2012) The N-end rule pathway. *Annu. Rev. Biochem.* **81**, 261-289

44. Dougan, D. A., Micevski, D., and Truscott, K. N. (2011) The N-end rule pathway: from recognition by N-recognins to destruction by AAA+ proteases. *Biochim. Biophys. Acta* **1823**, 83-91

45. Varshavsky, A. (2008) Discovery of cellular regulation by protein degradation ("Reflections" article). *J. Biol. Chem.* **283**, 34469-34489

46. Lee, K. E., Heo, J. E., Kim, J. M., and Hwang, C. S. (2016) N-terminal acetylation-targeted N-end rule proteolytic system: the Ac/N-end rule pathway. *Mol. Cells* **39**, 169-178

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

48. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2009) Two proteolytic pathways regulate DNA repair by cotargeting the Mgt1 alkylguanine transferase. *Proc. Natl. Acad. Sci. USA* **106**, 2142-2147

49. Heck, J. W., Cheung, S. K., and Hampton, R. Y. (2010) Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. *Proc. Natl. Acad. Sci. USA* **107**, 1106-1111

50. Eisele, F., and Wolf, D. H. (2008) Degradation of misfolded proteins in the cytoplasm by the ubiquitin ligase Ubr1. *FEBS Lett.* **582**, 4143-4146

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

52. Inobe, T., and Matouschek, A. (2014) Paradigms of protein degradation by the proteasome. *Curr. Op. Struct. Biol.* **24**, 156-164

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

54. Mogk, A., Schmidt, R., and Bukau, B. (2007) The N-end rule pathway of regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends Cell Biol.* **17**, 165-172

55. Rivera-Rivera, I., Román-Hernández, G., Sauer, R. T., and Baker, T. A. (2014) Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. *Proc. Natl. Acad. Sci. USA* **111**, E3853-E3859.

56. Piatkov, K. I., Vu, T. T. M. H., C.-S., and Varshavsky, A. (2015) Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. *Microbial Cell* **2**, 376-393

57. Humbard, M. A., Surkov, S., De Donatis, G. M., Jenkins, L., and Maurizi, M. R. (2013) The N-degradome of Escherichia coli: limited proteolysis in vivo

generates a large pool of proteins bearing N-degrons. *J. Biol. Chem.* **288**, 28913-28924

58. Graciet, E., Hu, R. G., Piatkov, K., Rhee, J. H., Schwarz, E. M., and Varshavsky, A. (2006) Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. *Proc. Natl. Acad. Sci. USA* **103**, 3078-3083

59. Shemorry, A., Hwang, C. S., and Varshavsky, A. (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. *Mol. Cell* **50**, 540-551

60. Park, S. E., Kim, J. M., Seok, O. H., Cho, H., Wadas, B., Kim, S. Y., Varshavsky, A., and Hwang, C. S. (2015) Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. *Science* **347**, 1249-1252

61. Aksnes, H., Hole, K., and Arnesen, T. (2015) Molecular, cellular, and physiological significance of N-terminal acetylation. *Int. Rev. Cell. Mol. Biol.* **316**, 267-305

62. Dorfel, M. J., and Lyon, G. J. (2015) The biological functions of Naa10 - From amino-terminal acetylation to human disease. *Gene* **567**, 103-131

63. Starheim, K. K., Gevaert, K., and Arnesen, T. (2012) Protein N-terminal acettyltransferases: when the start matters. *Trends Biochem. Sci.* **37**, 152-161

64. Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2012) The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. *Proc. Natl. Acad. Sci. USA* **109**, E1839-E1847

65. Piatkov, K. I., Oh, J.-H., Liu, Y., and Varshavsky, A. (2014) Calpain-generated natural protein pragments as short-lived substrates of the N-end rule pathway. *Proc. Natl. Acad. Sci. USA* **111**, E817-E826

66. Brower, C. S., Piatkov, K. I., and Varshavsky, A. (2013) Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. *Mol. Cell* **50**, 161-171

67. Yamano, K., and Youle, R. J. (2013) PINK1 is degraded through the N-end rule pathway. *Autophagy* **9**, 1758-1769

68. Cha-Molstad, H., Sung, K. S., Hwang, J., Kim, K. A., Yu, J. E., Yoo, Y. D., Jang, J. M., Han, D. H., Molstad, M., Kim, J. G., Lee, Y. J., Zakrzewska, A., Kim, S. H., Kim, S. T., Kim, S. Y., Lee, H. G., Soung, N. K., Ahn, J. S., Ciechanover, A., Kim, B. Y., and Kwon, Y. T. (2015) Amino-terminal arginylation targets endoplasmic reticulum chaperone BiP for autophagy through p62 binding. *Nat. Cell Biol.* **17**, 917-929

69. Wang, H., Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. *Mol. Cell* **34**, 686-695

70. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. *Nat. Cell Biol.* **12**, 1177-1185

71. Varshavsky, A. (2004) Spalog and sequelog: neutral terms for spatial and sequence similarity. *Curr. Biol.* **14**, R181-R183

72. Hassink, G., Kikkert, M., van Voorden, S., Lee, S.-J., Spaapen, R., van Laar, T., Cloleman, C. S., Bartee, E., Früh, K., Chau, V., and Wiertz, E. (2005) TEB4 is a C4HC3 RING-finger containing ubiquitin ligase of the endoplasmic reticulum. *Biochem. J.* **388**, 647-655

73. Foresti, O., Ruggiano, A., Hannibal-Bach, H. K., Ejsing, C. S., and Carvalho, P. (2013) Sterol homeostasis requires regulated degradation of squalene monooxygenase by the ubiquitin ligase Doa10/Teb4. *Elife* **2**, e00953

74. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (2010) *Current Protocols in Molecular Biology.*, Wiley-Interscience, New York

75. Guthrie, C., and Fink, G. R. (1991) Guide to Yeast Genetics and Molecular Biology. *Meth. Enzymol.* **194**, 3-933

76. Andrews, B., Boone, C., Davis, T. N., and Fields, S. (2016) *Budding Yeast (a laboratory manual).* Cold Spring Harbor Press, Cold Spring Harbor, NY

77. Gietz, R. D., and Woods, R. A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Meth. Enzymol.* **350**, 87-96

78. Huang, Z., Deng, J., and Borjigin, J. (2005) A novel H28Y mutation in LEC rats leads to decreased NAT protein stability in vivo and in vitro. *J. Pineal Res.* **39**, 84-90

79. Sikorski, R. S., and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in S. cerevisiae. *Genetics* **122**, 19-27

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

81. Kushnirov, V. V. (2000) Rapid and reliable protein extraction from yeast. *Yeast* **16**, 857-860

82. Swerdlow, P. S., Finley, D., and Varshavsky, A. (1986) Enhancement of immunoblot sensitivity by heating of hydrated filters. *Anal Biochem* **156**, 147-153

83. Xiao, Q., Zhang, F., Nacev, B. A., Liu, J. O., and Pei, D. (2010) Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. *Biochemistry* **49**, 5588-5599

84. Varland, S., Osberg, C., and Arnesen, T. (2015) N-terminal modifications of cellular proteins: The enzymes involved, their substrate specificities and biological effects. . *Proteomics* **15**, 2385-2401

85. Perrot, M., Massoni, A., and Boucherie, H. (2008) Sequence requirements for Nalpha-terminal acetylation of yeast proteins by NatA. *Yeast* **25**, 513-527

86. Lee, D. H., and Goldberg, A. L. (1996) Selective inhibitors of the proteasome-dependent and vacuolar pathways of protein degradation in Saccharomyces cerevisiae. *J. Biol. Chem.* **271**, 27280-27284

87. Fleming, J. A., Lightcap, E. S., Sadis, S., Thoroddsen, V., Bulawa, C. E., and Blackman, R. K. (2002) Complementary whole-genome technologies reveal the cellular response to proteasome inhibition by PS-341. *Proc. Natl. Acad. Sci. USA* **99**, 1461-1466

88. Lee, S. J., Liu, T., Chattoraj, A., Zhang, S. L., Wang, L., Lee, T. M., Wang, M. M., and Borjigin, J. (2009) Posttranscriptional regulation of pineal melatonin synthesis in Octodon degus. *J. Pineal Res.* **47**, 75-81

89. Falcon, J., Coon, S. L., Besseau, L., Cazamea-Catalan, D., Fuentes, M., Magnanou, E., Paulin, C. H., Boeuf, G., Sauzet, S., Jorgensen, E. H., Mazan, S., Wolf, Y. I., Koonin, E. V., Steinbach, P. J., Hyodo, S., and Klein, D. C. (2014) Drastic neofunctionalization associated with evolution of the timezyme AANAT 500 Mya. *Proc. Natl. Acad. Sci. USA* **111**, 314-319

90. Ganguly, S., Mummaneni, P., Steinbach, P. J., Klein, D. C., and Coon, S. L. (2001) Characterization of the Saccharomyces cerevisiae homolog of the melatonin rhythm enzyme arylalkylamine N-acetyltransferase. *J. Biol. Chem.* **276**, 47239-47247

91. Kim, J. M., and Hwang, C. S. (2014) Crosstalk between the Arg/N-end and Ac/N-end rule. *Cell Cycle* **13**, 1366-1367

92. Slater, M., Hartzell, D., Hartnett, J., Wheeler, S., Stecha, P., and Karassina, N. (2008) Achieve the protein expression level you need with the mammalian HaloTag 7 Flexi vectors. *Promega Notes* **100**, 16-18

93. Wu, C., and Li, S. S. (2009) CelluSpots: a reproducible means of making peptide arrays for the determination of SH2 domain binding specificity. . *Meth. Mol. Biol.* **570**, 197-202

94. Komander, D., and Rape, M. (2012) The ubiquitin code. *Annu. Rev. Biochem.* **81**, 203-229

95. Inobe, T., Fishbain, S., Prakash, S., and Matouschek, A. (2011) Defining the geometry of the two-component proteasome degron. *Nat. Chem. Biol.* **7**, 161-167

96. van Heusden, G. P., and Steensma, H. Y. (2006) Yeast 14-3-3 proteins. *Yeast* **23**, 159-171

97. Obsil, T., and Obsilova, V. (2011) Structural basis of 14-3-3 protein functions. *Sem. Cell Dev. Biol.* **22**, 663-672

98. Kakiuchi, K., Yamauchi, Y., Taoka, M., Iwago, M., Fujita, T., Ito, T., Song, S. Y., Sakai, A., Isobe, T., and Ichimura, T. (2007) Proteomic analysis of in vivo 14-3-3 interactions in the yeast Saccharomyces cerevisiae. *Biochemistry* **46**, 7781-7792

99. Obsilova, V., Kopecka, M., Kosek, D., Kacirova, M., Kylarova, S., Rezabkova, L., and Obsil, T. (2014) Mechanisms of the 14-3-3 protein function: regulation of protein function through conformational modulation. *Physiol. Res.* **63, Suppl. 1**, S155-S164

100. Yang, J., Kamide, K., Kokubo, Y., Takiuchi, S., Tanaka, C., Banno, M., Miwa, Y., Yoshii, M., Horio, T., Okayama, A., Tomoike, H., Kawano, Y., and Miyata, T. (2005) Genetic variations of regulator of G-protein signaling 2 in hypertensive patients and in the general population. *J. Hypertens.* **8**, 1497-1505

101. Matsuo, M., Coon, S. L., and Klein, D. C. (2013) RGS2 is a feedback inhibitor of melatonin production in the pineal gland. *FEBS Lett.* **587**, 1392-1398

102. Lynch, M. (2007) *The Origins of Genome Architecture*, Sinauer Associates, Inc., Sunderland, MA

CHAPTER 3:

CONTROL OF MAMMALIAN G PROTEIN SIGNALING BY N-TERMINAL ACETYLATION AND THE N-END RULE PATHWAY

From Park, S.-E., Kim, J.-M., Seok, O.-H., Cho, H., Wadas, B., Kim, S.-Y., Varshavsky, A., and Hwang, C.-S. (2015) *Science* 347(6227):1249-1252

ABSTRACT

Rgs2, a regulator of G proteins, lowers blood pressure by decreasing signaling through G±$_q$. Human patients expressing Met-Leu-Rgs2 (**ML**-Rgs2) or Met-Arg-Rgs2 (**MR**-Rgs2) are hypertensive compared to people expressing wild-type Met-Gln-Rgs2 (**MQ**-Rgs2). Here we found that wild-type **MQ**-Rgs2 and its mutant **MR**-Rgs2 were destroyed by the Ac/N-end rule pathway, which recognizes N$^{±}$-terminally acetylated (Nt-acetylated) proteins. The shortest-lived mutant **ML**-Rgs2 was targeted by both the Ac/N-end rule and Arg/N-end rule pathways. The latter pathway recognizes unacetylated N-terminal residues. Thus, the Nt-acetylated Ac-**MX**-Rgs2 (**X**=Q, L, R) proteins are specific substrates of the mammalian Ac/N-end rule pathway. Furthermore, the Ac/N-degron of Ac-**MQ**-Rgs2 was conditional, and Teb4, an endoplasmic reticulum membrane-embedded ubiquitin ligase, was able to regulate G protein signaling by targeting Ac-**MX**-Rgs2 proteins for degradation through their N$^{±}$-terminal acetyl group.

RESULTS AND DISCUSSION

Regulators of G protein signaling (RGSs) bind to specific G± subunits of heterotrimeric G proteins (G±²³) and accelerate the hydrolysis of G±-bound GTP, thereby abrogating the signaling by G proteins (*1-4*). The mammalian Rgs2 protein regulates stress responses, translation, circadian rhythms, Ca^{2+} channels, specific hormones, and cardiovascular homeostasis (*3-10*). Blood pressure-increasing vasoconstrictors such as norepinephrine and angiotensin II are upregulated by activated G±$_q$ proteins. The latter are deactivated by Rgs2 (*7*). Both *Rgs2*$^{-/-}$ and heterozygous

Rgs2$^{+/-}$ mice are strongly hypertensive (*8, 9*). Human patients with decreased Rgs2 signaling are hypertensive as well (*10*).

In some hypertensive patients, one of two *Rgs2* genes encoded Met-Leu-Rgs2 (**ML**-Rgs2), in which Gln at position 2 of wild-type Met-Gln-Rgs2 (**MQ**-Rgs2) had been replaced by Leu. Another hypertension-associated Rgs2 mutant is Met-Arg-Rgs2 (**MR**-Rgs2) (*10*). The Q2L and Q2R mutations are not detected in the general population (*10*). All three Rgs2 proteins are upregulated by a proteasome inhibitor, suggesting that they may be targeted by a proteasome-dependent proteolytic system (*11*).

The N-end rule pathway recognizes proteins containing N-terminal (Nt) degradation signals called N-degrons, polyubiquitylates these proteins and thereby causes their degradation by the proteasome (fig. A2.S1) (*12-20*). The main determinant of an N-degron is a destabilizing Nt-residue of a protein. Recognition components of the N-end rule pathway, called N-recognins, are E3 ubiquitin ligases that can target N-degrons. Regulated degradation of proteins by the N-end rule pathway mediates a broad range of biological functions (fig. A2.S1) (*12-20*).

The N-end rule pathway consists of two branches. One branch, the Arg/N-end rule pathway, targets unacetylated destabilizing Nt-residues (*12, 14, 16*). The Nt-residues Arg, Lys, His, Leu, Phe, Tyr, Trp, and Ile, as well as Nt-Met (if it is followed by a bulky hydrophobic ($) residue) are directly recognized by N-recognins (*16*). In contrast, the unacetylated Asn, Gln, Asp, and Glu (as well as Cys, under some conditions) Nt-residues are destabilizing owing to their preliminary enzymatic modifications (fig. A2.S1D).

The pathway's other branch, called the Ac/N-end rule pathway, targets proteins through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues (fig. A2.S1A, C) (*13, 15,*

16). Degrons and E3 ubiquitin ligases of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively. Approximately 90% of human proteins are cotranslationally and irreversibly Nt-acetylated by ribosome-associated Nt-acetylases (*21, 22*). (In contrast, acetylation of internal Lys residues is reversible and largely posttranslational.) Doa10, an endoplasmic reticulum (ER) membrane-embedded E3 ubiquitin ligase of the yeast *Saccharomyces cerevisiae* (*23, 24*), functions as an Ac/N-recognin (*13*). Not4, a cytosolic and nuclear E3, is another yeast Ac/N-recognin (*15*).

The Arg/N-end rule pathway is present in all examined eukaryotes, from fungi to mammals and plants (fig. A2.S1D) (*12, 18, 19*). In contrast, the Ac/N-end rule pathway (fig. A2.S1C) has been identified in *S. cerevisiae* (*13, 15, 16*), but its presence in mammals and other multicellular eukaryotes has been conjectural so far.

Owing to genetic tractability of *S. cerevisiae*, we began by examining wild-type human **MQ**-Rgs2 and its **ML**-Rgs2 mutant using cycloheximide (CHX) chases in yeast (*13, 15, 16*). The C-terminally HA-tagged **MQ**-Rgs2$_{ha}$ was short-lived ($t_{1/2}$ H30 min) in wild-type yeast, and was stabilized in both *naa20"* and *doa10"* cells, which lacked, respectively, the cognate NatB Nt-acetylase and the Doa10 Ac/N-recognin (*13*) (Figs. 1A, C, S1C and S2). Thus wild-type **MQ**-Rgs2$_{ha}$ was degraded, in yeast, largely by the Ac/N-end rule pathway (*13, 15, 16*).

The mutant human **ML**-Rgs2$_{ha}$ was also short-lived in wild-type *S. cerevisiae* ($t_{1/2}$ < 15 min), and was partially stabilized in both *naa30"* and *ubr1"* cells, which lacked, respectively, the cognate NatC Nt-acetylase and the Ubr1 N-recognin of the Arg/N-end rule pathway (Figs. 1B, D, S1C, S2, and S3A, D). Tellingly, **ML**-Rgs2$_{ha}$ was

nearly completely stabilized in double-mutant *naa30" ubr1"* cells, including strongly elevated time-zero (pre-chase) levels of **ML**-Rgs2$_{ha}$ (Figs. 1B, D and S3A, D).

Many cellular proteins are partially Nt-acetylated (*21*). Non-Nt-acetylated yeast **M$** -type proteins are eliminated by the Arg/N-end rule pathway, while the Nt-acetylated counterparts of these proteins are destroyed by the Ac/N-end rule pathway (*15, 16*). This dual-targeting pattern was also observed with human **ML**-Rgs2$_{ha}$ (Figs. 1B, D and S3A, D). The contribution of the Ubr1 N-recognin to the degradation of **ML**-Rgs2$_{ha}$ (Figs. 1B and S3A) indicated its incomplete Nt-acetylation in *S. cerevisiae*, similarly to results with natural **M$** -type yeast proteins (*15, 16*). Both wild-type **MQ**-Rgs2 and the second hypertension-associated mutant **MR**-Rgs2 were targeted (after their Nt-acetylation) solely by the Ac/N-end rule pathway, because the unacetylated Nt-Met followed by a non-$ residue such as Gln or Arg is not recognized by the Arg/N-end rule pathway (in contrast to an **M$** -type protein such as **ML**-Rgs2$_{ha}$) (fig. A2.S1D) (*16*).

Wild-type **MQ**-Rgs2 is a predicted substrate of the NatB Nt-acetylase (fig. A2.S2), in agreement with stabilization of **MQ**-Rgs2$_{ha}$ in *naa20" S. cerevisiae*, which lack NatB (Fig.1A, C). We analyzed, using mass spectrometry (*13*), the **MQ**-Rgs2 protein from human HEK293 cells, confirming its Nt-acetylation (fig. A2.S5A).

Considerably unequal levels of transiently expressed **MQ**-Rgs2$_{ha}$, **MR**-Rgs2$_{ha}$, and **ML**-Rgs2$_{ha}$ in human HeLa cells suggested their different stabilities (**MQ > MR** > **ML**) in these cells, an interpretation consistent with near-equal levels of the corresponding *MX-Rgs2* mRNAs (fig. A2.S4C, F, G). Indeed, CHX-chases in HeLa cells showed that **MR**-Rgs2$_{ha}$ ($t_{1/2}$ < 8 min) and **ML**-Rgs2$_{ha}$ ($t_{1/2}$ < 5 min) were shorter-lived than the also unstable wild-type **MQ**-Rgs2$_{ha}$ ($t_{1/2}$ H15 min) (Figs. 1E, F and S3B, E). The

same (**MQ > MR** > **ML**) order of degradation rates was observed in ^{35}S-pulse-chases of **MQ**-Rgs2$_{ha}$, **MR**-Rgs2$_{ha}$, and **ML**-Rgs2$_{ha}$ (Figs. 1J and S4J).

In contrast to Nt-acetylatable and (therefore) short-lived wild-type **MQ**-Rgs2$_{ha}$, the otherwise identical **PQ**-Rgs2$_{ha}$ (Pro-Gln-Rgs2$_{ha}$) (generated cotranslationally from **MPQ**-Rgs2$_{ha}$) was neither Nt-acetylated nor recognized by the Arg/N-end rule pathway (figs. S1 and S2). **PQ**-Rgs2$_{ha}$ was long-lived in HeLa cells (Fig. 1G, H). This result was an additional, conceptually independent evidence for the targeting of Ac-**MQ**-Rgs2$_{ha}$ by the Ac/N-end rule pathway (fig. A2.S1C).

Remarkably, the exogenous (overexpressed) wild-type **MQ**-Rgs2$_{ha}$ was much shorter-lived ($t_{1/2}$ H15 min) than the endogenous **MQ**-Rgs2 ($t_{1/2}$ H3 hr) in HeLa cells that did not overexpress **MQ**-Rgs2$_{ha}$ (Figs. 1E, F, 2B, S3B, E and S4A, B). These results, with human cells, agreed with the recent demonstration of the biologically relevant conditionality of Ac/N-degrons in *S. cerevisiae*, owing to steric shielding of these degrons in cognate protein complexes (*15*). Given this understanding with natural Ac/N-end rule substrates in yeast (*15, 16, 25*), the present results (Figs. 1E, F, 2B, S3B, E and S4A, B) are what one would expect if Ac-**MQ**-Rgs2, in human cells that do not overexpress it, can be (reversibly) shielded from the Ac/N-end rule pathway soon after Nt-acetylation of **MQ**-Rgs2. This shielding would occur through formation of a physiologically relevant protective complex(es) between Ac-**MQ**-Rgs2 and its protein ligand(s). In contrast to the long half-life of the endogenous, “stoichiometrically” expressed **MQ**-Rgs2, its overexpression in HeLa cells would make the resulting “unprotectable” excess of Ac-**MQ**-Rgs2$_{ha}$ molecules vulnerable to destruction by the Ac/N-end rule pathway, thereby accounting for the large difference between the slowly

degraded endogenous **MQ**-Rgs2 (in cells that do not overexpress **MQ**-Rgs2) and the short half-life of overexpressed **MQ**-$Rgs2_{ha}$ (Figs. 1E, F, 2B, S3B, E and S4A, B). Indeed, the short-lived **MQ**-Rgs2 was strongly stabilized by co-overexpression of one of its binding partners, the $G\pm_{q}$ protein (Fig. 3A, B).

A candidate Ac/N-recognin of the mammalian Ac/N-end rule pathway was Teb4, an ER membrane-embedded E3 ubiquitin ligase that polyubiquitylates proteins retrotranslocated from the ER. Teb4 is similar to the *S. cerevisiae* Doa10 Ac/N-recognin (*13, 23, 24*). We found that human Teb4 was indeed an Ac/N-recognin:

(i) Transiently expressed **MQ**-$Rgs2_{ha}$ was upregulated by a proteasome inhibitor, while coexpression of human Teb4 ($Teb4_{3f}$) and **MQ**-$Rgs2_{ha}$ in HeLa cells down-regulated **MQ**-$Rgs2_{ha}$ (Fig. 2A, F). In addition, incrementally higher levels of $Teb4_{3f}$ resulted in incrementally lower levels of endogenous **MQ**-Rgs2 (Fig. 2G).

(ii) The level of **MQ**-$Rgs2_{ha}$ in HeLa cells was not decreased if **MQ**-$Rgs2_{ha}$ was coexpressed with $Teb4_{3f}^{C9A}$, a missense mutant that is inactive as a ubiquitin ligase (*24*) (Fig. 2F). Analogous assays with **MR**-$Rgs2_{ha}$ and **ML**-$Rgs2_{ha}$ gave similar results (fig. A2.S4D, E).

(iii) The in vivo polyubiquitylation of **MQ**-$Rgs2_{ha}$ in HeLa cells was increased by coexpression of $Teb4_{3f}$, but not by coexpression of the mutant $Teb4_{3f}^{C9A}$ (Fig. A2.S3F).

(iv) In CHX-chases, endogenous **MQ**-Rgs2 was stabilized by RNAi-mediated knockdowns of either endogenous Teb4 E3 or endogenous cognate NatB Nt-acetylase (Figs. 2B and S4A). Similar results were obtained with exogenous (overexpressed) **MQ**-$Rgs2_{ha}$, confirming the targeting of **MQ**-Rgs2 by Teb4 and indicating that this targeting required Nt-acetylation of **MQ**-Rgs2 by the NatB Nt-acetylase (Figs. 2B and S4A, B). In

addition, overexpressed **MR**-Rgs2$_{ha}$ as well as the also tested **MK-**Rgs2$_{ha}$ (containing a different basic residue at position 2) were stabilized by RNAi-based knockdown of Naa60, the catalytic subunit of cognate NatF Nt-acetylase (*21*, *22*), but these **MX-**Rgs2$_{ha}$ proteins were not stabilized by knockdown of the non-cognate NatB (Naa20) Nt-acetylase (Figs. 1I and S4H, I).

(v) Crosslinking-coimmunoprecipitation experiments indicated that Teb4 interacted with **MQ**-Rgs2 (more accurately, with Nt-acetylated Ac-**MQ-**Rgs2$_{ha}$, as shown below) (Fig. 2H, I).

(vi) **PQ**-Rgs2$_{ha}$ was neither Nt-acetylated nor recognized by the Arg/N-end rule pathway, and was a long-lived protein, in contrast to short-lived **MQ**-Rgs2$_{ha}$ (Figs. 1G, H and S1C). In agreement with these in vivo results, Teb4$_{3f}$ was coimmunoprecipitated with **MQ**-Rgs2$_{ha}$ (Ac-**MQ**-Rgs2$_{ha}$) but not with **PQ**-Rgs2$_{ha}$ (Figs. 2H, I and S3C).

(vii) GST-pulldown assays showed that Teb4$_{f}$ interacted with Ac-**MQ**-Rgs2$^{1-10}$-[GST] but not with **MQ**-Rgs2$^{1-10}$-[GST] or GST alone, indicating that the binding of Teb4 to Ac-**MQ**-Rgs2$^{1-10}$-[GST] required its Nt-acetyl group (Figs. 2C-E and S5B).

By increasing the rate of deactivation of G$\pm_q$, Rgs2 can down-regulate G$\pm_q$-activated protein kinases, including the growth-promoting kinase Erk1/2 (*3*, *4*). We asked whether the Teb4 Ac/N-recognin could regulate the activation of Erk1/2 through the degradation of Nt-acetylated Ac-**MQ**-Rgs2. The G$\pm_q$-coupled M3 acetylcholine receptor was expressed in HeLa cells either alone or together with **MQ**-Rgs2$_{ha}$. By activating the M3 receptor-coupled G$\pm_q$, the agonist carbachol strongly increased the level of activated Erk1/2 (measured by detecting its phosphorylation by the "upstream" Mek1/Mek2 kinases (*3*, *4*)). As predicted by the model in which the Teb4-Rgs2 circuit regulates the

activation of Erk1/2, this effect of carbachol on Erk1/2 was decreased in cells that also expressed **MQ**-Rgs2$_{ha}$, and was further diminished upon the RNAi-mediated knockdown of Teb4, a change that stabilized **MQ**-Rgs2$_{ha}$ and thereby further elevated its level (Fig. 3C).

Thus, wild-type human Ac-**MQ**-Rgs2, its hypertension-associated natural mutants Ac-**ML**-Rgs2 and Ac-**MR**-Rgs2, and its engineered mutant Ac-**MK**-Rgs2 are conditionally short-lived physiological substrates of the mammalian Ac/N-end rule pathway, and the Teb4 ubiquitin ligase acts as an Ac/N-recognin by targeting these proteins for degradation. Teb4 promotes G protein signaling by destroying the Ac-**MX**-Rgs2 proteins (**X**=Q, L, R), which are targeted for degradation through their N$^{\pm}$-terminal acetyl group. The faster degradation of **ML**-Rgs2 and **MR**-Rgs2 (in comparison to wild-type **MQ**-Rgs2), and their consequently lower levels can account, at least in part, for hypertensive phenotypes of these mutants. The resulting understanding of Rgs2 vis-á-vis the N-end rule pathway is summarized in fig. A2.S6.

Identification of Teb4 as an Ac/N-recognin of the mammalian Ac/N-end rule pathway suggests that some previously characterized substrates of Teb4, including specific transmembrane proteins ((*23, 24, 26*) and references therein), may be recognized by Teb4 at least in part though their Nt-acetylated Nt-residues (Ac/N-degrons), a testable proposition.

More than 30 human proteins contain an RGS-type domain (fig. A2.S7) (*3, 4*). The N-terminal Cys residue, a common feature of Rgs4, Rgs5, and Rgs16, can be oxidized in vivo by nitric oxide (NO) and oxygen. The resulting Nt-arginylation of oxidized Cys and the ensuing degradation of these RGSs by the Arg/N-end rule pathway

mediate its previously described function as a sensor of NO and oxygen (fig. A2.S1D) (reviewed in (*12, 18, 19*)). The present study identified Rgs2 (**MQ**-Rgs2) as a different kind of N-end rule substrate, an Nt-acetylated one. Given their inferred N-terminal sequences, it is possible, indeed likely, that a number of other mammalian RGS proteins (in addition to Rgs2, Rgs4, Rgs5 and Rgs16) may also prove to be substrates of either the Ac/N-end rule pathway, or the Arg/N-end rule pathway, or both of these proteolytic systems (fig. A2.S7).

Many cellular proteins contain Ac/N-degrons, having acquired them at birth (*13*). These degradation signals, recognized by the Ac/N-end rule pathway, tend to be conditional, owing to their shielding in cognate protein complexes (*15*). In addition, the Ac/N-end rule pathway and the Arg/N-end rule pathway are functionally complementary (*16*). These insights suggest the medically relevant possibility of controlling the levels (and thus the activities) of many different proteins, including Rgs2 and other RGSs, by modulating their Ac/N-degrons or specific components of the Ac/N-end rule pathway.

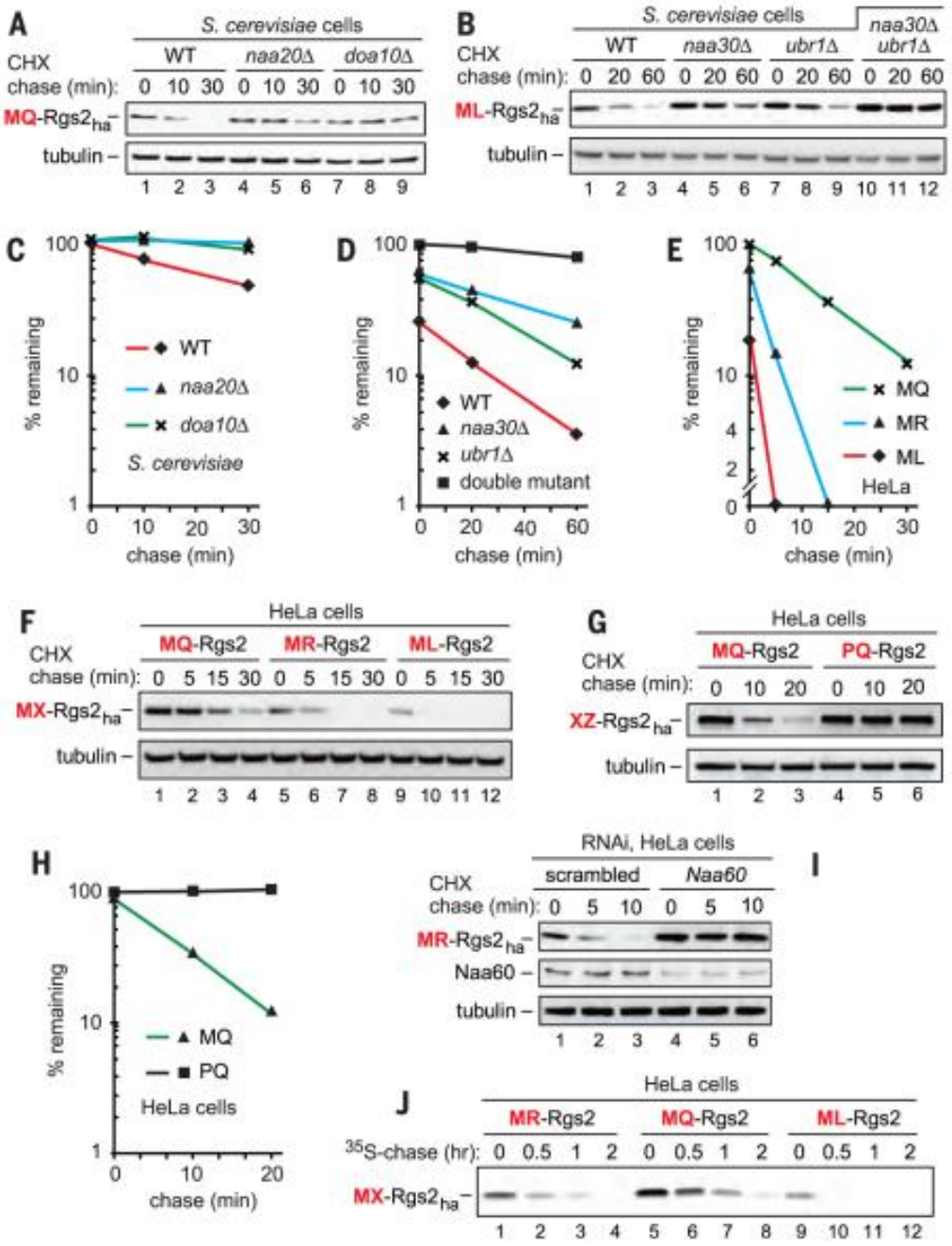


Fig. 3.1. Rgs2 as an N-end rule substrate. (**A**) CHX chases with wild-type human **MQ**-Rgs2$_{ha}$ in wild-type, *naa20"*, and *doa10" S. cerevisiae*. (**B**) As in **A** but with **ML**-Rgs2$_{ha}$ in wild-type, *naa30"*, *ubr1"*, and *naa30" ubr1"* strains. (**C, D, E**) Quantification of data in **A**, **B** and **F**, respectively. See the legend to fig. A2.S3 for definitions of "100%" levels at zero time. (**F**) CHX-chases with exogenously expressed **MX**-Rgs2$_{ha}$ (**X**=Q, R, L) in HeLa cells. (**G**) As in **F** but with **MQ**-Rgs2$_{ha}$ vs. **PQ**-Rgs2$_{ha}$. (**H**) Quantification of data in **G**. (**I**) CHX-chases with exogenously expressed **MR**-Rgs2 in HeLa cells subjected to RNAi for a either a "scrambled" target or Naa60. (**J**) ^{35}S-pulse-chases with **MX**-Rgs2$_{ha}$ (**X**=R, Q, L) in HeLa cells.

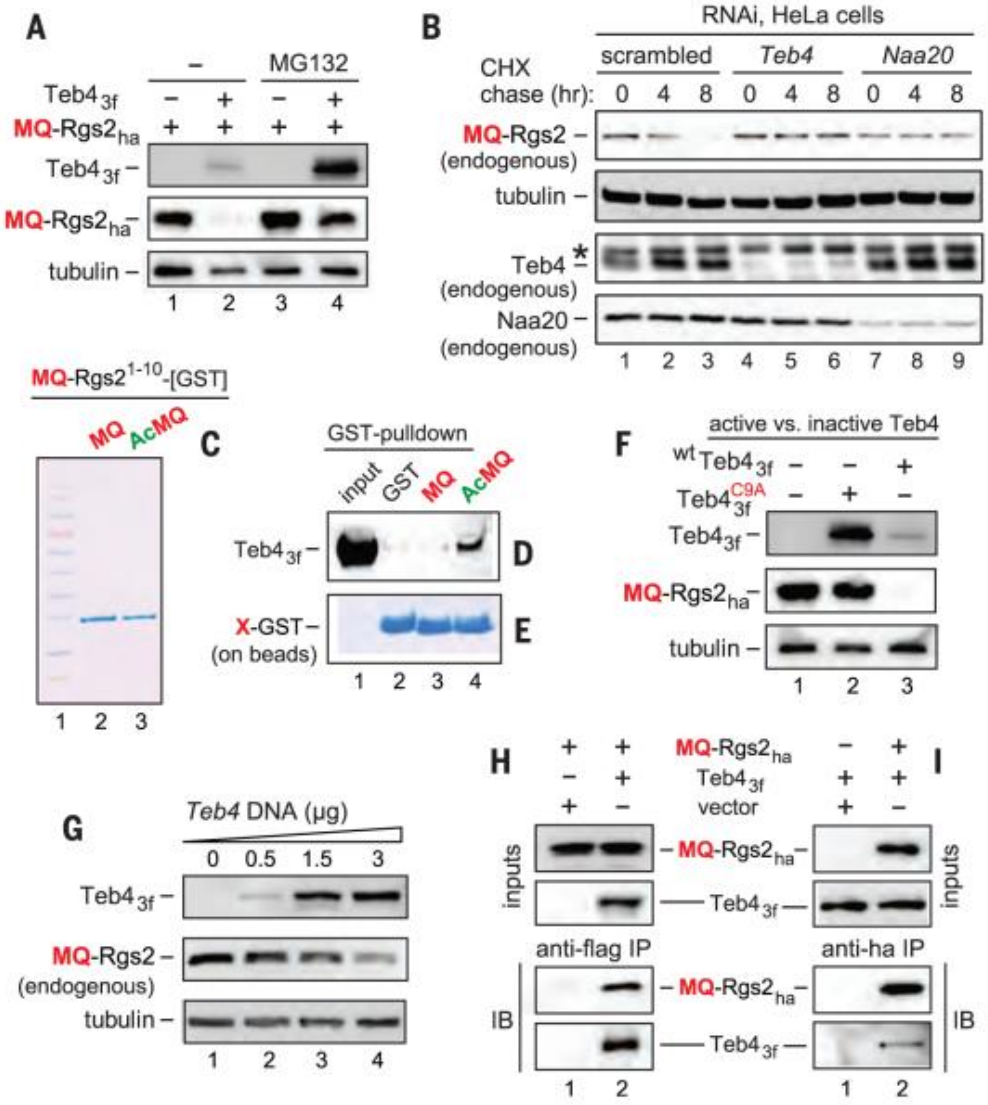


Fig. 3.2. Teb4 as an Ac/N-recognin. (**A**) **MQ**-Rgs2$_{ha}$ in HeLa cells with or without Teb4$_{3f}$ or the MG132 proteasome inhibitor. (**B**) CHX-chases with endogenous **MQ**-Rgs2, Teb4, and Naa20 subjected to RNAi for a "scrambled" target, or Teb4, or Naa20. The asterisk indicates a protein crossreacting with anti-Teb4. (**C**) M_r standards, purified **MQ**-Rgs2$^{1-10}$-[GST] and purified Ac-**MQ**-Rgs2$^{1-10}$-[GST], respectively. (**D**) Lane 1, Teb4$_{3f}$ input. Lanes 2-4, GST-pulldowns with GST alone, **MQ**-Rgs2$^{1-10}$-[GST], and Ac-**MQ**-Rgs2$^{1-10}$-[GST], respectively. (**E**) As in **D** but Coomassie-stained GST fusions released from beads. (**F**) **MQ**-Rgs2$_{ha}$ in HeLa cells overexpressing wild-type Teb4$_{3f}$ or Teb4$_{3f}^{C9A}$. (**G**) Increases in Teb4$_{3f}$ led to decreases in **MQ**-Rgs2$_{ha}$. (**H**) HeLa cells expressing **MQ**-Rgs2$_{ha}$ alone (lane 1) or together with Teb4$_{3f}$ (lane 2) were treated with a cell-penetrating crosslinker, followed by immunoprecipitations with anti-FLAG, reversal of crosslinks, SDS-PAGE, and immunoblotting with anti-HA and anti-FLAG. (**I**) As in **H** but Teb4$_{3f}$ alone in lane 1, and immunoprecipitations with anti-HA.

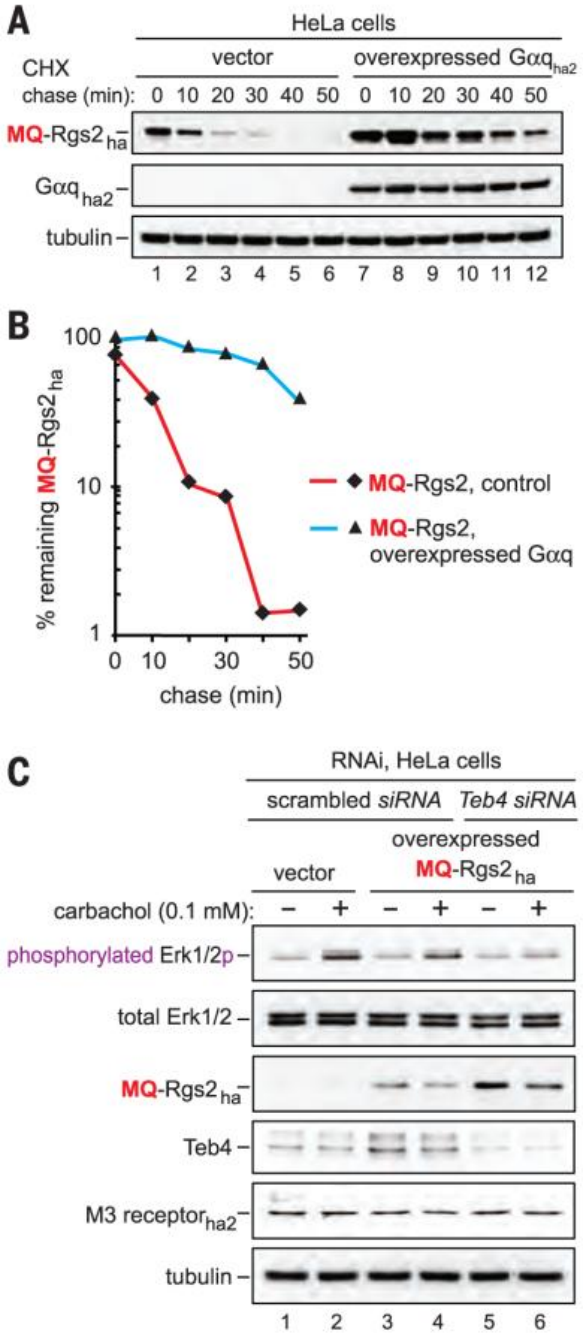


Fig. 3.3. G±q stabilizes Rgs2 while the Teb4-mediated degradation of Rgs2 increases signaling by G±q. (**A**) CHX-chase of **MQ**-Rgs2ha in HeLa cells that did not express or overexpressed the HA-tagged G±q. (**B**) Quantification of data in **A**. See the legend to fig. A2.S3 for definitions of "100%" levels at zero time. (**C**) HeLa cells were subjected to RNAi either for a "scrambled" target or for Teb4. The M3 receptor was transiently expressed either alone or together with **MQ**-Rgs2ha. The levels of indicated proteins, including the levels of either total Erk1/2 or the activated (specifically phosphorylated) Erk1/2p, were determined by SDS-PAGE and immunoblotting of cell extracts that had been prepared 10 min after treatment of cells with carbachol.

REFERENCES

1. R. J. Lefkowitz, A brief history of G-protein coupled receptors. Angew. Chem. 52, 6366-6378 (2013).
2. B. Kobilka, The structural basis of G-protein-coupled receptor signaling. Angew. Chem. 52, 6380-6388 (2013).
3. B. Sjögren, L. L. Blazer, R. R. Neubig, Regulators of G protein signaling proteins as targets for drug discovery. Prog. Mol. Biol. Transl. Sci. 91, 81-119 (2010).
4. A. J. Kimple, D. E. Bosch, P. M. Giguère, D. P. Siderovski, Regulators of G-protein signaling and their G± substrates: promises and challenges in their use as drug discovery targets. Pharmacol. Rev. 63, 728-749 (2011).
5. P. Chidiac, A. J. Sobiesiak, K. N. Lee, R. Gros, C. H. Nguyen, The eIF2B-interacting domain of RGS2 protects against GPCR agonist-induced hypertrophy in neonatal rat cardiomyocytes. Cell Signal. 26, 1226-1234 (2014).
6. M. Matsuo, S. L. Coon, D. C. Klein, RGS2 is a feedback inhibitor of melatonin production in the pineal gland. FEBS Lett. 587, 1392-1398 (2013).
7. M. R. Nance et al., Structural and functional analysis of the regulator of G protein signaling 2-(G±)q complex. Structure 21, 438-448 (2013).
8. K. M. Tang et al., Regulator of G-protein signaling-2 mediates vascular smooth muscle relaxation and blood pressure. Nat. Med. 12, 1506-1512 (2003).
9. S. P. Heximer et al., Hypertension and prolonged vasoconstrictor signaling in Rgs2-deficient mice. J. Clin. Investig. 111, 1259 (2003).
10. J. Yang et al., Genetic variations of regulator of G-protein signaling 2 in hypertensive patients and in the general population. J. Hypertens. 8, 1497-1505 (2005).

11. J. Bodenstein, R. K. Sunahara, R. R. Neubig, N-terminal residues control proteasomal degradation of Rgs2, Rgs4, and Rgs5 in human embryonic kidney 293 cells. Mol. Pharmacol. 71, 1040-1050 (2007).

12. A. Varshavsky, The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20, 1298-1345 (2011).

13. C.-S. Hwang, A. Shemorry, A. Varshavsky, N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327, 973-977 (2010).

14. C.-S. Hwang, A. Shemorry, A. Varshavsky, The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 12, 1177-1185 (2010).

15. A. Shemorry, C.-S. Hwang, A. Varshavsky, Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50, 540-551 (2013).

16. H.-K. Kim, R. R. Kim, J.-H. Oh, H. Cho, A. Varshavsky, C.-S. Hwang, The N-terminal methionine of cellular proteins as a degradation signal. Cell 156, 158-169 (2014).

17. J. M. Kim, C. S. Hwang, Crosstalk between the Arg/N-end and Ac/N-end rule. Cell Cycle 13, 1366-1367 (2014).

18. T. S. Tasaki, S. M. Sriram, K. S. Park, Y. T. Kwon, The N-end rule pathway. Annu. Rev. Biochem. 81, 261-289 (2012).

19. D. J. Gibbs, J. Bacardit, A. Bachmair, M. J. Holdsworth, The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. Trends Cell Biol. 24, 603-611 (2014).

20. A. Mogk, R. Schmidt, B. Bukau, The N-end rule pathway of regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 17, 165-172 (2007).

21. K. K. Starheim, K. Gevaert, T. Arnesen, Protein N-terminal acettyltransferases: when the start matters. Trends Biochem. Sci. 37, 152-161 (2012).

22. P. Van Damme et al., NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation. PLoS Genet. 7, e1002169 (2011).

23. S. G. Kreft, L. H.-C. Wang, M. Hochstrasser, Membrane topology of the yeast endoplasmic reticulum-localized ubiquitin ligase Doa10 and comparison with its human ortholog TEB4 (MARCH-VI). J. Biol. Chem. 281, 4646-4653 (2006).

24. G. Hassink et al., TEB4 is a C4HC3 RING-finger containing ubiquitin ligase of the endoplasmic reticulum. Biochem. J. 388, 647-655 (2005).

25. J. K. Monda et al., Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes. Structure 21, 1-12 (2012).

26. N. Zelcer et al., The E3 ubiquitin ligase MARCH6 degrades squalene monooxygenase and affects 3-hydroxy-3-methyl-glutaryl coenzyme A reductase and the cholesterol synthesis pathway. Mol. Cell. Biol. 34, 1262-1270 (2014)

CHAPTER 4:

ANALYZING N-TERMINAL ARGINYLATION THROUGH THE USE OF PEPTIDE ARRAYS AND PROTEIN DEGRADATION ASSAYS

INTRODUCTION

N-terminal arginylation (Nt-arginylation) of proteins is catalyzed by the Ate1 arginyltransferase (Arg-tRNA-protein transferase; R-transferase). This enzyme, present in all eukaryotes, from fungi to mammals and plants, is a component of the Arg/N-end rule pathway of protein degradation. R-transferase can Nt-arginylate "secondary" destabilizing N-terminal residues. The latter are a part of N-degrons that can be recognized by the Arg/N-end rule pathway. The "canonical" (definitively identified and characterized) secondary destabilizing residues are N-terminal Asp and Glu, and Cys, the latter after its oxidation to Cys-sulfinate or Cys-sulfonate through reactions involving nitric oxide (NO), oxygen, and N-terminal Cys oxidases [1, 2]. Therefore N-terminal Cys is, formally, a "tertiary" destabilizing residue, as it has to undergo two chemical modification steps (oxidation and Nt-arginylation) before the recognition of a corresponding N-degron by N-recognins (Ub ligases) of the Arg/N-end rule pathway. (For a complete introduction to the N-end rule pathway see *Introduction* to the thesis, beginning on page 6, and Figs. 1.3 and 4.1). In mammals such as the mouse, there are at least six isoforms of R-transferase, produced through alternative splicing of the *Ate1* pre-mRNA (Fig. 4.1B and C). Differences in substrate specificity have been reported between R-transferase isoforms, with disagreements, on this score, among different laboratories [3-5]. While there are more than 10 mammalian proteins (or natural protein fragments) that have already been directly identified as physiological substrates of the Ate1 R-transferase and the rest of the Arg/N-end rule pathway, the number of mammalian proteins (or protein fragments) that can be predicted to be Nt-arginylated by

R-transferase measures in the hundreds, if predictions are based on the presence, in a protein or its natural fragment, of the N-terminal Glu, Asp, or Cys residues [6-10].

The first natural proteins that were shown to contain N-terminal Cys (Nt-Cys)-based N-degrons were the regulators of G proteins Rgs4, Rgs5, and Rgs16. They were initially identified as N-end rule substrates through the use of an expression-cloning screen in reticulocyte extract [11]. Subsequent analyses of these natural substrates revealed that prior to their recognition and Nt-arginylation by R-transferase, the N-terminal Cys residue had to be oxidized to Cys-sulfonate ($CysO_3$) [12]. This finding, as well as the observation that Rgs4, Rgs5, and Rgs16 each contain a basic residue immediately following their N-terminal Cys, led, initially, to the suggestion that a basic residue at position 2 (not counting the initially present and cotranslationally removed N-terminal Met) would be required for efficient in vivo Cys-oxidation and subsequent Nt-arginylation. More recently, several other Ate1-dependent Arg/N-end rule substrates bearing N-terminal Cys but lacking a position-2 basic residue, have been identified, indicating that N-terminus-proximal amino acid sequence motifs that facilitate the oxidation and subsequent Nt-arginylation of N-terminal Cys are more complex than the presence of a basic residue at position 2 [6].

Several studies, largely by the Kashina laboratory, over the last decade have claimed that R-transferase can also arginylate, in vivo, several “non-canonical” (non-Asp, non-Glu, non-Cys) Nt-residues [13-17]. Whether these claims are correct remained unclear, in part because the cited data were non-definitive, and also because other laboratories have, so far, failed to find evidence for a broader specificity of Ate1 R-transferase [18-20]. Furthermore, another laboratory described a novel form of the 13-

residue neurotensin hormone that apparently contained the arginylated internal Glu-4 residue. In this posttranslational modification, the Arg residue was apparently conjugated, via an isopeptide bond, to the ^{3}-carboxyl of the Glu-4 residue [21]. While the study's authors did not identify the enzyme catalyzing this reaction, a later work by Kashina and colleagues produced evidence consistent with the possibility that the Ate1 R-transferase is capable of arginylating not only alpha-amino groups of specific N-terminal residues but also side-chain carboxylate groups of internal (non-N-terminal) Asp and Glu residues [22]. These still extant, but neither definitively confirmed nor refuted lines of evidence [13-17, 22-25], illustrate the current uncertainty about "non-canonical" substrates of R-transferase, outside of its unambiguously defined specificities for N-terminal Asp, Glu or (oxidized) Cys.

In the present study, we employed arrays of immobilized short peptides to address the substrate specificity of the mouse Ate1 R-transferase. We also used reticulocyte-based pulse-chase degradation assays to identify novel Nt-Cys substrates of the Arg/N-end rule pathway. We report here that the four major splicing-derived isoforms of the mouse Ate1 R-transferase have similar Nt-arginylation specificities in vitro. We also demonstrate that the sequence context of a substrate's N-terminus, particularly the identity of a residue at position 2, can significantly influence the rate of arginylation activity by R-transferase. That rate, in turn, would influence the overall kinetics of degradation of a corresponding protein by the Ate1-dependent arginylation branch of the Arg/N-end rule pathway. Furthermore, as described below, we found no evidence for any significant activity of the Ate1 R-transferase toward previously claimed non-canonical N-

terminal amino acid sequences, suggesting that the earlier evidence about non-canonical arginylation specificities of R-transferase [14, 23] is incorrect.

EXPERIMENTAL PROCEDURES

Plasmids, cDNAs, and Primers

A cDNA library was generated from wild-type C57BL/6J embryonic day 12 mouse embryos (The Jackson Laboratory, Bar Harbor, ME). Briefly, a single embryo was homogenized by bead-beating in Buffer RLT (Qiagen, Hilden, Germany) for 5 cycles (30 s each, speed 6.5, resting 2 min on ice between cycles) using a FastPrep-24 (MP Biomedicals, Santa Ana, CA). RNA was purified from embryo homogenates using an RNeasy Mini kit, following manufacturer's instructions (Qiagen). Total embryonic cDNA was synthesized using a Transcriptor First Strand cDNA Synthesis kit (Roche Life Sciences). cDNA was synthesized following manufacturer's instructions, with 1 µg total RNA used as substrate with a mixture of random hexamer and oligo(dT) primers. NEB Turbo *Escherichia coli* (New England Biolabs (NEB), Ipswich, MA) was used for cloning and maintaining plasmids. Phusion High-Fidelity DNA polymerase (NEB) was used for PCR. All constructed plasmids were verified by DNA sequencing. pKP496, the parental plasmid used for constructing Ub-fusion-technique (URT) plasmids, has been described previously [8].

Mdm2 – The mouse *Mdm2* open reading frame (ORF) was amplified using total mouse embryo cDNA and primers 316 and 319 for Cys-Mdm2, primers 317 and 319 for Val-Mdm2, or primers 318 and 319 (Table 4.4) for ArgCys-Mdm2. The resulting

fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW117, pBW130, and pBW131, respectively.

JunB – The mouse *JunB* ORF was amplified using total mouse embryo cDNA and primers 312 and 315 for Cys-JunB, primers 313 and 315 for Val-JunB, or primers 314 and 315 (Table 4.4) for ArgCys-JunB. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW116, pBW128, and pBW129, respectively.

Cdk14 – The mouse *Cdk14* ORF was amplified using total mouse embryo cDNA and primers 304 and 307 for Cys-Cdk14, primers 305 and 307 for Val-Cdk14, or primers 306 and 307 (Table 4.4) for ArgCys-Cdk14. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW115, pBW167, and pBW168, respectively.

B2L14– The mouse *B2L14* ORF was amplified using total mouse embryo cDNA and primers 300 and 303 for Cys-B2L14, primers 301 and 303 for Val-B2L14, or primers 302 and 303 (Table 4.4) for ArgCys-B2L14. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW114, pBW165, and pBW166, respectively.

Usp27x– The mouse *Usp27x* ORF was amplified using total mouse embryo cDNA and primers 320 and 323 for Cys-Usp27x, primers 321 and 323 for Val-Usp27x, or primers 322 and 323 (Table 4.4) for ArgCys-Usp27x. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW118, pBW169, and pBW170, respectively.

hMDM4 – The human MDM4 ORF was amplified from Dharmacon cDNA (Clone ID: 8144109) and primers 841 and 844 for Cys^{362}-MDM4, primers 842 and 844 for Val^{362}-MDM4, or primers 843 and 844 (Table 4.4) for $ArgCys^{362}$-MDM4. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, generating the plasmids pBW448, pBW449, and pBW450, respectively.

hMDM2 – The human MDM2 ORF was amplified from Dharmacon cDNA (Clone ID:3049213) and primers 860 and 863 for Cys^{362}-MDM2, primers 861 and 863 for Val^{362}-MDM2, or primers 862 and 863 (Table 4.4) for $ArgCys^{362}$-MDM2. The resulting fragments were cut with SacII and ClaI and cloned into SacII/ClaI-cut pKP496, yielding the plasmids pBW448, pBW449, and pBW450, respectively.

Purification of Mouse Ate1 R-Transferase from *E. coli*

E. coli plasmids expressing the untagged mouse *Ate1* cDNAs, encoding Ate1 isoforms, have been described previously [26] (Table 4.3). These plasmids pCB407 ($Ate1^{1B7A}$), pCB408 ($Ate1^{1B7B}$), pCB409 ($Ate1^{1A7A}$), and pCB410 ($Ate1^{1A7B}$) expressed His_{10}-Ub Ate1 fusions that could be purified and deubiquitylated in vitro using a previously described technique [27].

1-liter cultures of BL21(DE2) *E. coli* transformed with one of the above plasmids were grown to an A_{600} of ~0.6 at 37°C in Luria Broth (LB) containing 50 μg/ml ampicillin. Cultures were cooled on ice for 45 min, then induced by the addition of isopropyl-² -D-thiogalactoside (IPTG) to the final concentration of 0.25 mM for 6 h at room temperature. Cells were harvested by centrifugation at 2,000g for 10 min at 4°C. Pelleted cells were resuspended in 15 ml PBS containing 30% (vol/vol) glycerol, 0.3 M

NaCl, 12 mM imidazole, 20 mM ² -mercaptoethanol, and 1 mg/mL lysozyme, followed by freezing in liquid nitrogen. Samples were then thawed slowly on ice and centrifuged at 27,000g for 30 min at 4°C. The N-terminal His_{10}-Ub fusions of Ate1 isoforms were then purified by Ni-NTA chromatography, with elution by 0.3 M imidazole. Eluted proteins were dialyzed overnight against 5% (vol/vol) glycerol, 0.3 M NaCl, 2 mM ² -mercaptoethanol, and 50 mM Na_2HPO_4-NaH_2PO_4 (pH 8.0), containing 1X Complete EDTA-free protease inhibitor cocktail (Roche). The N-terminal His_{10}-Ub moiety was cleaved off by incubating dialyzed samples for 1 h at 37°C with purified Usp2cc deubiquitylase [27]. The latter was added at a 1:10 molar ratio. The resulting samples were dialyzed against 5% (vol/vol) glycerol, 0.15 M NaCl, 10 mM ² -mercaptoethanol, and 50 mM Tris (pH 7.4), followed by further purification of (deubiquitylated) Ate1 using Mono-S chromatography (Pharmacia Biotech: Id #9723125), with an elution gradient between 0.15 M and 1 M NaCl in the above dialysis buffer. Peak Ate1 fractions were pooled and dialyzed overnight at 4°C against 30% (vol/vol) glycerol, 0.15 M NaCl, 10 mM ² -mercaptoethanol, and 50 mM Tris-HCl (pH 7.4).

CelluSpots-Based Arginylation Assay

CelluSpots™ peptide arrays spotted on glass slides were synthesized by Intavis (Köln, Germany). Each slide contained two side-by-side identical copies of the array (Fig. 4.2A). Ate1-dependent arginylation assays were performed as described previously [4], with slight modifications. The peptide arrays were “outlined” with ImmEdge hydrophobic barrier pen (Vector Labs, Burlingame, CA) to minimize the required reaction volume. 90-µl reactions containing 5.8 µM ^{14}C-Arg (Perkin Elmer

NEC267E050UC; 346 mCi/mmol), purified *Escherichia coli* tRNA (0.6 mg/mL), and *E. coli* aminoacyl-tRNA synthestases (800 U/mL; Sigma-Aldrich) in 5 mM ATP, 0.15 M KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 50 mM Tris-HCl (pH 8.0) were preincubated at 37°C for 15 min to allow the formation of Arg-tRNA. Thereafter a specific purified Ate1 isoform (2 µg) was added and the resulting mixtures (100 µl final volumes) were added to each array. Arginylation reactions were allowed to proceed for 1 h at 37°C in a humidified chamber with gentle rocking. Reactions were quenched by rinsing the arrays twice with a large volume excess (10 ml) of Tris-Buffered Saline (TBS) containing also 1% Tween-20 and RNase A (10 µg/µl). Arrays were then washed 3 times for 20 min each in the same buffer (including RNase A) at 37°C, followed by an overnight (~16 h) wash in TBS containing 1% Tween-20 (without RNase A) at room temperature. The resulting slides were washed once with 10 mM Tris, pH 7.5, and were allowed to air-dry, followed by a 2-week autoradiograhy, using a Phosphorimager-type screen. The exposed screens were scanned using a Storm-860 Scanner (Molecular Dynamics, Caesarea, Israel). Quantification of ^{14}C spot intensities was performed using the array analysis module of ImageQuant TL software (GE Healthcare, Little Chalfont, UK).

In Vitro Transcription–Translation–Degradation Assay

The TNT T7 Coupled Transcription/Translation System (Promega, Madison, WI) was used to carry out transcription–translation–degradation assays, largely as described previously [8, 9, 28]. Nascent proteins in reticulocyte extract were pulse-labeled with L-[^{35}S]methionine (0.55 mCi/ml, 1,000 Ci/ mmol; MP Biomedicals, Santa Ana, CA) for 10

min at 30°C in the total volume of 30 µl. The labeling was quenched by the addition of cycloheximide and unlabeled methionine to the final concentrations of 0.1 mg/ml and 5 mM, respectively, bringing the total volume to 40 µl. Samples of 10 µl were removed at the indicated time points, and the reaction was terminated by the addition of 80 µl of TSD buffer (1% SDS, 5 mM DTT, 50 mM Tris-HCl, pH 7.4), snap-frozen in liquid nitrogen, and stored at -80°C until use. Samples were then heated at 95°C for 10 min and then diluted with 1 ml of TNN buffer (0.5% NP40, 0.25 M NaCl, 5 mM Na-EDTA, 50 mM Tris-HCl, pH 7.4) containing 1x Complete Protease Inhibitor Cocktail (Roche Diagnostics, Indianapolis, IN). Samples were immunoprecipitated using 10 µl of anti-FLAG M2 magnetic beads (Sigma-Aldrich) with rocking at 4°C for 4 h, followed by four washes in TNN buffer, resuspension in 20 µl SDS sample buffer, and heating at 95°C for 10 min followed by SDS 4–15% PAGE and autoradiography. Quantification of autoradiograms was carried out using Storm 860 Scanner (Molecular Dynamics).

RESULTS and DISCUSSION

Use of CelluSpots Peptide Arrays to Analyze the Specificity of R-Transferase

CelluSpots arrays are produced by the synthesis of peptides on a modified cellulose support that is dissolvable in trifluoroacetic acid (TFA) solution, thereby allowing the synthesized peptides to be (repeatedly) spotted onto coated microscope slides. This technology makes possible the synthesis of many replicate copies of a peptide array from a single set of peptide-synthesizing reactions. CelluSpots arrays have been successfully used to examine the specificity of histone-modifying enzymes for amino acid sequences in histones, to analyze the binding specificity of antibodies, to

characterize the binding specificity of SH2 (Src homology 2 domain) protein regions for phosphotyrosine-containing peptides, and to survey the specificities of kinases and kinase inhibitors [29-32].

To analyze the in vitro substrate specificity of the mouse Ate1 R-transferase, we performed ^{14}C-arginylation assays using a set of identical CelluSpots peptide arrays, each of which comprised 96 11-residue peptides. The readily attainable amino acid sequence diversity of these peptide-based assays, the availability of nearly unlimited identical copies of a given peptide array, and technically straightforward biochemical conditions of in vitro arginylation made CelluSpots a particularly promising approach to defining the in vitro specificity of mouse Ate1 R-transferase isoforms [4, 12, 33]. Amino acid sequences of peptides in our CelluSpots arrays were designed to examine several substrate-recognition aspects of R-transferase, including effects of a second-position residue, effects of charge clusters immediately downstream of the (arginylatable) N-terminal residue, effects of the Pro residue (an imino acid) close to N-terminal residue, and the influence of oxidation of N-terminal Cys on its ability to act as an acceptor of Arg in Ate1-catalyzed reactions. We also wished to address, using the technically direct and conceptually independent peptide-array technology, the earlier claims about arginylation, by the Ate1 R-transferase, of non-canonical amino acid sequences [14, 21-23] (Fig 4.2 and Table 4.1).

As a part of initial quality controls, and also to verify that peptides on CelluSpots arrays could actually serve as efficacious substrates of R-transferase, the arginylation reaction was optimized on slides containing duplicate copies of the 96-peptide array, in either the presence or absence of the purified mouse Ate1 R-transferase (Fig. 4.2). In

addition, the expected “positive” (a priori arginylatable) peptides and “negative” (a priori non-arginylatable, control) peptides were interspersed with other peptides throughout a CelluSpots array. Several peptides were included in duplicate, making it possible to assess the extent of variability of these arginylation assays within each experiment (see *Methods*). Preliminary experiments confirmed that CelluSpots peptides could serve as efficacious Ate1-dependent arginylation substrates, as illustrated by the incorporation of ^{14}C-Arg vis-à-vis each peptide spot and by the strict dependence of this incorporation on the presence of added Ate1 (Fig. 4.2B and *Methods*).

Different Ate1 Isoforms Exhibit Similar Substrate Preferences

The first subset of peptides on the array was designed to examine the effects of a second-position amino acid residue on Ate1-dependent arginylation. The 11-mer peptides for this experimental set, termed GSG, contained the Gly-Ser-Gly sequence at positions 3-5, followed by a quasi-randomly chosen sequence, Phe-Gly-Pro-Ala-Ser-Gly, at positions 6-11 (Fig. 4.3). The resulting sequence (positions 3-11) is not significantly similar to sequences in natural mouse proteins. Twenty 11-residue peptides containing the above “downstream” 9-residue sequence were designed to bear N-terminal Asp, a residue that is apparently the most efficacious acceptor of Arg (among N-terminal Asp, Glu or oxidized Cys), followed by a variable residue at position 2. In the resulting notation, specific 11-residue peptides of this set are denoted as $^{D(X)}$GSG, with “D” signifying N-terminal Asp. A second set of twenty otherwise identical peptides, termed $^{R(X)}$GSG, contained, each, the N-terminal Arg residue, which is not expected to be arginylated by Ate1. Peptides of this ($^{R(X)}$GSG) set were used as negative controls.

Preliminary experiments indicated that each of the twenty a priori arginylatable $^{D(X)}$GSG peptides, with varied residues at position 2, could indeed be arginylated by the added R-transferase, whereas none of the otherwise identical twenty $^{R(X)}$GSG peptides (negative controls) were arginylated above background levels (Fig. 4.2B).

Four major splicing-derived isoforms of the mouse Ate1 R-transferase have been described. They differ by the presence or absence of exon 1a versus 1b and exon 7a versus exon 7b. (Fig. 4.1B, 4.1C, and Introduction). To determine whether these Ate1 isoforms are comparably sensitive to the identity of the second residue in a substrate peptide, we examined each purified Ate1 isoform using CelluSpots arrays. The relative activities of all four mouse Ate1 R-transferase isoforms toward the same test peptides in four identical peptide arrays were found to be approximately the same (Fig. 4.3). Among the four isoforms, Ate1^{1A7B} exhibited a significantly lower overall enzymatic activity (per µg of added enzyme) than the other three isoforms, but the substrate preferences of Ate1^{1A7B}, vis-à-vis the examined set of peptides, were still similar to those of the other three isoforms (Fig. 4.3). All four Ate1 isoforms exhibited a significant preference for either a hydrophobic or a basic residue at position 2, in comparison to acidic residues or the Pro residue at this position (Fig. 4.3).

The Rate of Arginylation Is Influenced by the Overall Charge of a Region Adjacent to N-Terminal Residue

To examine whether the overall charge of a substrate's N-terminal region significantly affects arginylation of, for example, N-terminal Asp by Ate1, we assayed specific isoforms of purified mouse R-transferase using a $^{D(X)}$GSG peptide array in which

"X" was either a single residue or a stretch of four consecutive Asp, Arg, or Ala residues, starting at position 2. In the peptides where only a single Asp, Arg, or Ala residue was present (at position 2), the adjacent 3-residue sequence Gly-Ser-Gly remained at positions 3-5. In the peptides bearing a stretch of four consecutive Asp, Arg, or Ala residues, both position 2 and the Gly-Ser-Gly motif in positions 3-5 were replaced by these runs of identical residues (Fig. 4.5C).

All four Ate1 R-transferase isoforms exhibited significantly decreased efficacies of arginylation with peptides in which N-terminal Asp was followed by four basic (Arg) residues, in comparison to a single Arg at position 2 (Fig. 4.5A, columns 1 and 2; ^{DR1}GSG versus ^{DR4}GSG). All four Ate1 isoforms were also significantly less efficacious when N-terminal Asp was followed by four acidic (Asp) residues, in comparison to a single Asp at position 2 (Fig. 4.5A, columns 5 and 6; ^{DD1}GSG versus ^{DD4}GSG). In contrast, a difference in the efficacy of arginylation was much smaller when the N-terminal Asp residue was followed by four small uncharged (Ala) residues, in comparison to a single Ala at position 2 (Fig. 4.5A, columns 9 and 10; ^{DA1}GSG versus ^{DA4}GSG). Notably, even a single downstream (second-position) Asp made the resulting peptide a significantly less efficacious arginylation substrate than the otherwise identical peptide in which N-terminal Asp was followed by either a single positive (Arg) residue or a single small uncharged (Ala) residue (Fig 4.5A, compare column 1 with columns 5 and 9).

We next asked whether specific single-residue changes at position 2 altered the efficacy of Nt-arginylation of peptides containing the "canonical" arginylatable N-terminal residues, specifically Asp, Glu, or (oxidized) Cys. With peptides bearing N-

terminal Asp or Glu, their Ate1-mediated arginylation was found to be decreased (other parameters being equal) when either an acidic residue (Asp), or the Pro residue was present at position 2, in comparison to otherwise identical peptides bearing either Ser or Lys at position 2 (Fig. 4.6, columns 2, 4, 10, and 12). The isoform Ate1^{1B7B} was more sensitive to the presence of Asp at position 2 than the other three Ate1 isoforms, a rare example of detectable specificity differences among Ate1 isoforms (Fig. 4.6A and B). In this set of experiments, peptides bearing N-terminal Cys were not significantly Nt-arginylated nearly irrespective of the identity of a residue at position 2, in agreement with the previously identified requirement for a preliminary oxidation of N-terminal Cys [1, 12]. A small increase over background ^{14}C-Arg incorporation could be consistently observed with peptides bearing unmodified Cys when the second residue was basic (Arg) (Fig. 4.6A, columns 5-8). This effect could be caused either by a weak increase in the otherwise negligible activity of R-transferase toward unoxidized N-terminal Cys when the second residue was basic (Arg) or, non-alternatively, by a (possibly) increased propensity of N-terminal Cys to be partially (and nonenzymatically) oxidized to Cys-sulfinate or Cys-sulfonate by oxygen in the air if the second residue was basic.

The Rate of Ate1-Dependent Protein Degradation is Sensitive to the Overall Charge of a Region Adjacent to N-terminal Residue

To determine whether the observed influence of second-position residues and charge clusters on the efficacy of Nt-arginylation in CelluSpots assays were relevant to the rate of Ate1-dependent protein degradation, we carried out degradation assays using the previously described transcription-translation-degradation reticulocyte extract that

contains the Arg/N-end rule pathway [7-9]. Our Arg/N-end rule substrates were based on Rgs4, a regulator of specific G proteins that has been identified as an Ate1-dependent physiological substrate of the Arg/N-end rule pathway [1, 11, 34]. Mammalian Rgs4 proteins bear N-terminal Cys, after the cotranslational removal of their N-terminal Met by MetAPs. N-terminal Cys of Rgs4 has been shown to be efficaciously oxidized, by NO and oxygen, either in vivo or in reticulocyte extract to (ultimately) Cys-sulfonate, which can be Nt-arginylated by the Ate1 R-transferase. The resulting Nt-arginylated Rgs4 is processively destroyed to short peptides by the post-arginylation part of the Arg/N-end rule pathway (Fig. 4.4) [1, 12].

Our degradation assays in reticulocyte extract employed the Ub reference technique (URT), derived from the Ub fusion technique [35]. In this method, the cotranslational cleavage of a URT-based protein fusion by deubiquitylases in reticulocyte extract produces, at the initially equimolar ratio, a test protein with a desired N-terminal residue and a "reference" protein, in the present case ^{f}DHFR-Ub^{K48R}, a flag-tagged derivative of the mouse dihydrofolate reductase (Fig. 4.4A). In URT-based pulse-chase assays, a pulse-labeled test protein is quantified, during chase, by measuring its level relative to that of a stable reference protein at each time point, thereby increasing the accuracy of pulse-chase assays [35-37]

To simplify the mechanics of initial targeting, i.e., to bypass the N-terminal Cys-oxidation step required for the degradation of wild-type Rgs4 [1, 12, 34], its natural N-terminal Cys was mutated to Asp, in the context of a URT fusion (Fig 4.4A). The resulting Rgs4^{C2D}was a direct substrate of the Ate1 R-transferase. In rabbit reticulocyte extract, ^{DA}RgsS4$_{f}$ (bearing the N-terminal sequence Asp-Ala, i.e., bearing two N-terminal

mutations vis-à-vis the wild-type ^{CK}Rgs4 protein)) was degraded with a $t_{1/2}$ < 10 min (Fig. 4.4B, lanes 5-8), whereas the otherwise identical ^{RA}Rgs4$_f$ (bearing the N-terminal sequence Arg-Ala) was degraded even more rapidly ($t_{1/2}$ < 1 min; Fig 4.4B, lanes 1-4), because in the latter case the preliminary step of Nt-arginylation was neither possible nor necessary (Fig. 4.1A). In contrast, the otherwise identical ^{VA}RGS4$_f$ protein (bearing the N-terminal sequence Val-Ala, which is not recognized by the Arg/N-end rule pathway) was stable during the chase ($t_{1/2}$ >> 45 min; Fig. 4.4B, lanes 9-12). In sum, this set of otherwise identical test proteins, which differ solely at their positions 1 and 2, is a particularly suitable setting for examining effects of N-terminus-proximal changes on the rate of protein degradation by the Arg/N-end rule pathway (Fig 4.4B and C).

We also constructed derivatives of RGS4^{C2D} that contained either one or four Asp residues at positions 2-5 or, alternatively, either one or four Arg residues at the same positions, to determine whether the overall charge in the immediate vicinity of protein's N-terminus can affect the degradation of a protein by the Arg/N-end rule pathway. Remarkably, a stretch of four acidic (Asp) residues after N-terminal Asp (^{DD4}RGS4$_f$) rendered the test substrate long-lived, despite the fact that the N-terminal Asp residue, when it is present in favorable downstream sequence contexts (see above), is apparently the best substrate for Ate1-mediated arginylation (Fig. 4.5D, lanes 13-16, and Fig. 4.5E). In contrast, when four basic (Arg) residues followed Asp, the rate of degradation of the resulting ^{DR4}RGS4$_f$ was nearly identical to that of ^{DR1}RGS4$_f$, which contained a single Arg residue after N-terminal Asp (Fig. 4.5D, lanes 5-8 versus lanes 1-4, and Fig. 4.5E). These protein degradation-based findings were in agreement with the results of peptide-based Nt-arginylation assays with peptides whose N-terminal sequences were identical to

those of XZ-Rgs4$_f$ test proteins. Together, the results with CelluSpots peptide arrays and protein degradation assays in reticulocyte extract (Fig. 4.5A and B) indicated that acidic residues immediately downstream of an otherwise Nt-arginylatable N-terminal residue (e.g., Asp) of a protein can strongly attenuate its R-transferase-mediated Nt-arginylation and, consequently, the rate of its degradation by the Arg/N-end rule pathway.

We also compared degradation-rate effects of a second-position basic residue (Arg) versus acidic residue (Asp) and versus the imino residue Pro in the corresponding Asp-Z-Rgs4 proteins (Z=Asp, Pro, Lys). In agreement with a partial inhibition of arginylation of the analogous CelluSpot peptides (Fig. 4.6A), either an acidic residue (Asp) or the Pro residue at position 2 of the above Asp-Z-Rgs4$_f$ reporters significantly decreased the rate of their degradation , in comparison to otherwise identical Asp-Z-Rgs4$_f$ proteins that contained a basic (Lys) residue at position 2 (Fig. 4.6C and D). These results indicated that the identity of a residue at position 2 of an arginylation-dependent N-degron in a protein can strongly influence its degradation rate, at least in part by altering the rate of its (pre-degradation) Nt-arginylation.

Oxidation of N-terminal Cys in CelluSpots Peptides Allows Their Arginylation

Earlier studies of N-degrons containing N-terminal Cys have shown that the Ate1-mediated Nt-arginylation of Cys required its preliminary oxidation to Cys-sulfinate or Cys-sulfonate [1, 12]. In contrast, other earlier studies suggested that the Ate1 isoforms Ate1^{1A7A} and Ate1^{1A7B} might be R-transferases that are relatively specific for N-terminal Cys [3]. Our results with CelluSpots arrays indicated that the unmodified N-terminal Cys is either a poor arginylation substrate of the Ate1 R-transferase or not a substrate at all.

The latter possibility would obtain if the observed weak (barely above background) Nt-arginylation of N-terminal Cys-bearing CelluSpots peptides resulted from their partial nonenzymatic oxidation during preparation and handling of peptide arrays (Fig. 4.6A, columns 5-8).

To compare specific Ate1 R-transferase isoforms by their ability to arginylate the unmodified versus oxidized N-terminal Cys residue, we used Rgs4-derived ("wild-type" or modified) N-terminal peptides that contained either an unmodified (i.e., not deliberately oxidized) N-terminal Cys or oxidized N-terminal Cys (Cys-sulfonate) that was substituted for unmodified Cys during the synthesis of these peptides. As expected (given the previous evidence about oxidized N-terminal Cys being a prerequisite for its Nt-arginylation [1, 12]), we found that peptides bearing the N-terminal Cys-sulfonate residue (incorporated during peptide synthesis) were much better arginylation substrates than the otherwise identical peptides bearing the unoxidized (more accurately, not deliberately oxidized) N-terminal Cys residue (Fig. 4.7A, column 4 versus column 3).

Significantly, peptides bearing the oxidized (at the stage of peptide synthesis) N-terminal Cys-sulfonate were still not as efficacious acceptors of Arg in the Ate1-catalyzed conjugation reaction as were the otherwise identical peptides bearing N-terminal Asp (Fig. 4.7A, column 4 versus column 2). As one would expect, the negative-control peptides, bearing the N-terminal Arg-Cys sequence (a mimic of Nt-arginylated oxidized Cys) were not arginylated at all (Fig. 4.7A, column 1). All four examined isoforms of the mouse Ate1 R-transferase exhibited indistinguishable relative efficacies of arginylation in this N-terminal Cys-based setting. This result overtly disagreed with

the earlier assertion that some among these isoforms of Ate1 are relatively specific for N-terminal Cys [3].

Results with CelluSpots Assays Contradict the Earlier Claim About Arginylation of Proteins with N-terminal Residues That Are Distinct From Asp, Glu and Cys

Papers by Kashina and colleagues [14, 15, 21] have claimed, largely on the basis of mass spectrometric (MS) data, that several proteins (or protein fragments) were arginylated in vivo despite the absence of the canonical arginylatable N-terminal residues (Asp, Glu, Cys) at the N-termini of those proteins. Inasmuch as these claims were not based on a technically definitive evidence, and given the utility of CelluSpots arrays in addressing the substrate specificity of arginylation, we carried out arginylation assays with 11-residue peptides whose sequences were identical to N-terminal sequences of several mammalian proteins (or protein fragments) that Kashina and colleagues described as in vivo arginylation substrates [14].

Of the ten examined CelluSpots peptides, corresponding to N-terminal sequences of the ten previously described putative in vivo arginylation substrates [14], only one of them exhibited a (possible) weak above-background arginylation in CelluSpots arginylation assays (Fig. 4.7C, lanes 2-11, and Table 4.1). The sequence of the (weakly) positive 11-residue peptide (Fig. 4.7A, column 6) was identical to the first 11 residues of the Lys448-Lys464 fragment of the mouse MPAST2 protein (MPAST2$^{K448\text{-}K464}$ that was detected by the earlier MS analyses and classified as an arginylated fragment [14]. The observed weakly-above-background arginylation of the MPAST2-derived peptide, in CelluSpots assays, was weaker than the extent of arginylation of unmodified Cys test

peptides (Fig. 4.6A) and considerably weaker than the arginylation of "positive" control peptides in the same row of peptides (Fig. 4.7C, columns 1 and 12). The nine other peptides, whose amino acid sequences were identical to the N-terminal sequences of peptides from the nine other proteins that have been claimed, by Kashina and colleagues, to be arginylated [14], did not exhibit any arginylation above background levels of CelluSpots arginylation assays (Fig. 4.7C, lanes 2-11, and Table 4.1).

In one of the above-cited earlier works, a peptide corresponding to the Met589-Lys606 fragment of the mouse Myh9 protein (termed $Myh9^{M589\text{-}K606}$), was detected by MS and was classified as a peptide bearing the arginylated "non-canonical" Met-589 residue, with the underlying assumption that the Met-589 residue was the N-terminal residue of a corresponding (in vivo-formed) fragment of the Myh9 protein [14]. We carried out additional CelluSpots assays to determine whether, for example, an altered chemical state of N-terminal Met of the Met589-Lys606 fragment might have been responsible for the earlier data indicating that fragment's arginylation [14]. Specifically, we asked whether the Ate1 R-transferase could arginylate any of the 11-residue peptides whose sequences were either identical to the N-terminal sequence of the $Myh9^{M589\text{-}K606}$ peptide [14] (bearing unmodified N-terminal Met), or contained the same sequence save for an N-terminal Met-sulfoxide (instead of unmodified Met), or Met-sulfone, or Arg-Met. The latter would be an a priori non-arginylatable N-terminal sequence, used as a negative control. We found that none of these peptides, including the "wild-type" sequence of $Myh9^{M589\text{-}K606}$, generated a ^{14}C-Arg arginylation signal above background of CelluSpots arginylation assays, indicating the inability of these peptides to serve as Ate1 substrates at least in vitro (Fig. 4.7B, columns 1-4).

In another earlier study, a short peptide called neurotensin (it is produced through the processing of a larger precursor protein in the endoplasmic reticulum (ER) and Golgi, i.e., outside the currently known locations of the cytosolic/nuclear Ate1 R-transferase), was observed to be arginylated at its internal Glu-4 residue, presumably through the conjugation of Arg to the ^{3}-carboxyl group of Glu-4 [21]. We examined, using CelluSpots arginylation assays, a set of 11-residue peptides designed to mimic the N-terminal sequence of neurotensin that encompasses its Glu-4 residue. These peptides were synthesized to contain either the above Glu-4 of wild-type neurotensin or one of several other residues (Asp, Glu, Asn, or Gln) at the same position 4. Because the natural N-terminal sequence of neurotensin starts with the Gln residue, a nonenzymatic deamidation of that Gln would result in N-terminal Glu, a canonical arginylatable N-terminal residue (Fig. 4.1A) [38, 39]. To preclude the possibility of Nt-arginylation of neurotensin-like CelluSpots peptides, their N-termini were designed to bear Ala instead of wild-type Gln. We found that none of these neurotensin peptide mimics generated a ^{14}C-Arg arginylation signal above background of CelluSpots assays, indicating their inability to serve as Ate1 substrates at least in vitro (Fig. 4.7B, columns 5-8).

To the best of our knowledge, the most plausible alternative explanation of the overt discrepancy between our results and the earlier reports by Kashina and colleagues (cited above) that might “rescue” the arginylation-based interpretation of the earlier evidence is that our arginylation assays, which employed purified mouse Ate1 R-transferase isoforms, might have been missing a relevant (protein-based?) modifier of R-transferase activity that would have endowed Ate1 with the claimed (expanded) substrate specificity.

To address this possibility, we carried out otherwise identical CelluSpots arginylation assays in which gently prepared extracts from mouse Neuro2A cells were added to reaction mixes either in the absence or the presence of additionally added purified Ate1^{1B7A}. This modification of CelluSpots assays did not alter any of the findings described above, making the above (potential) explanation much less likely but not definitively refuted, given the in vitro nature of CelluSpots assays.

Analyzing Natural N-terminal Cys-Bearing Proteins for Their Degradation by the Arg/N-end rule pathway

At present, Rgs4, Rgs5 and Rgs16 proteins are the only known physiological substrates of the mammalian Arg/N-end rule pathway that bear N-terminal Cys as a result of the cotranslational removal of N-terminal Met, by MetAPs, from nascent counterparts of these proteins [1, 11, 34]. In an attempt to identify other physiological Arg/N-end rule substrates whose N-terminal Cys is, initially, a position-2 residue, we employed the Ub-reference technique (URT) (Fig. 4.4A) to express specific natural proteins whose N-termini begin with a Met-Cys sequence. A search, in the UniprotKB/Swiss-Prot databases, for human proteins whose ORFs encode the N-terminal Met-Cys sequence brought up ~180 such proteins (excluding splice variants; ScanProsite). As described below, we examined five proteins from this set (Fig. 4.8). These proteins were chosen for pulse-chase assays in reticulocyte extract in part because their (encoded) Met-Cys N-termini were highly (essentially completely) conserved among at least mammals (data not shown). Despite this conservation, and in contrast to results with Rgs4, Rgs5, and Rgs16, the URT-based pulse-chase assays with the above five N-terminal Cys-bearing proteins

showed that none them were significantly more unstable when expressed with N-terminal Cys versus N-terminal Val (the latter residue is not recognized by the Arg/N-end rule pathway at least in reticulocyte extract [6, 7, 9]).

An N-degron recognized by the Arg/N-end rule pathway comprises at least two determinants: a destabilizing N-terminal residue and a targetable (usually conformationally mobile) internal Lys residue(s) [40]. We asked whether the lack of degradation of the above proteins (Fig. 4.8) by the Arg/N-end rule pathway (which can confer a short half-life on, for example, Cys-Rgs4, a previously identified N-terminal Cys-bearing Arg/N-end rule substrate) was caused by the failure to arginylate their N-terminal Cys residues in reticulocyte extract, or was, instead, the result of a block at a post-arginylation step, owing, for example, to the absence of an efficaciously ubiquitylatable internal Lys and/or to inefficient degradation of a targeted and polyubiquitylated protein by the proteasome. To distinguish between some of these possibilities, each of the proteins in Fig. 4.8 was also expressed as its “pre-arginylated” Arg-Cys counterpart.

Tellingly, in each of these cases the Arg-Cys-bearing versions of the original N-terminal Cys-bearing proteins were short-lived in reticulocyte extract, in contrast to otherwise identical proteins bearing N-terminal Cys (Figs. 4.8B-J). We conclude that the lack of rapid degradation of the examined N-terminal Cys-bearing proteins that are cited in Fig. 4.8 was caused by their slow (possibly negligible) Nt-arginylation, implying that their N-terminal Cys residues could not be efficaciously oxidized under conditions in which the N-terminal Cys of, for example, Rgs4 could be rapidly oxidized and arginylated [11, 34] (Fig. 4.8).

With some of the tested N-terminal Arg-Cys-bearing proteins, a decrease in the level of an Arg-Cys-protein band during the chase was accompanied by the accumulation of high molecular mass derivatives of that proteins, suggesting that even when "pre-arginylated" Arg-Cys proteins were polyubiquitylated, some of them encountered additional kinetic inefficiencies at a stage that followed their polyubiquitylation, e.g., the stage of their actual degradation by the proteasome (data not shown).

Identification of a Natural Fragment of hMDM4 as an N-terminal Cys-Bearing Substrate of the Arg/N-end rule pathway

A Cys residue that resides at position 2 of a nascent protein is made N-terminal cotranslationally, through the (usually) complete removal of the initially present Met residue by ribosome-associated MetAPs [41]. In contrast, an internal Cys residue can be made N-terminal through a sequence-specific cleavage of a protein by a nonprocessive protease, such as a caspase, a calpain, or a separase. These and analogous intracellular proteases would cleave, posttranslationally, a subset of a given protein's molecules at a cleavage site in which the above Cys would be a P1' residue. Both N-terminal and internal Cys residues are susceptible not only to oxidation (described above) but also, for example, to palmitoylation (a reversible modification) or, alternatively, to Nt-acetylation, an apparently irreversible modification that can make a protein a target for degradation by the Ac/N-end rule pathway (Fig. 4.1) [42-47]. In sum, the actual metabolic fate of a newly exposed (cotranslationally or posttranslationally) N-terminal Cys is a function ofseveral kinetically competing, mutually exclusive chemical modifications. This complexity accounts, in part, for the current difficulty of predicting whether a specific N-

terminal Cys-bearing protein would be an efficacious substrate of the arginylation branch of the Arg/N-end rule pathway (positive examples: the Rgs4, Rgs5, Rgs16), or whether such a protein would remain long-lived, despite the presence of N-terminal Cys (negative examples: the proteins cited in Fig. 4.8).

Our laboratory has previously described two (out of two actually tested) natural C-terminal protein fragments that are produced through posttranslational cleavages of full-length proteins by caspases, bear (as a result) N-terminal Cys, and could be shown to be short-lived substrates of the Arg/N-end rule pathway [6]. It is possible that posttranslationally (as distinguished from cotranslationally) generated protein fragments bearing N-terminal Cys would be, in general, more likely to be efficacious (short-lived) substrates of the Arg/N-end rule pathway, owing, for example, to their relatively slow posttranslational Nt-acetylation (since the bulk of Nt-acetylases are associated with the ribosomes, resulting in their relative low levels in the bulk solvent). To address this possibility, we searched CutDB [48], a database of proteolytic cleavages, for posttranslationally generated C-terminal protein fragments bearing N-terminal Cys, identifying 39 C-terminal fragments of natural proteins that are produced by single posttranslational proteolytic cleavages and bear N-terminal Cys (Table 4.2). Two of these N-terminal Cys-bearing proteins, Traf1 and Ripk1, were previously identified by our laboratory as short-lived Arg/N-end rule substrates [8].

Four of the newly identified (through the above CutDB search), posttranslationally generated C-terminal protein fragments bearing N-terminal Cys also contained a second-position basic residue. The resulting Cys-basic residue motif is identical to the one in, e.g., Rgs4, a known arginylated and short-lived substrate of the

Arg/N-end rule pathway [1, 11, 34]. We used pulse-chase assays in reticulocyte extract to examine two of the above four N-terminal Cys-bearing, caspase-generated protein fragments specifically the Mdm4 protein (mouse double minute 4 homolog) and the Mdm2 protein, an E3 Ub ligase that targets, in particular, the p53 protein (Fig. 4.9).

URT-based pulse-chases indicated that the N-terminal Cys-bearing fragment of human Mdm4 was, in fact, short-lived in reticulocyte extract, in contrast to its otherwise identical N-terminal Val-bearing counterpart (Fig. 4.9B). In vivo pulse-chases, in human HEK293T cells, of the same set of Mmd4 fragments, bearing N-terminal Cys, Val, or ArgCys, confirmed the instability of the fragments bearing N-terminal Cys and N-terminal ArgCys, in comparison to their N-terminal Val-bearing counterpart (a control) (Fig. 4.9C). We conclude that this N-terminal Cys-bearing fragment of human Mdm4 is indeed, most likely, a physiological substrate of the Arg/N-end rule pathway.

In contrast to instability of the human Mdm4 C-terminal fragment, the caspase-generated, N-terminal Cys-bearing fragment of human Mdm2, which also contained the N-terminal Cys-basic residue motif, was found to be long-lived in reticulocyte extract (Fig. 9A). Together, these findings indicated, yet again, that the metabolic instability of an N-terminal Cys-bearing protein (or a protein fragment) is a function of multiple kinetic variables, including the timing of fragment's formation (e.g., cotranslational versus posttranslational), the nature of a sequence immediately downstream of N-terminal Cys, and other parameters as well including specific aspects of a (potential) N-degron outside the immediate vicinity of the fragment's N-terminus. Further studies, including those that use wild-type versus Ate1-lacking mice, will be required to verify,

definitively, whether the caspase-generated, N-terminal Cys-bearing C-terminal fragment of Mdm4 is a bona fide Arg/N-end rule substrate in specific mouse tissues.

Experimental approaches described above employed CelluSpots peptide arrays as a useful tool for analyzing Ate1-mediated arginylation and its specificity. By examining a broad range of N-terminal sequences and by taking advantage of the “repeatability” (multiple copies) of CelluSpots assays, we were able to demonstrate that there is little if any qualitative difference, at least in vitro, between the four mouse Ate1 R-transferase isoforms in their specificities of arginylation. We also verified and confirmed, using the CelluSpots strategy, the canonical arginylatable N-terminal residues Asp, Glu, and (oxidized) Cys are, in fact, the most preferred (and possibly virtually exclusive) substrates of Ate1-encoded R-transferases. We also analyzed specific sequence requirements immediately downstream of an (arginylatable) N-terminal residue, and showed, among other things, that the identity of a residue at position 2 can significantly influence the efficacy of arginylation by R-transferase and ultimately, therefore, the degradation rate of a corresponding protein (or a protein fragment) by the Arg/N-end rule pathway.

Tellingly, we found no evidence that the mouse Ate1 R-transferase can arginylate non-canonical N-terminal residues (as described above, such an ability was claimed in earlier studies by another laboratory [14]). We also found no supporting evidence for another earlier claim, by the same laboratory [22], of the possibility of Ate1-mediated arginylation of gamma-carboxyl side chains of either N-terminal or internal Asp or Glu residues. These negative results are in agreement with other recent reports that failed to confirm the above earlier claims [18-20]. Although our attempts to determine whether the

specificity of purified R-transferase can be altered (expanded?) by a factor present in mouse cell extracts have also yielded negative conclusions (see above), a caveat to these otherwise detailed and rigorous arginylation data (Figs. 4.5-4.7) is that they are largely based on in vitro assays, and therefore may still overlook a setting in which the specificity of R-transferase may be altered relative to the one observed in vitro. While unlikely, this possibility cannot be precluded, at present.

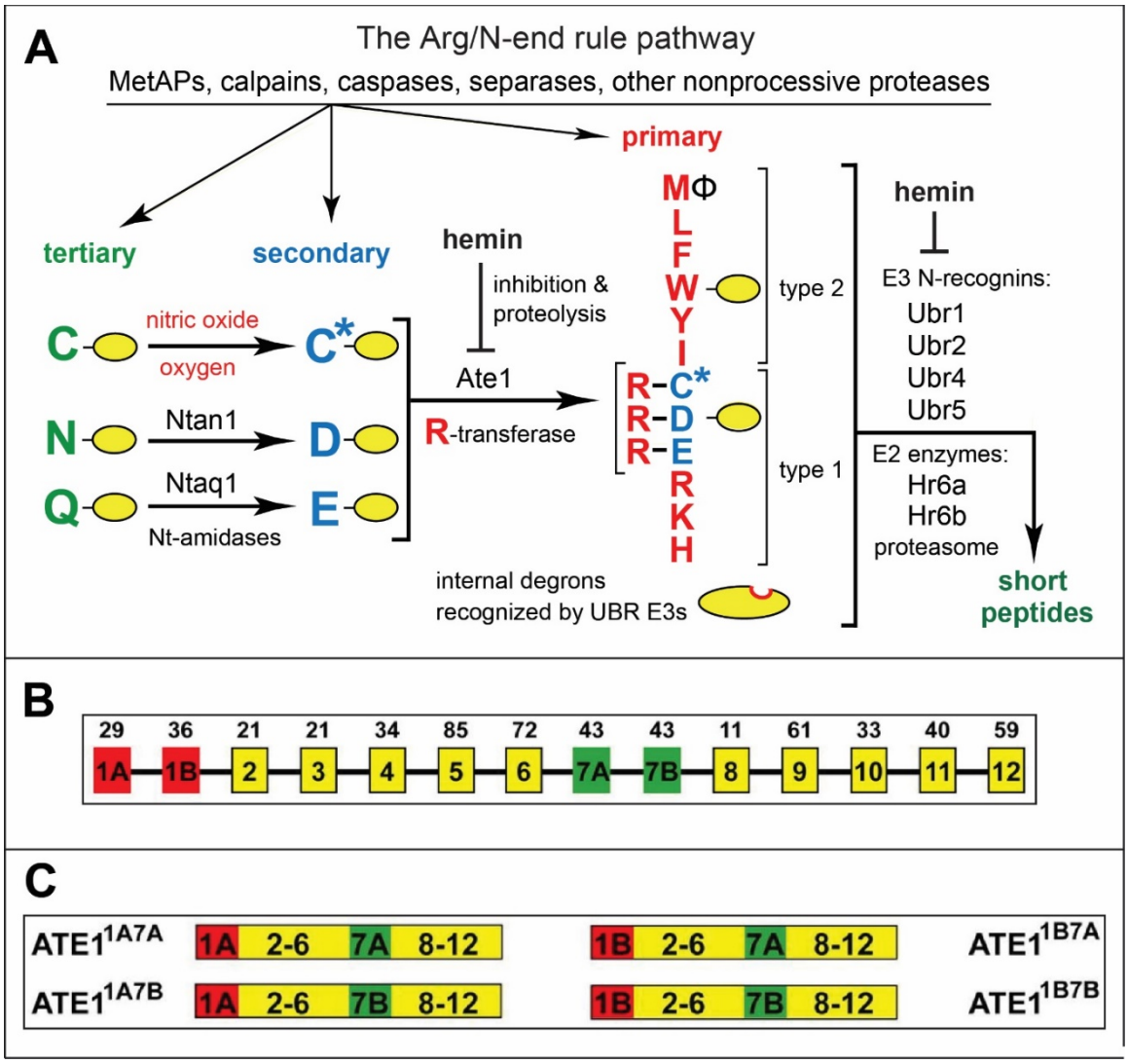


Fig. 4.1. The mammalian Arg/N-end rule pathway and *Ate1* arginyltransferase isoforms. A. The Arg/N-end rule pathway. The prefix "Arg" in the pathway's name refers to N-terminal arginylation (Nt-arginylation) of N-end rule substrates by the Ate1 arginyltransferase (R-transferase), a significant feature of this proteolytic system. The Arg/N-end rule pathway targets specific unacetylated N-terminal residues. In the yeast *S. cerevisiae*, the Arg/N-end rule pathway is mediated by the Ubr1 N-recognin, a 225 kDa RING-type E3 ubiquitin ligase and a part of the targeting apparatus comprising a complex of the Ubr1-Rad6 and Ufd4-Ubc4/5 holoenzymes. In multicellular eukaryotes several functionally overlapping E3 ubiquitin ligases (Ubr1, Ubr2, Ubr4, Ubr5) function as N-recognins of this pathway. An N-recognin binds to the "primary" destabilizing N-terminal residues Arg, Lys, His, Leu, Phe, 9 Tyr, Trp, and Ile. In contrast, the N-terminal Asn, Gln, Asp, and Glu residues (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications. These modifications include the Nt-deamidation of N-terminal Asn and Gln by the Ntan1 and Ntaq1 Nt-amidases, respectively, and the Nt-arginylation of N-terminal Asp and Glu by the Ate1 R-transferase, which can also Nt-arginylate oxidized Cys, either Cys-sulfinate or Cys-sulfonate. The latter derivatives of Cys can form in animal and plant cells through oxidation of Cys by

nitric oxide (NO) and oxygen, and also the catalysis by by N-terminal Cys-oxidases. In addition to its type-1 and type-2 binding sites that recognize, respectively, the basic and bulky hydrophobic unacetylated N-terminal residues, an N-recognin such as Ubr1 contains other substrate-binding sites as well, the ones that recognize internal (non-N-terminal) degrons. As shown in the diagram, the unacetylated Met of the N-terminal M$ motif can be recognized by both yeast and mammalian Ubr1; this capability greatly expands the substrate range of the Arg/N-end rule pathway.
B. The exons, including alternative exons, of the mouse *Ate1* gene, with deduced lengths of the corresponding peptide segments indicated on top. C. Mouse R-transferase isoforms encoded by the Ate1 mRNAs and the terminology of these isoforms (see the main text).

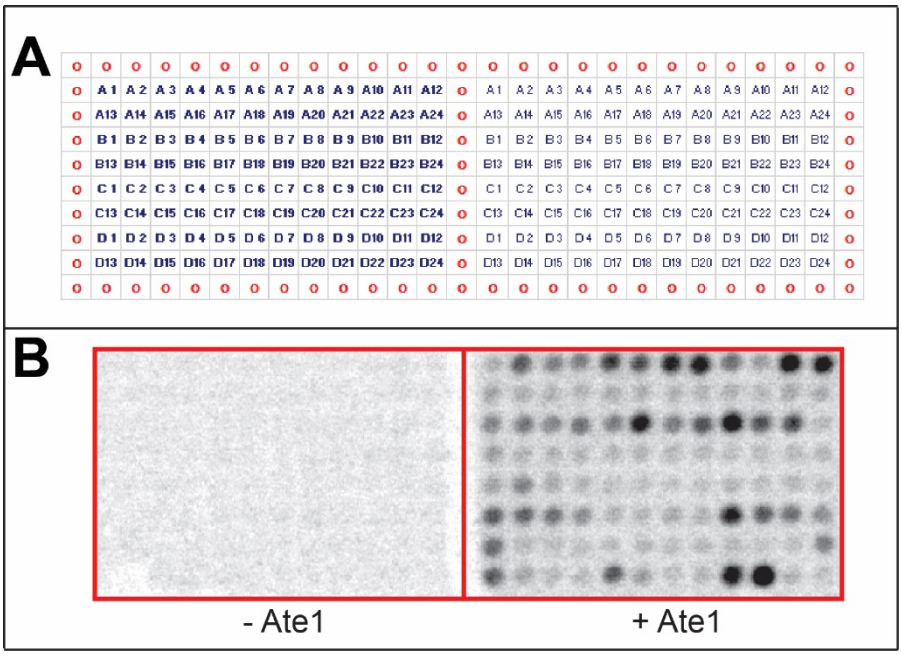


Fig. 4.2. Design of and validation of CelluSpots™-based arginylation assays. A. Layout of the CelluSpots array and assay described in panel B. Each glass slide contains a pair of identical CelluSpots peptide arrays. An individual array was "bounded" through the use of a hydrophobic pen, making it possible for each slide to be used for two separate enzymatic reactions (see *Methods*). For the identity of each of the 96 peptides, see Table 4.1 and Supplemental File 1. B. Arginylation assays performed on CelluSpots array either in the absence of purified mouse Ate1^{1B7A} R-transferase (left) or in the presence of Ate1^{1B7A} (right). See *Methods* for additional details.

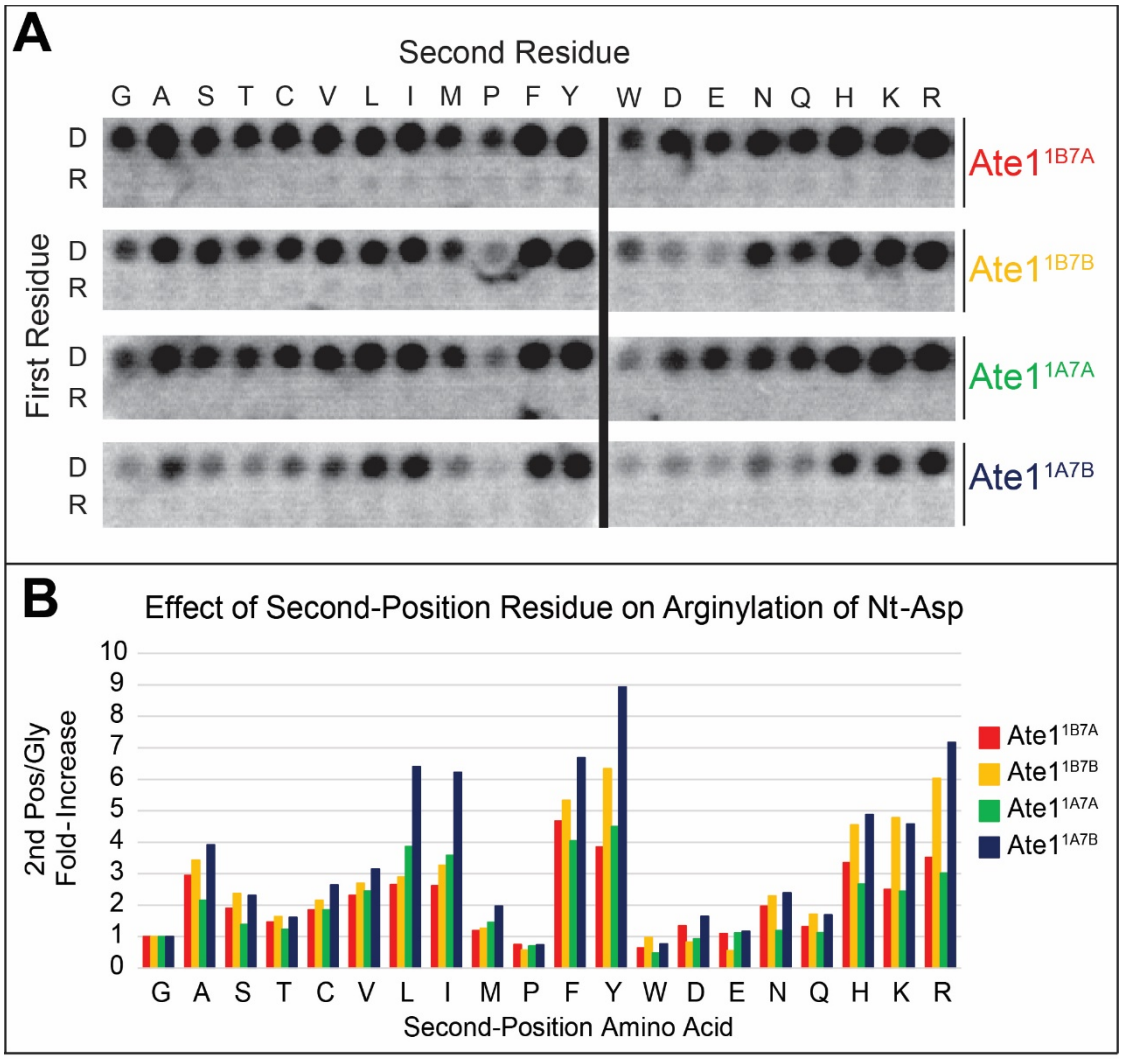


Fig. 4.3. Effect of a residue at position 2 on the arginylation activity of R-transferase. A. Effect of a residue at position 2 on the arginylation of $^{D(X)}$GSG and $^{R(X)}$GSG peptides in CelluSpots assays by each of the four major mouse Ate1 isoforms (see *Methods* and *Results* for specific notations and other details). Each Ate1 isoform was able to arginylate the full complement of $^{D(X)}$GSG peptides, whereas none of the isoforms were able to arginylate any of the N-terminal Arg-bearing $^{R(X)}$GSG control peptides. B. Quantification of data in (A) plotted as a function of Ate1 isoform and second-position residues of only the $^{D(X)}$GSG peptides.

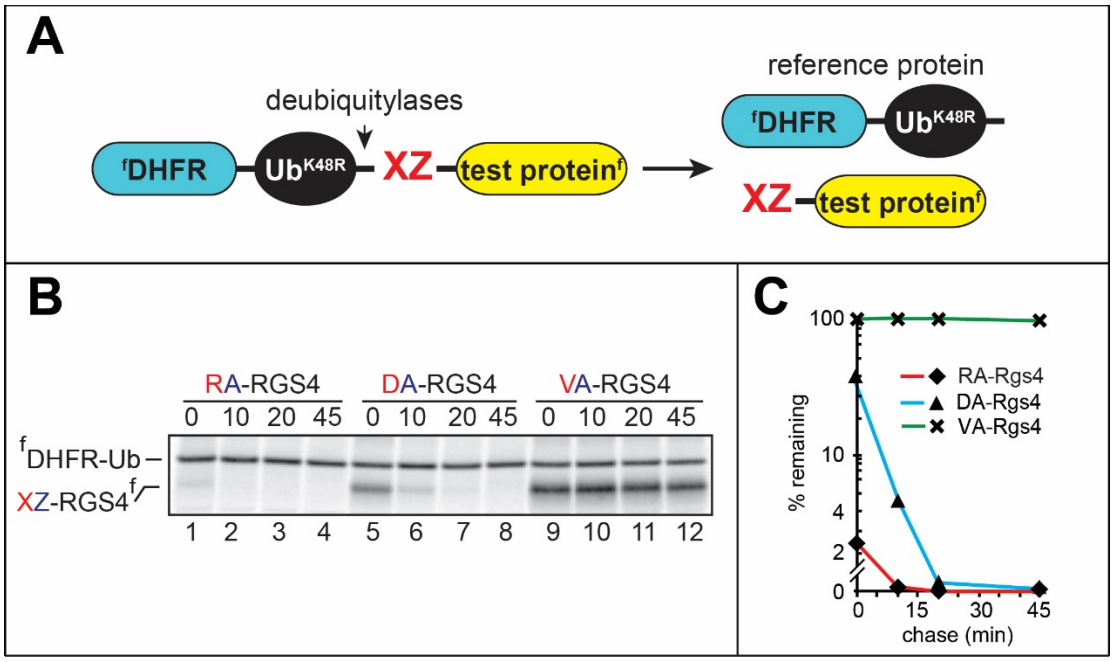


Fig. 4.4 Pulse-chase analyses of modified Rgs4, a physiological substrate of the arginylation branch of the Arg/N-end rule pathway, in rabbit reticulocyte extract. A. The Ub Reference Technique (URT) (see Results for its description) [35]. B. Pulse-chase analyses of XA-Rgs4 proteins (X=Arg, Asp, Val).The varied N-terminal residues are indicated in red. The Ala residue, in blue, replace, in these derivatives of wild-type Rgs4, its natural Lys residue at position 2. ^{RA}RGS4, ^{DA}RGS4, and ^{VA}RGS4 were expressed in rabbit reticulocyte extract and pulse-labeled with ^{35}S-Met/Cys for 10 min at 30°C, followed by a chase, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography. C. Quantification of data in B using the reference protein ^{f}DHFR-Ub^{K48R}.

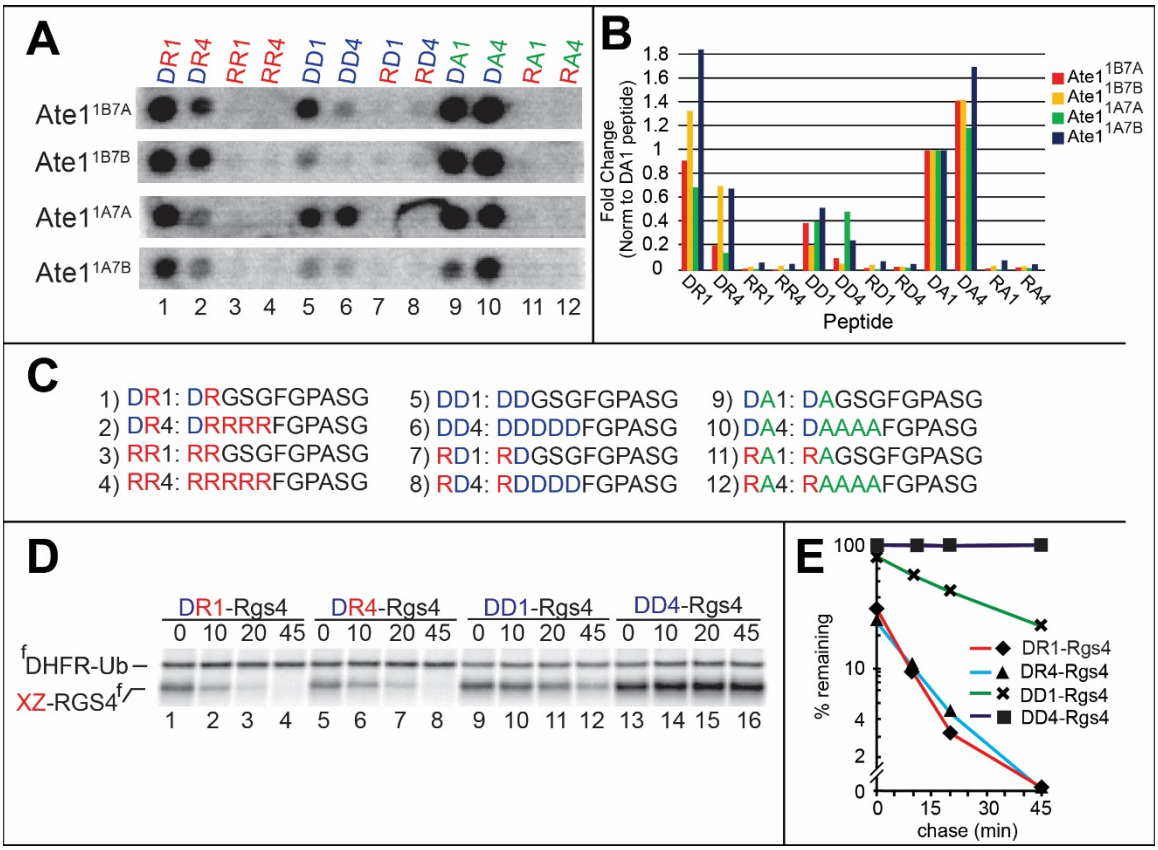


Fig. 4.5 Influence of charge clusters immediately downstream of N-terminal residue on the relative efficacy of N-terminal arginylation (Nt-arginylation) and protein degradation. A. CelluSpots arginylation assays with of GSG peptides (see Results for specific terms and notations) containing mutations within 4 residues from their N-terminal residues. B. Quantification of data in A. C. Sequences of 11-residue peptides used in A. D. Pulse-chase analyses, in reticulocyte extract, of Rgs4-based model substrates. Varied residues are indicated in red and blue (see also Results). ^{DR1}Rgs4, ^{DR4}Rgs4, ^{DD1}Rgs4, and ^{DD4}Rgs4 were expressed in rabbit reticulocyte extract and pulse-labeled with ^{35}S-Met/Cys for 10 min at 30°C, followed by a chase, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography. E. Quantification of data in D.

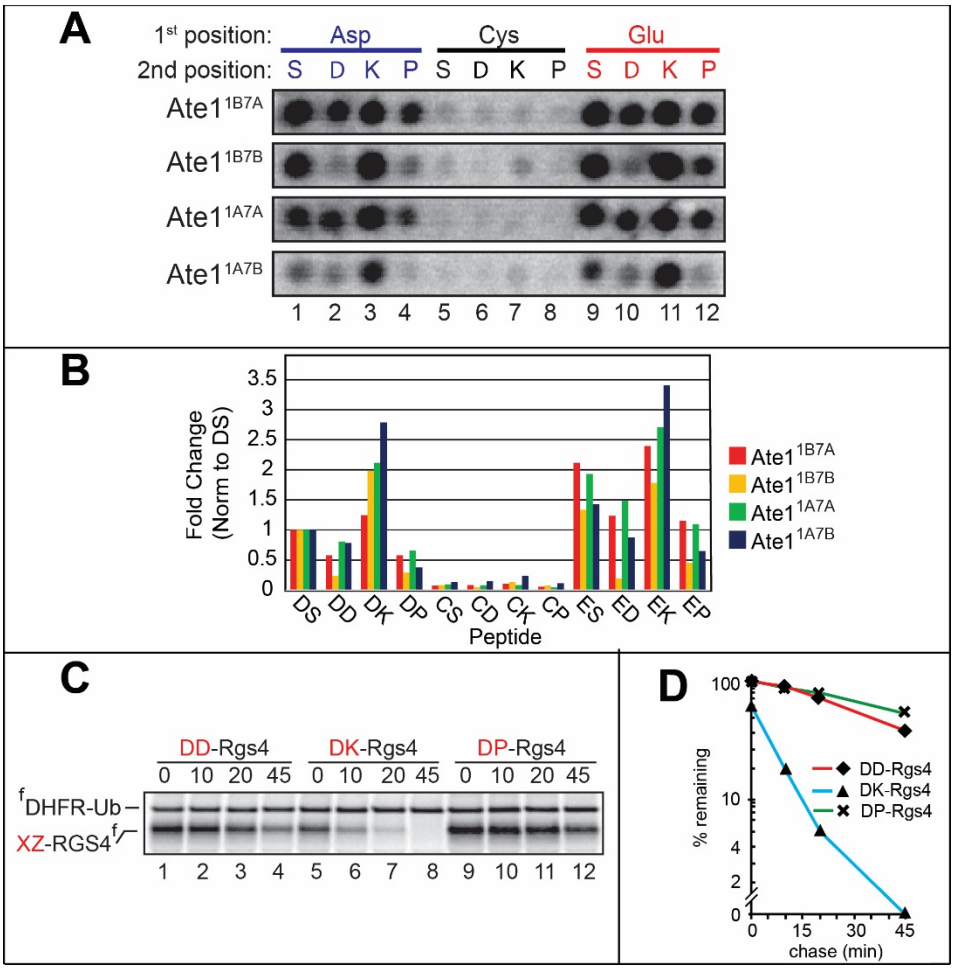


Fig. 4.6 Effect of a residue at position 2 on the relative efficacy of N-terminal arginylation. A. CelluSpots arginylation assays with GSG peptides (see Results for the terminology of 11-residue peptides) containing specific mutations at positions 1 and 2. B. Quantification of data in A. C. Degradation of Rgs4-based model substrates. Varied residues are indicated in red. ^{DD}Rgs4, ^{DK}Rgs4, and ^{DP}Rgs4 were expressed in rabbit reticulocyte extract and pulse-labeled with ^{35}S-Met/Cys for 10 min at 30°C, followed by a chase, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography. D. Quantification of data in C.

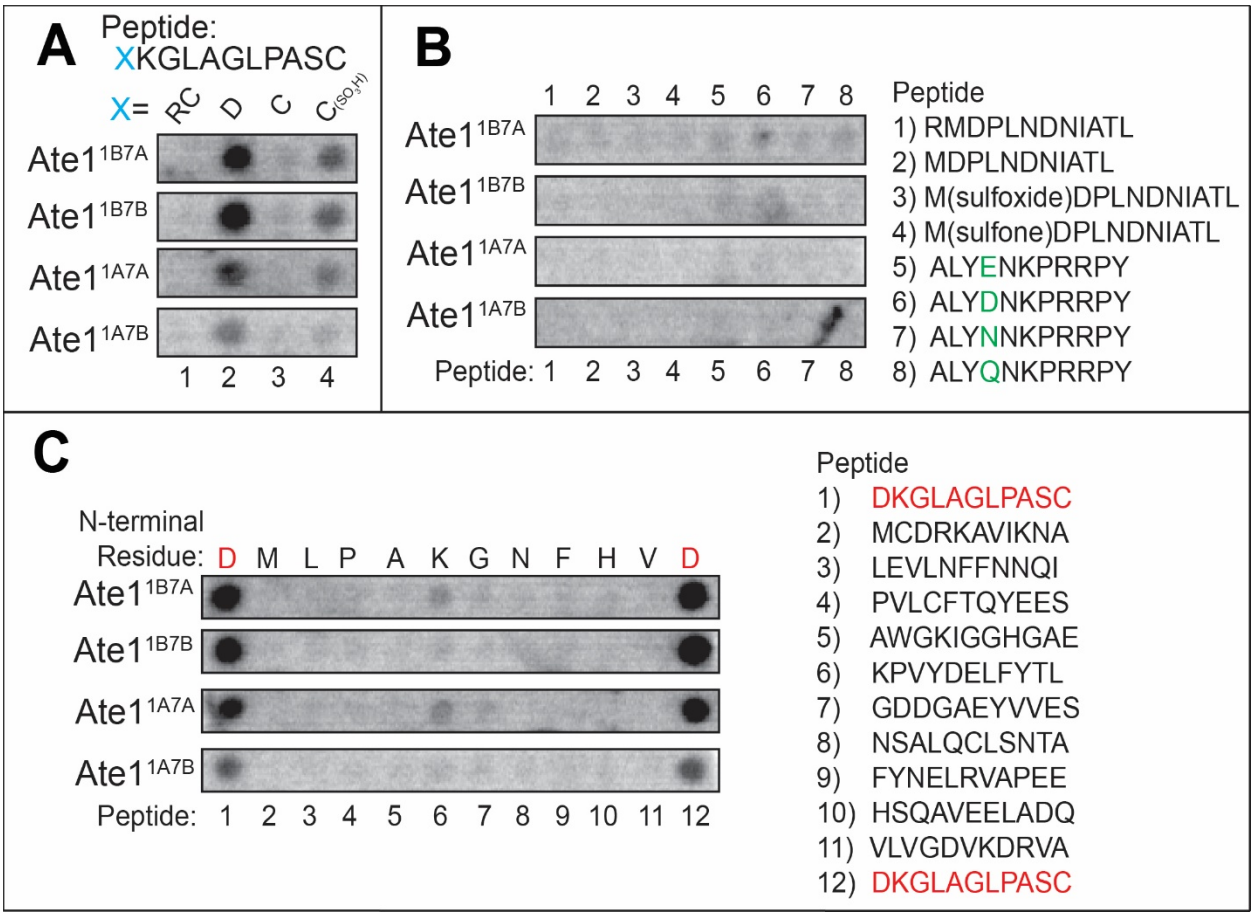


Fig. 4.7. CelluSpots arginylation assays that addressed earlier claims of protein arginylation at non-canonical residues (see Results for references and additional details). A. Arginylation assays with CelluSpots peptides derived from the first 11 residues of mouse Rgs4. These peptides bear either ArgCys, or Asp, or unmodified (unoxidized) Cys, or Cys-sulfonate at their N-termini. B. CelluSpots arginylation assays with 11-residue peptides whose amino acid sequences were derived from the Met^{589}-Leu^{599} fragment of the mouse Myh9 protein previously reported to be arginylated at $Myh9^{Met589}$ [14]. These peptides contained either N-terminal ArgMet, or unmodified N-terminal Met, or N-terminal Met sulfoxide, or N-terminal Met sulfone (peptide spots 1-4). The tested peptides included a set of otherwise identical 11-residue peptides whose sequences were identical to the N-terminal sequence neurotensin, with the indicated substitutions at the N-terminus (Ala instead of Gln) and at position 4 (see Results for additional details). C. CelluSpots arginylation assays with 11-residue peptides derived from sequences that have been previously reported to be arginylated on non-canonical residues in vivo [14]. Spots 1 and 12 are positive-control peptides derived from the N-terminus of mouse Rgs4, but containing the arginylatable N-terminal Asp.

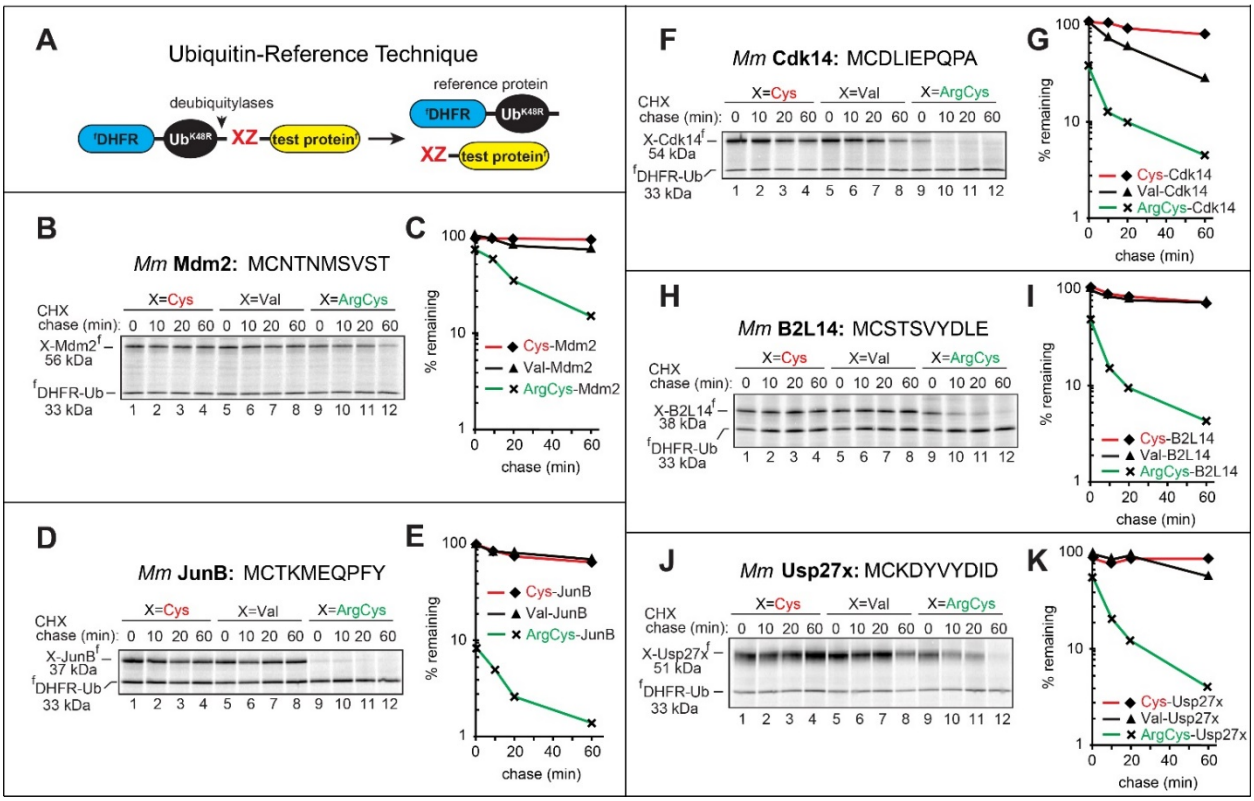


Fig. 4.8. Degradation assays with proteins whose ORFs encode N-terminal Met-Cys. A. The Ub Reference Technique (URT) (see Results for its description) [35]. B. Degradation of mouse Mdm2 and its derivatives ^{Cys}Mdm2, ^{Val}Mdm2, and ^{ArgCys}Mdm2. They were expressed in rabbit reticulocyte extract and labeled with ^{35}S-Met/Cys for 10 min at 30°C, followed by a chase, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography. C. Quantification of data B using the reference protein ^{f}DHFR-Ub^{K48R}. D. Same as B but with mouse X-JunB (X= Cys, Val, ArgCys). E. Quantification of D. F. Same as B but with mouse X-Cdk14 (X=Cys, Val, ArgCys). G. Quantification of F. H. Same as B but with mouse X-B2L14 (X=Cys, Val, ArgCys). I. Quantification of H. J. Same as B but with mouse X-Usp27x (X=Cys, Val, ArgCys). K. Quantification of J.

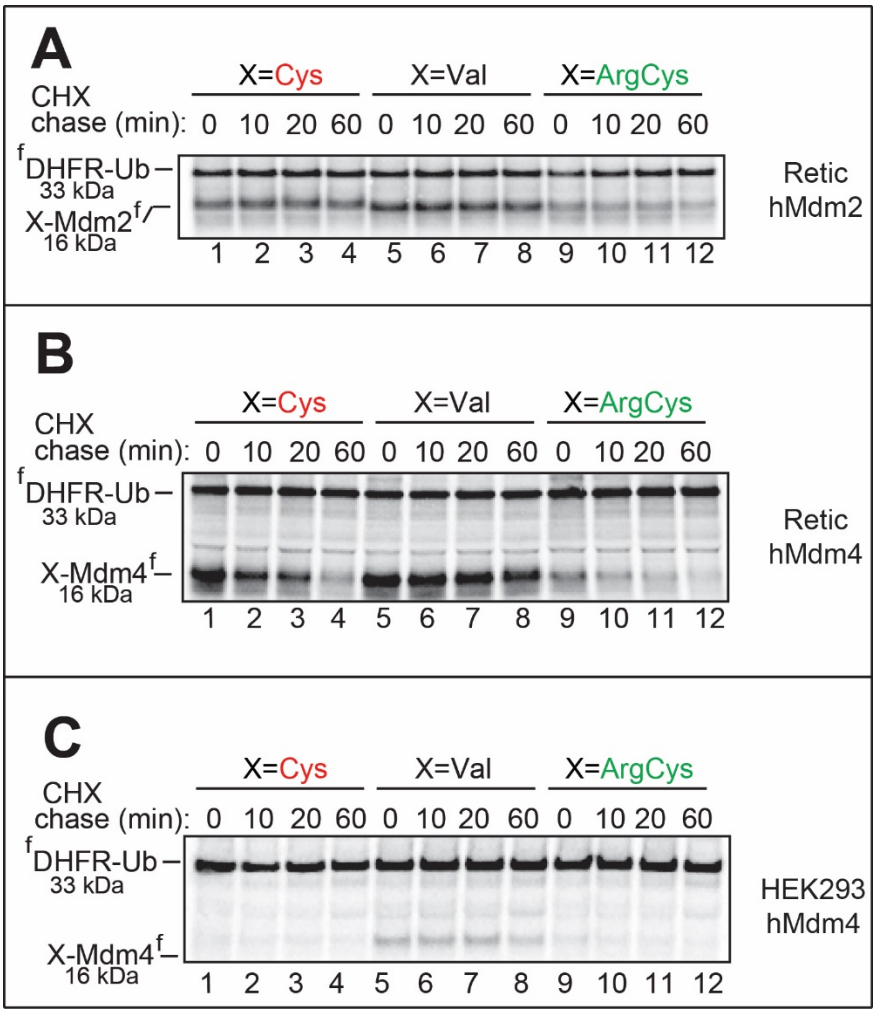


Fig. 4.9. Degradation of C-terminal fragments of human Mdm2 and Mdm4. A. Degradation assays with human Mdm2^{362} C-terminal fragment and its derivatives using the URT technique (see Fig. 4.4A and its legend) in reticulocyte extract. Cys-Mdm2^{362}, Val-Mdm2^{362}, and ArgCys-Mdm2^{362} were expressed in rabbit reticulocyte extract and labeled with ^{35}S-Met/Cys for 10 min at 30°C, followed by a chase, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography. B. Same as in A, but with human Mdm4^{362} C-terminal fragment (X= Cys, Val, ArgCys). C. In vivo degradation assays, in HEK293T human cells, and the human Mdm4^{362} C-terminal fragment and its indicated derivatives.Cys-Mdm4^{362}, Val-MDM2^{362}, and ArgCys-Mdm4^{362} were expressed in HEK293 cells by transient transfection and labeled with ^{35}S-Met/Cys for 20 min at 37°C, followed by a chase, preparation of lysates, immunoprecipitation with anti-FLAG antibody, SDS-PAGE, and autoradiography.

Table 4.1. Location and sequences of CelluSpots array peptides.

	Spot	***Sequence***	***Name***	***N-term.***
1	A 1	DGGSGFGPASG	DG-GSG	amine
2	A 2	DAGSGFGPASG	DA-GSG	amine
3	A 3	DSGSGFGPASG	DS-GSG	amine
4	A 4	DTGSGFGPASG	DT-GSG	amine
5	A 5	DCGSGFGPASG	DC-GSG	amine
6	A 6	DVGSGFGPASG	DV-GSG	amine
7	A 7	DLGSGFGPASG	DL-GSG	amine
8	A 8	DIGSGFGPASG	DI-GSGS	amine
9	A 9	DMGSGFGPASG	DM-GSG	amine
10	A10	DPGSGFGPASG	DP-GSG	amine
11	A11	DFGSGFGPASG	DF-GSG	amine
12	A12	DYGSGFGPASG	DY-GSG	amine
13	A13	RGGSGFGPASG	RG-GSG	amine
14	A14	RAGSGFGPASG	RA-GSG	amine
15	A15	RSGSGFGPASG	RS-GSG	amine
16	A16	RTGSGFGPASG	RT-GSG	amine
17	A17	RCGSGFGPASG	RC-GSG	amine
18	A18	RVGSGFGPASG	RV-GSG	amine
19	A19	RLGSGFGPASG	RL-GSG	amine
20	A20	RIGSGFGPASG	RI-GSG	amine
21	A21	RMGSGFGPASG	RM-GSG	amine
22	A22	RPGSGFGPASG	RP-GSG	amine
23	A23	RFGSGFGPASG	RF-GSG	amine
24	A24	RYGSGFGPASG	RY-GSG	amine
25	B 1	DWGSGFGPASG	DW-GSG	amine
26	B 2	DDGSGFGPASG	DD-GSG	amine
27	B 3	DEGSGFGPASG	DE-GSG	amine
28	B 4	DNGSGFGPASG	DN-GSG	amine
29	B 5	DQGSGFGPASG	DQ-GSG	amine
30	B 6	DHGSGFGPASG	DH-GSG	amine
31	B 7	DKGSGFGPASG	DK-GSG	amine
32	B 8	DRGSGFGPASG	DR-GSG	amine
33	B 9	DGPSGFGPASG	DGP3-GSG	amine
34	B10	DGGPGFGPASG	DGP4-GSG	amine
35	B11	DGGSPFGPASG	DGP5-GSG	amine
36	B12	DGGSGPGPASG	DGP6-GSG	amine

37	B13	RWGSGFGPASG	RW-GSGS	amine
38	B14	RDGSGFGPASG	RD-GSGS	amine
39	B15	REGSGFGPASG	RE-GSGS	amine
40	B16	RNGSGFGPASG	RN-GSGS	amine
41	B17	RQGSGFGPASG	RQ-GSGS	amine
42	B18	RHGSGFGPASG	RH-GSGS	amine
43	B19	RKGSGFGPASG	RK-GSGS	amine
44	B20	RRGSGFGPASG	RR-GSGS	amine
45	B21	RGPSGFGPASG	RGP3-GSG	amine
46	B22	RGGPGFGPASG	RGP4-GSG	amine
47	B23	RGGSPFGPASG	RGP5-GSG	amine
48	B24	RGGSGPGPASG	RGP6-GSG	amine
49	C 1	RCKGLAGLPASC	Rgs4 Arg	amine
50	C 2	DKGLAGLPASC	Rgs4 C2D	amine
51	C 3	CKGLAGLPASC	Rgs4	amine
52	C 4	C(sulfonic acid)KGLAGLPASC	Rgs4, Cys(sulfonic acid)	amine
53	C 5	RMDPLNDNIATL	Myh9, Arg	amine
54	C 6	MDPLNDNIATL	Myh9	amine
55	C 7	M(sulfoxide)DPLNDNIATL	Myh9, Met(sulfoxide)	amine
56	C 8	M(sulfone)DPLNDNIATL	Myh9, Met(sulfone)	amine
57	C 9	ALYENKPRRPY	NT1-11, Q1A	acetyl
58	C10	ALYDNKPRRPY	NT1-11, Q1A, E4D	acetyl
59	C11	ALYNNKPRRPY	NT1-11, Q1A, E4N	acetyl
60	C12	ALYQNKPRRPY	NT1-11, Q1A, E4Q	acetyl
61	C13	DSGSGFGPASG	GSG, G1D	amine
62	C14	DDGSGFGPASG	GSG, G1D,A2D	amine
63	C15	DKGSGFGPASG	GSG, G1D,A2K	amine
64	C16	DPGSGFGPASG	GSG, G1D,A2P	amine
65	C17	CSGSGFGPASG	GSG, G1C	amine
66	C18	CDGSGFGPASG	GSG, G1C,A2D	amine
67	C19	CKGSGFGPASG	GSG, G1C,A2K	amine
68	C20	CPGSGFGPASG	GSG, G1C,A2P	amine
69	C21	ESGSGFGPASG	GSG, G1C	amine
70	C22	EDGSGFGPASG	GSG, G1E,A2D	amine
71	C23	EKGSGFGPASG	GSG, G1E,A2K	amine
72	C24	EPGSGFGPASG	GSG, G1E,A2P	amine
73	D 1	DKGLAGLPASC	Rgs4, C2D	amine
74	D 2	MCDRKAVIKNA	dynein, cytoplasmic, light peptide	amine
75	D 3	LEVLNFFNNQI	Ras suppressor protein 1	amine

76	D 4	PVLCFTQYEES	Properdin factor, complement	amine
77	D 5	AWGKIGGHGAE	hemoglobin alpha 1 chain	amine
78	D 6	KPVYDELFYTL	EH domain containing protein MPAST2	amine
79	D 7	GDDGAEYVVES	similar to GAPDH	amine
80	D 8	NSALQCLSNTA	ubiquitin specific protease 4	amine
81	D 9	FYNELRVAPEE	actin, beta, cytoplasmic	amine
82	D10	HSQAVEELADQ	myosin, heavy polypeptide 9, non-muscle iso.	amine
83	D11	VLVGDVKDRVA	phosphoribosyl pyrophosphate synthetase	amine
84	D12	DKGLAGLPASC	Rgs4, C2D	amine
85	D13	DRGSGFGPASG	DR1-GSG	amine
86	D14	DRRRRFGPASG	DR4-GSG	amine
87	D15	RRGSGFGPASG	RR1-GSG	amine
88	D16	RRRRRFGPASG	RR4-GSG	amine
89	D17	DDGSGFGPASG	DD1-GSG	amine
90	D18	DDDDDFGPASG	DD4-GSG	amine
91	D19	RDGSGFGPASG	RD1-GSG	amine
92	D20	RDDDDFGPASG	RD4-GSG	amine
93	D21	DAGSGFGPASG	DA1-GSG	amine
94	D22	DAAAAFGPASG	DA4-GSG	amine
95	D23	RAGSGFGPASG	RA1-GSG	amine
96	D24	RAAAAFGPASG	RA4-GSG	amine

Table 4.2. Proteolytic cleavage sites, identified in the CutDB database, in specific proteins in which the Cys residue is present at the P1' position of a cleavage site.

Protease	Substrate		Substrate Species	Cleavage Sequence	Cleavage Site
caspase-3	MDM4	Mouse double minute 4 homolog	Homo sapiens	DVPD-CRRT	361-362
caspase-3	PKN2	Protein kinase n2	Homo sapiens	DITD-CPRT	117-118
caspase-3	VAV1	vav1 oncogene	Homo sapiens	DLYD-CVEN	161-162
caspase-3	CCNE1	cyclin E1 isoform 2	Homo sapiens	LDVD-CLEF	290-291
caspase-3	CCNE1	cyclin E1 isoform 2	Homo sapiens	LDVD-CLEF	275-276
caspase-3	MDM2	E3 ubiquitin-protein ligase Mdm2	Homo sapiens	DVPD-CKKT	361-362
caspase-3	GRIPAP1	GRIP1 associate protein 1 isoform 1	Homo sapiens	EERD-CHLK	580-581
caspase-3	GRIPAP1	GRIP1 associate protein 1 isoform 2	Homo sapiens	EERD-CHLK	501-502
caspase-3	SALL1	sal-like 1 isoform a	Homo sapiens	DQVD-CSDL	109-110
caspase-3	SALL1	sal-like 1 isoform b	Homo sapiens	DQVD-CSDL	109-110
caspase-3	SLC4A1AP	solute carrier family 4, member 1, adaptor protein	Homo sapiens	DYQD-CSKT	722-723
caspase-3	SSFA2	sperm specific antigen 2 isoform 1	Homo sapiens	DSTD-CLSL	517-518
caspase-3	SSFA2	sperm specific antigen 2 isoform 2	Homo sapiens	DSTD-CLSL	517-518
caspase-3	HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1 isoform A2	Homo sapiens	KLTD-CVVM	37-38
caspase-3	HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1 isoform B1	Homo sapiens	KLTD-CVVM	49-50
caspase-3	ATXN3	ataxin 3 isoform 1	Homo sapiens	DLPD-CEAD	171-172
caspase-3	TRAF1	Tnf receptor-associated factor 1	Homo sapiens	LEVD-CYRA	163-164
caspase-6	TRAF1	Tnf receptor-associated factor 1	Homo sapiens	LEVD-CYRA	163-164
caspase-8	TRAF1	Tnf receptor-associated factor 1	Homo sapiens	LEVD-CYRA	163-164
caspase-8	RIPK1	Receptor-interacting serine-threonine kinase-1	Homo sapiens	LQLD-CVAV	324-325
ADAMTS4 peptidase	BGN	biglycan preproprotein	Homo sapiens	RNMN-CIEM	186-197
cathepsin	SERPINB1	serine (or cystein) proteinase inhibitor, clase B member 1	Homo sapiens	IATF-CMLM	343-344
cathepsin B	SLPI	Secretory leukocyte peptidase inhibitor precursor	Homo sapiens	CPDT-CGIK	67-68
cathepsin S	SLPI	Secretory leukocyte peptidase inhibitor precursor	Homo sapiens	CPDT-CGIK	67-68
chymase	SERPINB1	serine (or cystein) proteinase inhibitor, clase B member 1	Homo sapiens	IATF-CMLM	343-344
chymotrypsin	SERPINB1	serine (or cystein) proteinase inhibitor, clase B member 1	Homo sapiens	IATF-CMLM	343-344
endothelin-converting enzyme 1	EDN1	endothelin-1	Homo sapiens	CVYF-CHLD	66-67
endothelin-converting enzyme 1	EDN1	endothelin-1	Homo sapiens	RSKR-CSCS	52-53
furin	EDN1	endothelin-1	Homo sapiens	RSKR-CSCS	52-53
furin	EDN1	endothelin-1	Homo sapiens	RENR-CQCA	108-109
furin	MMP2	matrix metalloproteinase 2 preproprotein	Homo sapiens	RKPR-CGNP	101-102
furin	MMP3	matrix metalloproteinase 3 preproprotein	Homo sapiens	RKPR-CGVP	91-92
granzyme M	SERPINB9	serine (or cystein) proteinase inhibitor, clase B member 9	Homo sapiens	VVAE-CCME	340-341
kallikrein h	SEMG2	Semenogelin ii precursor	Homo sapiens	RHLN-CGEK	359-360
matrix metallopeptidase-2	LGALS1	beta-galactoside-binding lectin precursor	Homo sapiens	KPGE-CLRV	16-17
neutrophil elastase	TIMP-1	metalloprotease inhibitor	Homo sapiens	MESV-CGYF	92-93
plasmin	MMP13	matrix metalloproteinase 13 preproprotein	Homo sapiens	KKPR-CGVP	95-96
procollagen C-peptidase	COL1A1	alpha 1 type 1 collagen preproprotein	Homo sapiens	EGEC-CPVC	91-92
proprotein convertase 7	EDN1	preproendothelin-1	Homo sapiens	RSKR-CSCS	52-53
S2P peptidase	ATF-6A	Cyclic-AMP-dependent transcription factor ATF-6 alpha	Homo sapiens	RRVV-CVMI	379-380
S2P peptidase	ATF-6B	Cyclic-AMP-dependent transcription factor ATF-6 beta	Homo sapiens	RKVV-CIMV	379-380
S2P peptidase	SREBF2	sterol regulatory element binding transcription factor 2	Homo sapiens	RILL-CVLT	484-485
S2P peptidase	SREBF1	sterol regulatory element binding transcription factor 1	Homo sapiens	RLAL-CTLV	490-491
thrombin	AANAT	Arylalkamine n-acetyltransferase	Homo sapiens	SEFR-CLTP	38-39

Table 4.3. Plasmids used in this study.

Plasmid	Description	Reference
pKP496	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-MCS-flag under the control of CMV promoter	Piatkov et al., 2012
pBW114	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-mB2L14-flag under the control of CMV promoter	This study
pBW115	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-mCdk14-flag under the control of CMV promoter	This study
pBW116	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-mJunB-flag under the control of CMV promoter	This study
pBW117	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-mMdm2-flag under the control of CMV promoter	This study
pBW118	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-Usp27x-flag under the control of CMV promoter	This study
pBW128	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-mJunB-flag under the control of CMV promoter	This study
pBW129	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-mJunB-flag under the control of CMV promoter	This study
pBW130	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-mMdm2-flag under the control of CMV promoter	This study
pBW131	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-mMdm2-flag under the control of CMV promoter	This study
pBW165	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-mB2L14-flag under the control of CMV promoter	This study
pBW166	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-mB2L14-flag under the control of CMV promoter	This study
pBW167	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-mCdk14-flag under the control of CMV promoter	This study
pBW168	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-mCdk14-flag under the control of CMV promoter	This study
pBW169	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-Usp27x-flag under the control of CMV promoter	This study
pBW170	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-Usp27x-flag under the control of CMV promoter	This study
pBW448	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys362-hMDM4-flag under the control of CMV promoter	This study
pBW449	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val362-hMDM4-flag under the control of CMV promoter	This study
pBW450	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys362-hMDM4-flag under the control of CMV promoter	This study
pBW463	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Cys2-hMDM362-flag under the control of CMV promoter	This study
pBW464	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-Val2-hMDM362-flag under the control of CMV promoter	This study
pBW465	Amp^R; Neo^R; pcDNA3-based plasmid encoding flag-DHFR-ha-Ub-ArgCys2-hMDM362-flag under the control of CMV promoter	This study

Table 4.4. PCR primers used in this study.

Primer	Sequence
BW300	5'-TCCGCGGAGGATGCAGCACCAGTGTGTATGA-3'
BW301	5'-TCCGCGGAGGAGTGAGCACCAGTGTGTATGA-3'
BW302	5'-TCCGCGGAGGAAGATGCAGCACCAGTGTGTATG-3'
BW303	5'-ATTATCGATGTCTACTTCTTCATGTGAGATCCCA-3'
BW304	5'-TCCGCGGAGGATGCGACCTCATTGAACCG-3'
BW305	5'-TCCGCGGAGGAGTGGACCTCATTGAACCG-3'
BW306	5'-TCCGCGGAGGAAGATGCGACCTCATTGAAC-3'
BW307	5'-ATTATCGATGTGTTTGCTGTTCGATAGGCT-3'
BW312	5'-TCCGCGGAGGATGCACGAAAATGGAACAGC-3'
BW313	5'-TCCGCGGAGGAGTGACGAAAATGGAACAGC-3'
BW314	5'-TCCGCGGAGGAAGATGCACGAAAATGGAAC-3'
BW315	5'-ATTATCGATGAAGGCGTGTCCCTTGAC-3'
BW316	5'-TCCGCGGAGGATGCAATACCAACATGTCTGTGTC-3'
BW317	5'-TCCGCGGAGGAGTGAATACCAACATGTCTGTGTC-3'
BW318	5'-TCCGCGGAGGAAGATGCAATACCAACATGTCTG-3'
BW319	5'-ATTATCGATGTTGAAGTAAGTTAGCACAATCATTTG-3'
BW320	5'-TCCGCGGAGGATGTAAAGACTATGTGTATGAC-3'
BW321	5'-TCCGCGGAGGAGTGAAAGACTATGTGTATGAC-3'
BW322	5'-TCCGCGGAGGAAGATGTAAAGACTATGTGTATG-3'
BW323	5'-ATTATCGATGTAGGCTTGTGGAGTCATTTC-3'
BW841	5'-TTTACCGCGGAGGATGTCGAAGAACCATTTCGG-3'
BW842	5'-TTTACCGCGGAGGAGTGCGAAGAACCATTTCGG-3'
BW843	5'-TTTACCGCGGAGGAAGATGTCGAAGAACCATTTCGG-3'
BW844	5'-TTTAATCGATTGCTATAAAAACCTTAATAACCAGCTG-3'
BW860	5'-TTTACCGCGGAGGATGTAAAAAAACTATAGTGAATGATTCCAG-3'
BW861	5'-TTTACCGCGGAGGAGTGAAAAAAACTATAGTGAATGATTCCAG-3'
BW862	5'-TTTACCGCGGAGGAAGATGTAAAAAAACTATAGTGAATGATTCCAG-3'
BW863	5'-ATTATCGATGGGGAAATAAGTTAGCACAATCA-3'

CHAPTER 5:

DEGRADATION OF THE SEPARASE-CLEAVED REC8, A MEIOTIC COHESIN SUBUNIT, BY THE N-END RULE PATHWAY

From Liu, Y.-J., Liu, C., Chang, Z., Wadas, B., Brower, C.S., Song, Z.-H., Xu, Z.-L., Shang, Y.-L., Liu, W.-X., Wang, L.-N., Dong, W., Varshavsky, A., Hu, R.-G., and Li, W. (2016) *J. Biol. Chem.* 291(14):7426-7438

ABSTRACT

The Ate1 arginyltransferase (R-transferase) is a component of the N-end rule pathway, which recognizes proteins containing N-terminal degradation signals called N-degrons, polyubiquitylates these proteins and thereby causes their degradation by the proteasome. Ate1 arginylates N-terminal Asp, Glu or (oxidized) Cys. The resulting N-terminal Arg is recognized by ubiquitin ligases of the N-end rule pathway. In the yeast *Saccharomyces cerevisiae*, the separase-mediated cleavage of the Scc1/Rad21/Mcd1 cohesin subunit generates a C-terminal fragment that bears N-terminal Arg and is destroyed by the N-end end rule pathway without requirement for arginylation. In contrast, the separase-mediated cleavage of Rec8, the mammalian meiotic cohesin subunit, yields a fragment bearing N-terminal Glu, a substrate of the Ate1 R-transferase. Here we constructed and used a germ cells-confined *Ate1*$^{-/-}$ mouse strain to analyze the separase-generated C-terminal fragment of Rec8. We show that this fragment is a short-lived N-end rule substrate, that its degradation requires N-terminal arginylation, and that male *Ate1*$^{-/-}$ mice are nearly infertile, owing to massive apoptotic death of *Ate1*$^{-/-}$ spermatocytes during the metaphase of meiosis I. These effects of *Ate1* ablation are inferred to be caused, at least in part, by the failure to destroy the C-terminal fragment of Rec8 in the absence of N-terminal arginylation.

INTRODUCTION

The N-end rule pathway is a set of proteolytic systems whose unifying feature is the ability to recognize and polyubiquitylate proteins containing degradation signals called N-degrons, thereby causing degradation of these proteins by the proteasome (Fig.

1A, B) (1-8). The main determinant of an N-degron is either an unmodified or chemically modified destabilizing N-terminal residue of a protein. The identity of the next residue, at position 2, is often important as well. A second determinant of an N-degron is a protein's internal Lys residue. It functions as the site of a protein's polyubiquitylation and tends to be located in a conformationally disordered region (4,9,10). Recognition components of the N-end rule pathway are called N-recognins. In eukaryotes, N-recognins are E3 ubiquitin (Ub) ligases that can target N-degrons (Fig. 1A, B). Bacteria also contain the N-end rule pathway, but Ub-independent versions of it (11-16).

Regulated degradation of proteins and their natural fragments by the N-end rule pathway has been shown to mediate a strikingly broad range of biological functions, including the sensing of heme, nitric oxide (NO), oxygen, and short peptides; the control, through subunit-selective degradation, of the input stoichiometries of subunits in oligomeric protein complexes; the elimination of misfolded or otherwise abnormal proteins; the degradation of specific proteins after their retrotranslocation to the cytosol from mitochondria or other membrane-enclosed compartments; the regulation of apoptosis and repression of neurodegeneration; the regulation of DNA repair, transcription, replication, and chromosome cohesion/segregation; the regulation of G proteins, autophagy, peptide import, meiosis, immunity, fat metabolism, cell migration, actin filaments, cardiovascular development, spermatogenesis, and neurogenesis; the functioning of adult organs, including the brain, muscle, testis, and pancreas; and the regulation of leaf and shoot development, leaf senescence, and many other processes in plants (Figs. 1A, B) ((4-7) and references therein).

In eukaryotes, the N-end rule pathway consists of two branches. One branch, called the Ac/N-end rule pathway, targets proteins for degradation through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues (Fig. 1B) (2,3,17-19). Degradation signals and E3 Ub ligases of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively. Nt-acetylation of cellular proteins is apparently irreversible, in contrast to acetylation-deacetylation of proteins' internal Lys residues. About 90% of human proteins are cotranslationally Nt-acetylated by ribosome-associated Nt-acetylases (20). Posttranslational Nt-acetylation is known to occur as well. Many, possibly most, Nt-acetylated proteins contain Ac/N-degrons (Fig. 1B).

The pathway's other branch, called the Arg/N-end rule pathway, targets specific unacetylated N-terminal residues (Fig. 1A) (3,21-25). The "primary" destabilizing N-terminal residues Arg, Lys, His, Leu, Phe, Tyr, Trp, and Ile are directly recognized by N-recognins. The unacetylated N-terminal Met, if it is followed by a bulky hydrophobic ($) residue, also acts as a primary destabilizing residue (Fig. 1A) (3). In contrast, the unacetylated N-terminal Asn, Gln, Asp, and Glu (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications, which include N-terminal deamidation (Nt-deamidation) of Asn and Gln and Nt-arginylation of Asp, Glu and oxidized Cys (Fig. 1A) (4-6,26). In the yeast *S. cerevisiae*, the Arg/N-end rule pathway is mediated by the Ubr1 N-recognin, a 225 kDa RING-type E3 Ub ligase and a part of the targeting complex comprising the Ubr1-Rad6 and Ufd4-Ubc4/5 E2-E3 holoenzymes (4,27). In multicellular eukaryotes, several E3 Ub ligases, including Ubr1, function as N-recognins of the Arg/N-end rule pathway (Fig. 1A).

Nt-arginylation is mediated by the *Ate1*-encoded arginyltransferase (Arg-tRNA-protein transferase; R-transferase), a component of the Arg/N-end rule pathway and one subject of the present study (Fig. 1A) (28-36). Alternative splicing of mouse *Ate1* pre-mRNAs yields at least six R-transferase isoforms, which differ in their Nt-arginylation activity (28,31). R-transferases are sequelogous (similar in sequence (37)) throughout most of their ~60 kDa spans from fungi to mammals (4). R-transferase can arginylate not only N-terminal Asp and Glu but also N-terminal Cys, if it has been oxidized to Cys-sulfinate or Cys-sulfonate, through reactions mediated by NO, oxygen, and N-terminal Cys-oxidases (5,6,30). Consequently, the Arg/N-end rule pathway functions as a sensor of NO and oxygen, through the conditional oxidation of N-terminal Cys in proteins such as the Rgs4, Rgs5, and Rgs16 regulators of G proteins in mammals (30,38) and specific transcriptional regulators in plants (reviewed in (5,6)). By now, more than 200 distinct proteins (including natural protein fragments) have been either shown or predicted to be Nt-arginylated (21,22,39-41). Many, possibly most, Nt-arginylated proteins are conditionally or constitutively short-lived substrates of the Arg/N-end rule pathway (Fig. 1A).

Identified and predicted substrates of the Arg/N-end rule pathway include cohesin subunits of the kleisin family. Rings of oligomeric cohesin keep sister chromatids together through a topological confinement (42-53). Physiological roles of cohesins include DNA replication, cohesion/segregation, repair, transcription; and control of apoptosis (54-61). In *S. cerevisiae*, the Scc1/Rad21/Mcd1 cohesin subunit is cleaved at a specific site, late in mitosis, by the nonprocessive Esp1 protease called separase, allowing the release of sister chromatids and the transition from metaphase to anaphase (48,62-64).

In mammals and other multicellular eukaryotes, the bulk of cohesin-mediated confinement of sister chromatids is removed during prophase through openings of cohesin rings that do not involve proteolytic cuts. Sister chromatids continue to be held together until the end of metaphase, to a large extent through still intact cohesin rings at the centromeres. Once the bipolar spindle attachment of chromosomes is achieved at the end of metaphase, the activated separase cleaves the mammalian Rad21 subunit (a sequelog of yeast Scc1) in the closed cohesin complexes, resulting in their opening and allowing the separation of sister chromatids (45,48,58,65). At least in mammals, the Rad21 cohesin subunit can be cleaved in vivo not only by separase but also by calpain-1 (a Ca^{2+}-activated protease) (58), and by caspases as well (56,57).

During meiosis, in which cohesins also play essential roles, the mitosis-specific cohesin subunit Rad21 is replaced by the sequelogous (similar in sequence (37)), meiosis-specific Rec8 subunit (43,59,66-70). (A second meiosis-specific kleisin-type subunit, called Rad21L, is expressed during early meiosis and disappears afterward (71).) Meiotic DNA replication is followed by two rounds of cell division to produce four haploid daughter cells. During the first meiotic cell division cycle (meiosis I), replicated homologous chromosomes pair and recombine with each other. The pairs of modified (recombined) homologous chromosomes are separated at the end of meiosis I, yielding two diploid daughter cells (72,73). During meiosis II, the replicated sister chromatids of each chromosome are pulled apart to produce haploid daughter cells. In the testis of male mice, meiosis I and II take place in meiotic spermatocytes, leading to the formation of haploid spermatids and later mature sperm cells (43,59,66,74).

In *S. cerevisiae*, the separase-generated C-terminal fragment of the Scc1 cohesin subunit bears N-terminal Arg. This fragment of Scc1 forms late in mitosis upon the activation of separase, and is rapidly destroyed by the Arg/N-end end rule pathway (62). A failure to eliminate this (normally short-lived) Scc1 fragment, for example, in a *ubr1"* mutant that lacks the Arg/N-end rule pathway, results in chromosome instability (62). The C-terminal fragment of Scc1 retains a physical affinity for the rest of cohesin complex (62). Therefore the plausible (but not proven) cause of chromosome instability in *ubr1"* cells is an interference with cohesin mechanics by the metabolically stabilized, cohesin-bound C-terminal fragment of Scc1. Since the yeast Scc1 fragment bears N-terminal Arg, the fragment's degradation by the Arg/N-end rule pathway does not require Nt-arginylation (62). In mammals, however, the separase-generated C-terminal fragments of the Rad21 subunit of mitotic cohesin and the Rec8 subunit of meiotic cohesin bear N-terminal Glu, a substrate of the Ate1 R-transferase (Fig. 1A, C) (47,75,76).

In the present work, we constructed an *Ate1*$^{-/-}$ mouse strain in which the ablation of *Ate1* was confined to germ cells. We show that the separase-generated C-terminal fragment of Rec8, a subunit of meiotic cohesin, is a short-lived physiological substrate of the Arg/N-end rule pathway, and that the degradation of this Rec8 fragment requires its Nt-arginylation. These and related results suggest that a failure to destroy this fragment in arginylation-lacking spermatocytes of *Ate1*$^{-/-}$ mice contributes to a greatly reduced male fertility of these mice, owing to the observed arrest and apoptotic death of *Ate1*$^{-/-}$ spermatocytes at the end of meiosis I. These results expand the already large set of functions of the Arg/N-end rule pathway. Together with earlier data about *Ate1*$^{-/-}$ and

Ubr2$^{-/-}$ mice (33,77), the present findings also indicate that perturbations of the Arg/N-end rule pathway may be among the causes of infertility in humans.

EXPERIMENTAL PROCEDURES

Mutant Mouse Strains – Conditional (*cre-lox*-based) *Ate1*$^{flox/flox}$ mice were described (33). *Tnap-Cre* mice, in which the Cre recombinase is selectively expressed in primordial germ cells (78), were purchased from Jackson Laboratory (Bar Harbor, ME). Germ-cell specific *Ate1*$^{-/-}$ mouse strains were constructed in the present study through matings of *Ate1*$^{flox/flox}$ mice (33) and *Tnap-Cre* mice (78). All animal experiments were approved by the Animal Research Panel of the Committee on Research Practice of the University of the Chinese Academy of Science.

Antibodies – Mouse anti-Sycp3 monoclonal antibody (SC-74569) and mouse anti-PLZF monoclonal antibody (SC-28319) were from Santa Cruz Biotechnology (Dallas, TX). Rabbit anti-Sycp3 polyclonal antibody (ab150292) and Rabbit anti-Rec8 polyclonal antibody were from Abcam (Cambridge, MA). Mouse anti-MLH1 monoclonal antibody (51-1327GR) was from BD Pharmingen (San Jose, CA). Rabbit anti-Sycp1 polyclonal antibody-Cy5 (NB300-228c) was from Novus Biologicals (Littleton, CO). Mouse anti-Cyclin B1 monoclonal antibody (MA5-14319) was from Thermo Scientific (Waltham, MA). FITC-conjugated mouse monoclonal anti-±-tubulin antibody (F2168) was from Sigma-Aldrich (St. Louis, MO). FITC-conjugated goat anti-rabbit antibody, and TRITC-conjugated goat anti-mouse antibody, as well as horseradish peroxidase (HRP)-conjugated secondary antibody were from Zhong Shan Jin Qiao (Beijing, China). Alexa Fluor 680-conjugated goat anti-mouse antibody and Alexa Fluor 680-conjugated goat anti-rabbit antibody were from Invitrogen (Carlsbad, CA).

Male Fertility Assays – Breeding assays with wild-type and *Ate1*$^{-/-}$ male mice were carried out previously described (79). Briefly, each examined male mouse (8-9 weeks old) was caged with two wild-type (CDI strain) female mice (6-8 weeks old), and their vaginal plugs were checked every morning. The number of pups produced by each pregnant female was counted within a week after birth. Each male was tested through six cycles of this breeding assay.

Epididymal Sperm Count – The *cauda epididymis* was dissected from adult mice. Sperm was released by cutting the cauda epididymis into pieces, placing them in 1 ml of phosphate-buffered saline (PBS), and incubating for 10 min at 37°C. Thereafter, 10-µl samples were transferred to a hemocytometer for counting sperm cells.

Immunoblotting and Related Procedures – Extracts from testes were prepared using a Dounce homogenizer in a cold high-salt RIPA-like buffer (0.5 M NaCl, 1% Nonidet-P40, 1% Na-deoxycholate, 0.1% SDS, 2 mM Na-EDTA, 25 mM Tris-HCl, pH 7.6) supplemented with 1mM phenylmethylsulfonyl fluoride (PMSF) and a protein inhibitor cocktail (Roche, Basel, Switzerland, 04693132001). Extracts were centrifuged at 13,000*g* for 15 min. Samples of supernatants were fractionated by SDS-PAGE (NuPAGE 4-12% Bis-Tris Gel; Invitrogen) and electro-transferred to a nitrocellulose membrane, followed by incubation for 2-4 hrs at room temperature (RT) with a primary antibody, a further incubation with the (HRP)-conjugated secondary antibody, washes, and the detection of immunoblotting patterns using chemilumenescence, largely as described (18,80).

The Ubiquitin Reference Technique (URT) and Analyses of Rec8 in Reticulocyte Extract – The mouse *Rec8* open reading frame (ORF) was amplified by PCR, using

cDNA from GE Healthcare Dharmacon (Clone ID: 6335959) (Pittsburgh PA). The resulting DNA fragments were cut with *Sac*II and *Cla*I and cloned into *Sac*II/*Cla*I-cut pKP496, yielding the plasmids pBW442, pBW443, and pBW444, respectively. pKP496 was an Amp^R, Neo^R, pcDNA3-based plasmid encoding flag-DHFR-ha-Ub^{K48R}-MCS (multiple cloning site)-flag under the control of the P_{CMV} promoter (21). pBW442, pBW443, and pBW444 expressed, respectively, ^{f}DHFR-ha-Ub^{K48R}-Glu^{455}-$Rec8^f$, ^{f}DHFR-ha-Ub^{K48R}-Val^{455}-$Rec8^f$, and ^{f}DHFR-ha-Ub^{K48R}-Arg-Glu^{455}-$Rec8^f$ from the P_{CMV} promoter. The plasmid pcDNA3MR8, expressing ^{f}DHFR-ha-Ub^{K48R}-Met^{206}-$Rec8^f$, was constructed similarly.

Rabbit reticulocyte-based degradation assays were carried out using the TNT T7 Coupled Transcription/Translation System (Promega, Madison, WI), largely as previously described (21-23). Nascent proteins translated in the extract were pulse-labeled with ^{35}S-L-methionine (0.55 mCi/mL, 1,000 Ci/mmol, MP Biomedicals, Santa Ana, CA) for 10 min at 30°C in the total volume of 30 µl. Labeling was quenched by the addition of 0.1 mg/ml cycloheximide, and 5 mM unlabeled methionine, bringing the final reaction volume to 40 µl. Samples of 10 µl were removed at the indicated time points and the reaction was terminated by the addition of 80 µl of TSD buffer (1% SDS, 5 mM dithiothreitol (DTT), 50 mM Tris-HCl, pH 7.4), snap-frozen in liquid nitrogen, and stored at -80°C until use. Following the collection of all time points, samples were heated at 95°C for 10 min, then diluted with 1 ml of TNN buffer (0.5% NP40, 0.25 M NaCl, 5 mM Na-EDTA, 50 mM Tris-HCl, pH 7.4) containing 1X complete Protease Inhibitor Cocktail (Roche), and immunoprecipitated using 5 µl of anti-flag M2 magnetic beads (Sigma-Aldrich). Samples were incubated with rotation at 4°C for 4 hrs, followed by three

washes in TNN buffer, a wash in 10 mM Tris-HCl, pH 8.5, and resuspension in 20 µl of SDS-sample buffer. Samples were heated at 95°C for 5 min, and fractionated SDS-10% PAGE, followed by autoradiography and quantification, using Typhoon-9500 Imager and ImageQuant (GE Healthcare).

Reticulocyte extract-based assays in which $Rec8^{f}$, its derivatives, and the $^{f}DHFR\text{-}ha\text{-}Ub^{K48R}$ reference protein were detected by immunoblotting were carried similarly, except that no ^{35}S-Met/Cys labeling was involved. The synthesis-deubiquitylation-degradation of URT-based test fusions was allowed to proceed for 1 hr at 30°C either in the absence of added dipeptides, or in the presence of either Arg-Ala (1 mM) or Ala-Arg (1 mM), followed by SDS-PAGE and immunoblotting with anti-flag antibody (Fig. 5E).

Histology and Immunohistochemistry – For histological analyses, testes and epididymis were fixed, after dissection, in 4% paraformaldehyde (formaldehyde, HCHO) (Solarbio, Beijing, China, P1110) overnight at 4°C, followed by standard procedures (79). Briefly, tissues were dehydrated through a series of ethanol washes and thereafter embedded in paraffin wax. Sections (5 µm thick) were produced using a microtome. Deparaffinized and rehydrated sections were stained with hematoxylin and eosin for histological observations. For immunohistochemical assays, deparaffinized and rehydrated sections were rinsed 3 times at room temperature (RT) in PBS (pH 7.4), followed by antigen retrieval by boiling for 15 min in 10 mM Na-citrate, pH 6.0. Sections were then incubated for 10 min at RT with 3% H_2O_2, followed by blocking of each section by incubating them for 30 min at RT in 5% bovine serum albumin (BSA; Sigma). Sections were incubated with a primary antibody at 4°C overnight, followed by

incubation with an HRP-conjugated secondary antibody at 37°C for 1 hr. Sections were then stained with 3, 3'-diaminobenzidine (DAB) according to manufacturer's instructions (Zhong Shan Jin Qiao, Beijing, China, ZL1-9018), and nuclei were stained with hematoxylin. Negative controls were processed identically but without the primary antibody. Sections were examined using a Nikon 80*i* inverted microscope with a charge-coupled camera.

Spreads of Spermatocyte Nuclei – They were prepared as previously described (81), with modifications. Briefly, testes were washed in PBS after dissection from mice. The tunicae were removed and adherent extratubular tissues were removed by rinsing the seminiferous tubules with PBS at room temperature. The tubules were placed in a hypotonic extraction buffer (50 mM sucrose, 17 mM Na-citrate, 0.5mM dithiothreitol (DTT), 0.5mM phenylmethylsulphonyl fluoride (PMSF), 3 mM Tris-HCl, pH 8.2) for 30-60 min. Thereafter approximately 1 inch of tubule was shredded to pieces by fine-tipped forceps in 20 µl of 0.1 M sucrose (pH adjusted to pH 8.2 by NaOH) on a clean glass slide. Another 20 µl of sucrose solution was then added, and a slightly cloudy suspension was prepared, using a 20-µl pipettor. Residual tubular tissues were removed. The suspension was transferred onto a new precoated (3-aminopropyl-triethoxysilane (Zhong Shan Jin Qiao, Beijing, China, ZL1-9002)) glass slides, each of them containing on the surface 0.1 ml of freshly made (and filtered through a 0.22 µm Rephile filter (Rephile, Shanghai, China, RJP3222SH)) solution of 1% paraformaldehyde (HCHO), 0.15% Triton-X100 (with pH adjusted to 9.2 using 10 mM Na-borate, pH 9.2). Each slide was gently rocked to mix the initial suspension with HCHO solution, followed by drying for at least 2 hr in a closed box with high humidity. To stain the resulting nuclei spreads,

slides were washed with 0.4% Photoflo (Kodak, Rochester, NY) 3 times and with PBS 3 times, then blocked in 5% BSA for 1 hr, incubated with a primary antibody in 1.5% BSA and 0.3% Triton-X100 overnight at 4°C, then incubated with secondary antibody in PBS for 1 hr at 37°C. The resulting slides were washed in PBS, and nuclei were stained with 4,6-diamidino-2-phenylindole (DAPI). Slides were examined using the LSM 780/710 microscope (Zeiss).

TUNEL Assays – They were carried out using the In Situ Cell Death Detection Kit (Roche Diagnostics, Basel, Switzerland, 11684795910) as described by us previously(82). Briefly, sections of testis were heated at 60°C for 2 hrs followed by washing in xylene and rehydration through a graded series of washes with ethanol and double-distilled water. Thereafter the sections were treated with proteinase K for 15 min at RT and rinsed twice with PBS. After adding the TUNEL reaction mixture, slides were incubated in a humidified atmosphere for 60 min at 37 °C in the dark, following by staining with DAPI (82). All experiments were repeated at least three times, with standard deviations shown.

RESULTS

Germ Cell-specific Ablation of the Ate1 R-transferase Strongly Decreases Fertility of Male Mice – Probing sections of mouse testis with affinity-purified antibody to mouse Ate1 indicated the presence of the Ate1 R-transferase in the testis, particularly in spermatocytes and spermatogonia (precursors of spermatocytes) (Fig. 2A). These immunohistochemical results were in agreement with *in situ* hybridization data about *Ate1* expression in spermatocytes (77). To produce mouse strains in which *Ate1* was

selectively ablated in primordial germ cells (PGCs), we mated the previously constructed *Ate1*$^{flox/flox}$ mice (33) with *Tnap-Cre* mice expressing Cre recombinase from the PGC-specific *Tnap* promoter (78) (see Experimental Procedures). Immunoblotting analyses of testis extracts from the resulting *Tnap-Ate1*$^{-/-}$ mice vs. *Ate1*$^{flox/flox}$ (wild-type) mice with anti-Ate1 antibody indicated a dramatic decrease of Ate1 in *Tnap-Ate1*$^{-/-}$ testes (Fig. 2B). Inasmuch as spermatogonia and spermatocytes (in which *Tnap-Cre* was selectively expressed) comprise a large fraction but not the entirety of testicular cells, these results (Fig. 2B) indicated that the Ate1 R-transferase was either completely or nearly completely absent from spermatogonia and spermatocytes of *Tnap-Ate1*$^{-/-}$ mice.

Fertility of *Tnap-Ate1*$^{-/-}$ and "wild-type" *Ate1*$^{flox/flox}$ males was assessed by mating three males of each strain with *Ate1*$^{flox/flox}$ females. For each male mouse, at least six plugged females were collected and the pregnancy rates were recorded. Only ~9% of plugged *Ate1*$^{flox/flox}$ females became pregnant after mating with *Tnap-Ate1*$^{-/-}$male mice, in comparison to the pregnancy rate of ~78% after mating with *Ate1*$^{flox/flox}$ males (Fig. 2D). In addition, the average number of pups born to *Ate1*$^{flox/flox}$ females that were mated with *Tnap-Ate1*$^{-/-}$males was only ~1.3, in contrast to ~7.3 pups that were born, on average, to Ate1$^{flox/flox}$ females mated to *Ate1*$^{flox/flox}$ males (Fig. 2E).

In agreement with low fertility of *Tnap-Ate1*$^{-/-}$males (Fig. 2D, E), hematoxylin/eosin (HE)-stained sections of their epididymides contained few apparently mature spermatozoa (~1×10^6 per epididymis), in contrast to a much larger number (~16×10^6 per epididymis) of mature spermatozoa in *Ate1*$^{flox/flox}$ males (Fig. 2G, I). In contrast to *Ate1*$^{flox/flox}$ males, the epididymides of *Tnap-Ate1*$^{-/-}$ males contained a number of round-shaped cells with large nuclei, possibly immature spermatids (Fig. 2F). In

agreement with these results, *Tnap-Ate1*$^{-/-}$ testes were considerably smaller and lighter than their *Ate1*$^{flox/flox}$ counterparts (Fig. 2C, H). Histological assays also showed that the average diameter of seminiferous tubules in *Tnap-Ate1*$^{-/-}$ testes was ~139 μm, in contrast to ~204 μm for *Ate1*$^{flox/flox}$ tubules, and that *Tnap-Ate1*$^{-/-}$ tubules contained a smaller average number of cells per section (~93 cells vs. ~177 cells, respectively) (Fig. 2K, L).

These results (Fig. 2) indicated that the Ate1 R-transferase was required for normal fertility levels in male mice. Given very low but still non-zero fertility of *Ate1*$^{-/-}$ males (Fig. 2D, E) as well as complete or nearly complete absence of the Ate1 R-transferase from germ cells in *Tnap-Ate1*$^{-/-}$ testes (Fig. 2B), it is formally possible that the total absence of the Ate1-mediated arginylation in spermatogonia and spermatocytes is still compatible with a low but non-zero probability of sperm maturation. The alternative and a priori more likely interpretation is that the low but detectable sperm maturation that underlies the residual fertility of *Tnap-Ate1*$^{-/-}$ mice (Fig. 2D, E) is made possible by rare spermatocytes of *Tnap-Ate1*$^{-/-}$ testes that retained at least one copy of the intact *Ate1*^{flox} gene.

The Absence of Arginylation Is Compatible with Early Stages of Germ Cell Development – Using an antibody to the promyelocytic leukemia zinc finger (Plzf), a spermatogonia-specific marker, we observed similar levels of Plzf in presumptive spermatogonia in either the absence or presence of Ate1 (~56 Plzf-positive cells, on average, per a section of seminiferous tubule in both *Tnap-Ate1*$^{-/-}$ and *Ate1*$^{flox/flox}$ mice) (Fig. 3A, C). Formation of the synaptonemal complex and synapsis of chromosomes during the prophase of meiosis I in spermatocytes is accompanied by expression of the synaptonemal complex proteins 1 and 3 (Sycp1 and Sycp3) (83-85). Antibodies to Sycp1

and Sycp3 stained meiotic chromosomes indistinguishably in chromosome spreads of either *Tnap-Ate1*$^{-/-}$ or *Ate1*$^{flox/flox}$ spermatocytes (Fig. 3B). Similar immunofluorescence assays with Mlh1, a marker for chromosome crossovers (86-89), also showed no significant differences between *Tnap-Ate1*$^{-/-}$ and *Ate1*$^{flox/flox}$ spermatocytes (Fig. 3B, D, E).

In addition, we could readily identify, cytologically, every major stage of the prophase of meiosis I in both *Tnap-Ate1*$^{-/-}$ and *Ate1*$^{flox/flox}$ spermatocytes, including leptotene, zygotene, pachytene and diplotene, and there were no statistical significant differences between the two genotypes vis-á-vis the percentages of each stage of prophase I (Figs. 3E, F and Fig. 4A, B). The average numbers of spermatocytes per a seminiferous tubule section in the prophase of meiosis I were also similar between *Tnap-Ate1*$^{-/-}$ and *Ate1*$^{flox/flox}$ mice (~61 and ~60 spermatocytes, respectively) (Fig. 3E). These results indicated that (at resolution levels of our assays) pre-metaphase stages of germ-cell development did not require the Ate1 R-transferase.

Lack of Arginylation Causes Metaphase Arrest in Meiosis I – In contrast to the absence of detectable defects in the progression of Ate1-lacking *Tnap-Ate1*$^{-/-}$ spermatocytes through the prophase of meiosis I, we found these cells to be arrested at the metaphase of meiosis I, followed by their death through apoptosis (Figs. 3G, H and 4C, D, F). In stage-XII seminiferous tubules, metaphase-I spermatocytes could be identified by their highly condensed (hematoxylin-stained) chromatin (74) (Fig. 3G, H). Both metaphase and later-stage (anaphase) spermatocytes (indicated by arrowheads and arrows, respectively) could be observed in Ate1-containing tubules of *Ate1*$^{flox/flox}$ testes (Fig. 3G). However, no anaphase spermatocytes could be detected in tubules of

Ate1-lacking *Tnap-Ate1*$^{-/-}$ testes, whose relative content of metaphase spermatocytes was ~9.9%, in contrast to ~2.6% of such cells in *Ate1*$^{flox/flox}$ testes (Fig. 3G, H). In agreement with these results, it was easy to detect spindle bodies marked by ±-tubulin (consistent with the arrest of metaphase I spermatocytes) in *Tnap-Ate1*$^{-/-}$ testes (Fig. 4E). We also examined, by immunoblotting, the expression of cyclin B1, a marker of metaphase (90). The level of cyclin B1 was significantly increased in *Tnap-Ate1*$^{-/-}$ testes (Fig. 2J), yet another indication of metaphase arrest of Ate1-lacking spermatocytes in meiosis I.

Apoptotic Death of Arginylation-lacking Spermatocytes in Metaphase of Meiosis I – The terminal deoxynucleotidyl transferase dUTP nick end labeling assay (TUNEL) was used to measure the extent of apoptosis of *Tnap-Ate1*$^{-/-}$ vs. *Ate1*$^{flox/flox}$ spermatocytes (Fig. 4C, D, F). In *Tnap-Ate1*$^{-/-}$ testes, on average ~33% of spermatocytes were overtly apoptotic (TUNEL-positive), in comparison to ~9% of such cells in Ate1-containing *Ate1*$^{flox/flox}$ testes (Fig. 4C). In addition, only ~1.8 apoptotic cells per a seminiferous tubule section were found, on average, in *Ate1*$^{flox/flox}$ testes, vs. ~8.3 apoptotic cells in *Tnap-Ate1*$^{-/-}$ testes (Fig. 4F). Together, these results are likely to account for the observed massive decrease in the content of mature sperm cells in the testes of Ate1-lacking *Tnap-Ate1*$^{-/-}$ males and the resulting very low fertility of these mice (Fig. 2C-E). As described below, the metabolic stabilization of a natural fragment of the meiosis-specific Rec8 cohesin subunit is likely to be at least a significant, and possibly the major, reason for this functional consequence of *Ate1* ablation in germ cells.

Arginylation-mediated Degradation of the Separase-produced Rec8 Fragment – Rec8 is the main meiosis-specific cohesin subunit of the kleisin family. Mouse Rec8 is sequelogous to both yeast and mammalian kleisin-type cohesin subunits (Fig. 1C and

Introduction). The cleavage of mouse Rec8 by the separase would be expected to generate a 15-kDa C-terminal fragment bearing N-terminal Glu, a secondary destabilizing residue and a substrate of the Ate1 R-transferase (Fig. 1A, C). Although separase can cleave mouse Rec8 in vitro at more than one location, the cleavage between Arg-454 and Glu-455 is by far the predominant one (Fig. 1C) (76).

To determine whether the separase-generated Rec8 fragment was a substrate of the Arg/N-end rule pathway, we used both steady-state and pulse-chase assays. Immunoblotting of extracts from wild-type and *Tnap-Ate1*$^{-/-}$ mouse testes with antibody to a C-terminal region of Rec8 showed the presence of both the full-length endogenous Rec8 protein and its fragment. The latter species migrated, upon SDS-PAGE, at a position close to the one expected for the 15 kDa Glu^{455}-Rec8 fragment (Figs. 1C and 5A, lane 2). Strikingly, however, while this endogenous Rec8 fragment was abundant in extracts from *Tnap-Ate1*$^{-/-}$ testes, it was virtually absent in wild-type extracts (Fig. 5A, lane 1 vs. lane 2).

A parsimonious interpretation of these results is that the separase-generated 15 kDa Glu^{455}-Rec8 fragment (Fig. 1C) (76) was arginylated by the Ate1 R-transferase in wild-type cells and thereafter rapidly destroyed by the “downstream” part of the Arg/N-end rule pathway, whereas in *Tnap-Ate1*$^{-/-}$ cells the Glu^{455}-Rec8 fragment was long-lived, as it could not be arginylated (Figs. 1A and 5A). Hence the virtual absence of the endogenous Glu^{455}-Rec8 fragment in wild-type testes at steady state (Fig. 5A, lane 1) and its accumulation in *Tnap-Ate1*$^{-/-}$ testes (Fig. 5A, lane 2). Interestingly, the level of the full-length Rec8 protein was also considerably higher in *Tnap-Ate1*$^{-/-}$ testes than in wild-type ones, at equal total protein loads (Fig. 5A and Discussion).

Degradation of the Glu^{455}-Rec8 fragment was also assayed directly, using ^{35}S-pulse-chases and the Ub reference technique (URT), derived from the Ub fusion technique (Fig. 5C) (4,21,22,91,92). Cotranslational cleavage of a URT-based Ub fusion by deubiquitylases that are present in all eukaryotic cells produces, at the initially equimolar ratio, both a test protein with a desired N-terminal residue and the reference protein $^{f}DHFR\text{-}Ub^{R48}$, a flag-tagged derivative of the mouse dihydrofolate reductase (Fig. 5C). In URT-based pulse-chase assays, the labeled test protein is quantified by measuring its levels relative to the levels of a stable reference at the same time point during a chase. In addition to being more accurate than pulse-chases without a built-in reference, URT also makes it possible to detect and measure the degradation of the test protein before the chase, i.e., during the pulse (17,91,92).

URT-based ^{35}S-pulse-chases with C-terminally flag-tagged Glu^{455}-$Rec8^{f}$ and its derivatives were performed in a transcription-translation-enabled rabbit reticulocyte extract, which contains the Arg/N-end rule pathway and has been extensively used to analyze this pathway (4,21,22). The indicated $^{f}DHFR\text{-}Ub^{R48}$-$\mathbf{X}^{455}$-$Rec8^{f}$ URT fusions (**X**=Glu; Val; ArgGlu) were labeled with ^{35}S-Met/Cys in reticulocyte extract for 10 min at 30°C, followed by a chase, immunoprecipitation with a monoclonal anti-flag antibody, SDS-PAGE, autoradiography, and quantification (Fig. 5B-D). The logic of these assays involves a comparison between the degradation rates of a protein bearing a destabilizing N-terminal residue and an otherwise identical protein with an N-terminal residue such as Val, which is not recognized by the Arg/N-end rule pathway (Fig. 1A).

The Glu^{455}-$Rec8^{f}$ fragment was short-lived in reticulocyte extract ($t_{1/2}$ of 10-15 min) (Fig. 5B, D). Moreover, ~30% of pulse-labeled Glu^{455}-$Rec8^{f}$ were degraded during

the pulse, i.e., before the chase, in comparison to the otherwise identical Val^{455}-$Rec8^{f}$, which was also completely stable during the chase (Fig. 5B, D). We also constructed and examined Arg-Glu^{455}-$Rec8^{f}$. This protein was a DNA-encoded equivalent of the posttranslationally Nt-arginylated Glu^{455}-$Rec8^{f}$ fragment of Rec8. As would be expected, given the immediate (cotranslational) availability of N-terminal Arg in the DNA-encoded Arg-Glu^{455}-$Rec8^{f}$, this protein was destroyed by the Arg/N-end rule pathway even more rapidly than the already short-lived Glu^{455}-$Rec8^{f}$. Specifically, nearly 80% of Arg-Glu^{455}-$Rec8^{f}$ was eliminated during the 10-min pulse (before the chase), in comparison to ~30% of Glu^{455}-Rec8, with the long-lived Val^{455}-$Rec8^{f}$ control being a part of the reference set (Fig. 5B, D).

The results of pulse-chase analyses (Fig. 5B-D) were in agreement with other measurements, in which the synthesis-deubiquitylation-degradation of ^{f}DHFR-Ub^{R48}-**X**-$Rec8^{f}$ fusions in reticulocyte extract was allowed to proceed for 1 hr, followed by detection of Glu^{455}-$Rec8^{f}$, of other test proteins, and of the ^{f}DHFR-Ub^{R48} reference protein by SDS-PAGE and immunoblotting with anti-flag antibody (Fig. 5E). In these assays, the samples were incubated in reticulocyte extract either without added dipeptides, or with 1 mM Arg-Ala (RA), bearing N-terminal Arg, a type-1 primary destabilizing residue, or with 1 mM Ala-Arg (AR), bearing N-terminal Ala, a residue that is not recognized by N-recognins of the Arg/N-end rule pathway (Figs. 1A and 5E).

Ubr1 and Ubr2, the two sequelogous (47% identical) and functionally overlapping 200-kDa N-recognins (E3 Ub ligases) of the mammalian Arg/N-end rule pathway, have several substrate-binding sites. These sites recognize (bind to) specific classes of N-degrons and specific internal (non-N-terminal) degrons. The two substrate-binding

sites that recognize N-degrons are the type-1 site, which specifically binds to the N-terminal basic residues Arg, Lys or His, and the adjacent but distinct type-2 site, which specifically binds to the N-terminal bulky hydrophobic residues Leu, Phe, Tyr, Trp, and Ile (4,6,93,94). Dipeptides bearing, for example, type-1 N-terminal residues can competitively and selectively inhibit the binding of Ubr1/Ubr2 to a type-1 N-degron but not to a type-2 N-degron in a test protein (4,6,95-97).

In agreement with the rapid degradation of the Glu^{455}-$Rec8^{f}$ fragment in ^{35}S-pulse-chase assays (Fig. 5B-D), the synthesis-deubiquitylation-degradation of the $^{f}DHFR\text{-}Ub^{R48}\text{-}Glu^{455}\text{-}Rec8^{f}$ fusion for 1 hr in reticulocyte extract either in the absence of added dipeptide, or in the presence of Ala-Arg (which does not bind to Ubr1/Ubr2) did not result in a detectable accumulation of the Glu^{455}-$Rec8^{f}$ fragment, at its expected (roughly 15 kDa) position in SDS-PAGE-based immunoblots (Fig. 5E, lanes 8 and 10). In striking contrast, the same assay but in the presence of the Arg-Ala dipeptide (which competitively inhibits the recognition of type-1 N-degrons) resulted in a prominent protein band at the expected position of Glu^{455}-$Rec8^{f}$ (Fig. 5E, lane 9 vs. lanes 8 and 10). Given the URT-based design of $^{f}DHFR\text{-}Ub^{R48}\text{-}Glu^{455}\text{-}Rec8^{f}$ (Fig. 5C), that protein band was inferred to be the Glu^{455}-$Rec8^{f}$ fragment that had been metabolically stabilized by the Arg-Ala dipeptide. In contrast, the larger Met^{206}-$Rec8^{f}$ fragment, corresponding to a putative (and at most a minor) separase cleavage site in the full-length Rec8 protein, was metabolically stable irrespective of the presence or absence of the Arg-Ala dipeptide (Fig. 5E, lanes 5-7 vs. lanes 8-10). The same was true of full-length $Rec8^{f}$. It should also be mentioned that full-length $Rec8^{f}$ was not converted into a smaller fragment in

reticulocyte extract, indicating (as would be expected) the absence of active separase in that extract (Fig. 5E, lanes 2-4).

DISCUSSION

The arginyltransferase (R-transferase) Ate1 is a component of the Arg/N-end rule pathway of protein degradation. Ate1 utilizes Arg-tRNA as a cosubstrate to arginylate N-terminal Asp, Glu or (oxidized) Cys of a targeted protein substrate. The resulting N-terminal Arg is recognized by E3 ubiquitin ligases (N-recognins) of the Arg/N-end rule pathway (Fig. 1A, Introduction, and references therein). In the present study, we constructed an *Ate1*$^{-/-}$ mouse strain in which the ablation of *Ate1* was confined to germ cells. We used this and other experimental tools to characterize Glu^{455}-Rec8, a specific C-terminal fragment of the kleisin-type, meiosis-specific Rec8 subunit of mouse cohesin. Rec8 is cleaved, late in meiosis I (see Introduction and references therein), by a nonprocessive protease called separase. The main C-terminal fragment of that cleavage, Glu^{455}-Rec8, bears N-terminal Glu, a substrate of Ate1 (Fig. 1A, C). We have shown that the mouse Glu^{455}-Rec8 fragment is a short-lived substrate of the Arg/N-end rule pathway and that the degradation of this fragment requires its Nt-arginylation by the Ate1 R-transferase (Figs. 5).

In *S. cerevisiae*, Scc1/Rad21/Mcd1 is the mitotic counterpart of the mammalian meiotic Rec8 subunit of cohesin. Similarly to the mouse Glu^{455}-Rec8 fragment, the separase-generated C-terminal fragment of yeast Scc1 is also a short-lived substrate of the Arg/N-end rule pathway, and the failure to destroy this fragment in *ubr1"* cells (which lack the Arg/N-end rule pathway) results in chromosome instability (62). The

separase-generated C-terminal fragment of yeast Scc1 retains, in part, the physical affinity of Scc1 for the rest of cohesin complex (62). The mouse Glu^{455}-Rec8 fragment would also be likely to interact in vivo with the rest of meiotic cohesin. If so, the failure to arginylate the Glu^{455}-Rec8 fragment in arginylation-lacking spermatocytes of *Ate1*$^{-/-}$ mice, and hence the failure to destroy this fragment (Figs. 1A and 5) would be expected to interfere with cohesin mechanics. This (at present hypothetical) interference would account, at least in part, for the observed arrest and apoptotic death of *Ate1*$^{-/-}$ spermatocytes at the end of meiosis I and the resulting strong decrease in the fertility of *Ate1*$^{-/-}$ males (Fig. 2C-L).

Immunoblotting analyses of the arginylation-dependent degradation of the endogenous Glu^{455}-Rec8 fragment showed that this fragment was virtually absent, at steady state, in wild-type mouse testes but accumulated in *Ate1*$^{-/-}$ testes (Fig. 5A). Interestingly, the steady-state level of the full-length Rec8 protein was also significantly higher in *Ate1*$^{-/-}$ testes than in wild-type ones, at equal total protein loads (Fig. 5A). A plausible but unproven interpretation of this result is that the cleavage of the full-length mouse Rec8 by separase may be subject to a product-mediated inhibition of separase in *Ate1*$^{-/-}$ spermatocytes, if the product, Glu^{455}-Rec8, is no longer eliminated by the Ate1-dependent arginylation branch of the Arg/N-end rule pathway (Figs. 1A and 5).

Although selective ablation of *Ate1* in mouse germ cells nearly abrogates the fertility of male mice (Fig. 2C-E), it is unlikely that specific cases of human infertility can be caused by unconditionally null mutations of human *Ate1*, inasmuch as global mouse *Ate1*$^{-/-}$ mutants are late embryonic lethals (29). In addition, a post-natal ablation of mouse *Ate1*, in adult mice (using *cre-lox* technology and a ubiquitously expressed Cre

recombinase), while compatible with mouse viability, causes a variety of abnormal phenotypes, including the loss of fat and hyperkinetic behavior (33). Nevertheless, partially active (hypomorphic) mutants of the human Ate1 R-transferase might underlie some, currently obscure, cases of human infertility.

The same disposition would obtain if the human Arg/N-end rule pathway would be partially inactivated "downstream" of the Ate1 R-transferase, at the level of pathway's Ub ligases (Fig. 1A). For example, unconditional *Ubr1*$^{-/-}$ mice, lacking one of two major N-recognins, Ubr1 and Ubr2, are viable and fertile, while exhibiting some abnormal phenotypes (98). The analogous *Ubr2*$^{-/-}$ mice, lacking the second major N-recognin (it is structurally and functionally similar to Ubr1), are also viable (in some strain backgrounds) but exhibit male infertility (77). This infertility is similar to the low-fertility phenotype of *Ate1*$^{-/-}$ mice in the present study, as in both cases the infertility is caused by apoptotic death, in meiosis I, of either *Ubr2*$^{-/-}$ or *Ate1*$^{-/-}$ spermatocytes (Figs. 2C-D and 4C, D, F). A parsimonious interpretation of these results is that a failure to rapidly destroy the separase-generated Glu^{455}-Rec8 fragment is a common mechanistic denominator of both *Ate1*$^{-/-}$ and *Ubr2*$^{-/-}$ infertility phenotypes.

No human *Ubr2*$^{-/-}$ mutants have been identified so far. In contrast, human patients with the previously characterized Johansen-Blizzard Syndrome (JBS) have been shown to be null *Ubr1*$^{-/-}$ mutants (99). It is unknown whether or not human JBS patients are fertile, in part because the overall phenotype of human JBS is more severe than analogous phenotype of *Ubr1*$^{-/-}$ mice. Abnormal phenotypes of human JBS (*Ubr1*$^{-/-}$) patients include anatomical malformations, an insufficiency and inflammation of the acinar pancreas, mental retardation, and deafness (4,99).

Although separase can cleave mouse Rec8 in vitro at more than one site, the cleavage between Arg-454 and Glu-455, resulting in the Glu^{455}-Rec8 fragment, is by far the predominant one (Fig. 1C) (76). Mammalian Rad21, the mitosis-specific counterpart of the meiotic Rec8 cohesin subunit, is also cleaved by separase, late in mitosis. Similarly to the cleavage of the meiosis-specific Rec8, the separase-mediated cleavage of the mammalian mitotic Rad21 subunit also yields the C-terminal fragment of Rad21 bearing N-terminal Glu (75). However, in contrast to the present results with the Glu^{455}-Rec8 fragment of meiotic Rec8 (Fig. 4B, C), our recent analyses of the N-terminal Glu-bearing mitotic Rad21 fragment using ^{35}S-pulse-chases and the URT method (Fig. 5A) indicated that this fragment was long-lived (B.W. and A.V., unpublished data). Further analyses of these unexpected (and therefore particularly interesting) results about the apparently stable mitotic Glu-Rad21 fragment vis-à-vis the short-lived meiotic Glu^{455}-Rec8 fragment (Fig. 4) are under way.

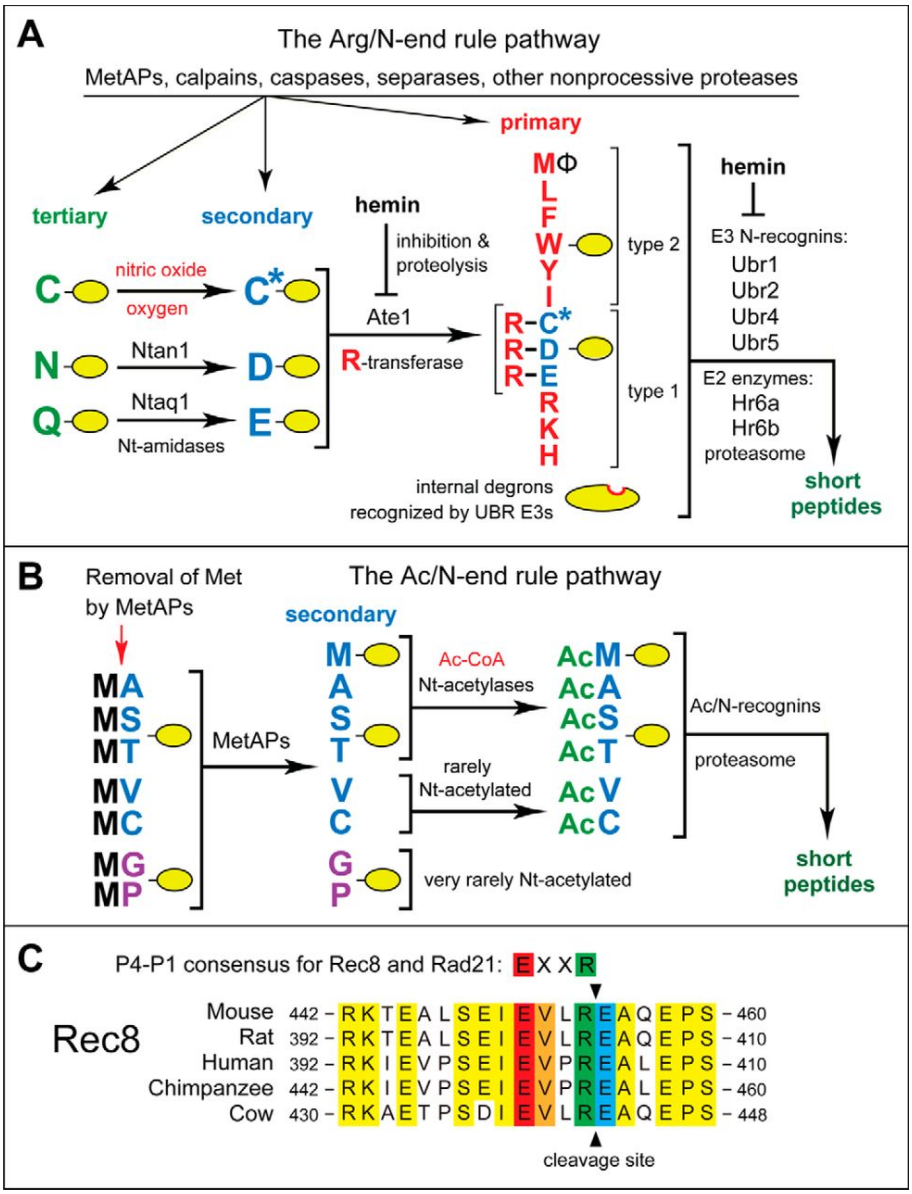


Fig. 5.1. **The mammalian N-end rule pathway and the separase cleavage site in Rec8, a meiosis-specific cohesin subunit.** See Introduction for references and descriptions of the pathway's mechanistic aspects and biological functions. Amino acid residues are denoted by single-letter abbreviations. (A) The Arg/N-end rule pathway. It targets proteins for degradation through their specific unacetylated N-terminal residues. A yellow oval denotes the rest of a protein substrate. "R-transferase" is the Ate1 arginyltransferase. "Primary", "secondary", and "tertiary" refer to mechanistically distinct classes of destabilizing N-terminal residues. "Type 1" and "type 2" refer to two sets of primary destabilizing N-terminal residues, basic (Arg, Lys, His) and bulky hydrophobic (Leu, Phe, Trp, Tyr, Ile, and Met followed by a bulky hydrophobic residue ($)), respectively. These sets of N-terminal residues are recognized by two distinct substrate-binding sites of N-recognins, the pathway's E3 ubiquitin ligases. (B) The Ac/N-end rule pathway. It targets proteins through their $N^{\pm}$-terminally acetylated

(Nt-acetylated) residues. Red arrow on the left indicates the cotranslational removal of the N-terminal Met residue by Met-aminopeptidases (MetAPs). N-terminal Met is retained if a residue at position 2 is larger than Val. (C) Alignments of amino acid sequences near the main separase cleavage site in mammalian Rec8, between Arg^{454} and Glu^{455} of mouse Rec8. The consensus sequence of this cleavage site (its P4-P1 residues) in both mitotic (Rad21) and meiotic (Rec8) kleisin-type cohesin subunits is also shown. Conserved residues of mammalian Rec8 near the cleavage site are in yellow. The conserved P4, P3, P1 and P1' residues of mammalian Rec8 at the cleavage site are in red, orange, green, and blue, respectively.

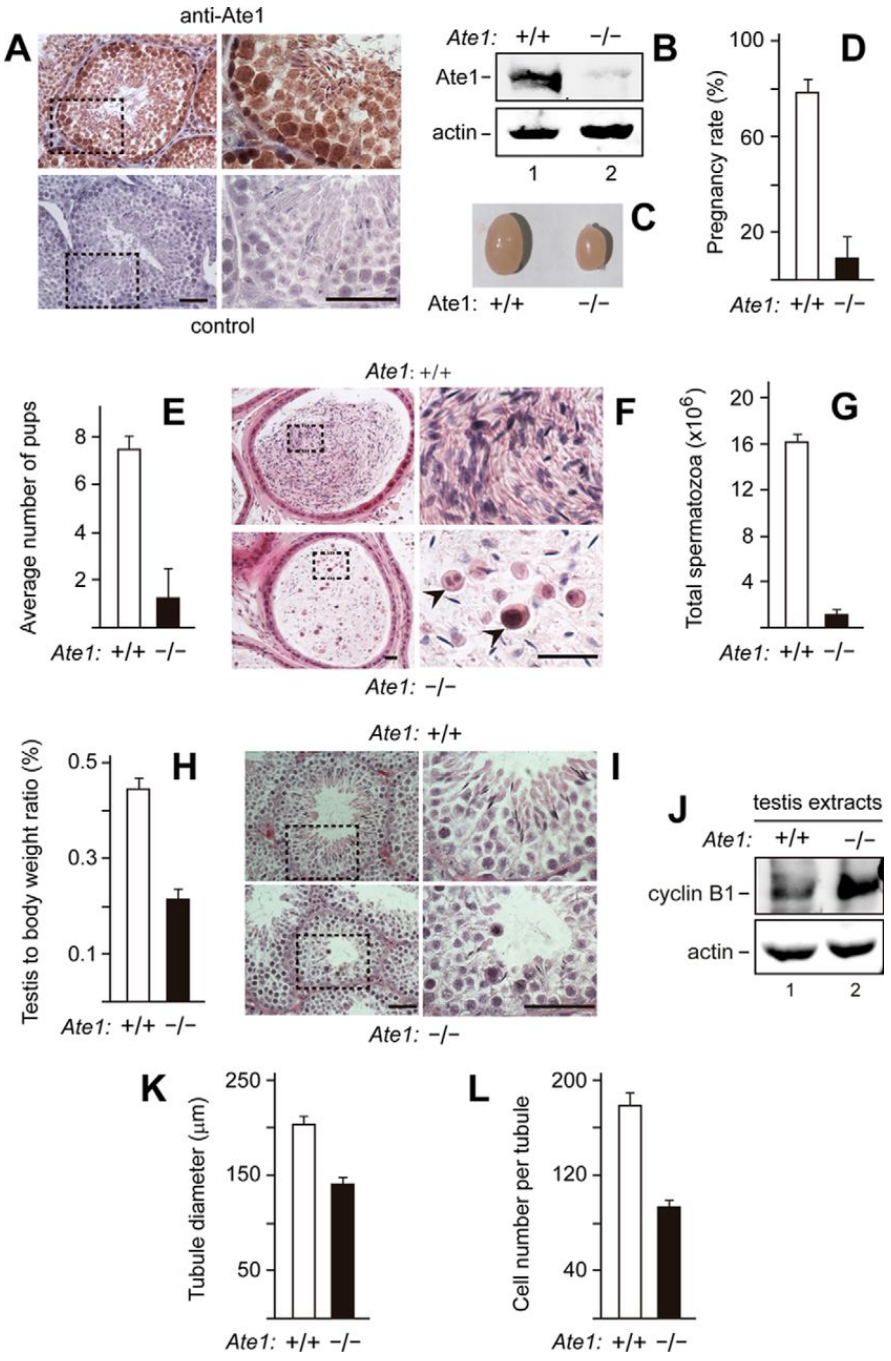


Fig. 5.2. ***Tnap-Ate1*$^{-/-}$ mice, in which the ablation of *Ate1* is confined to germ cells, are nearly infertile.** (A) Immunohistochemical detection of the Ate1 R-transferase in sections of wild-type (*Ate1*$^{flox/flox}$) seminiferous tubules in adult mice, using affinity-purified anti-Ate1 antibody. Sub-panels on the right are enlargements of the areas demarcated by dashed rectangles on the left. Upper sub-panels: staining with primary (anti-Ate1) and secondary antibodies. Lower sub-panels: secondary antibody alone (control). Scale bars, 50 µm. (B) Dramatic decrease of the Ate1 R-transferase in the testes

of *Tnap-Ate1*$^{-/-}$ mice, as determined by immunoblotting of testes extracts with anti-Ate1 antibody. (C) Decreased size of *Tnap-Ate1*$^{-/-}$ testes, in comparison to wild-type (*Ate1*$^{flox/flox}$) ones. (D) Pregnancy rates (%) of plugged wild-type females after matings with *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (E) Average numbers of pups per all plugged wild-type females after matings with *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (F) Histological appearance (hematoxylin-eosin) of sections through epididymis of *Tnap-Ate1*$^{-/-}$ vs. wild-type males. Sub-panels on the right are enlargements of areas (indicated by dashed rectangles) on the left. Arrowheads indicate abnormal cells, absent in sections of wild-type epididymis. Scale bar, 20 μm. (G) Calculated total sperm number in epididymis of *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (H) Ratio of testis mass to mouse body mass for *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (I) Histological appearance (hematoxylin-eosin) of sectioned seminiferous tubules in *Tnap-Ate1*$^{-/-}$ vs. wild-type males. Scale bars, 50 μm. (J) Immunoblotting analyses, using antibodies to actin and to cyclin B1 (the latter a marker for metaphase) of extracts from *Tnap-Ate1*$^{-/-}$ vs. wild-type testes (see the main text). (K) Average diameters of seminiferous tubules in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. (L) Average numbers of cells per seminiferous tubule in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. Standard deviations are indicated in D, E, G, H, K and L (the corresponding assays were carried out in triplicate).

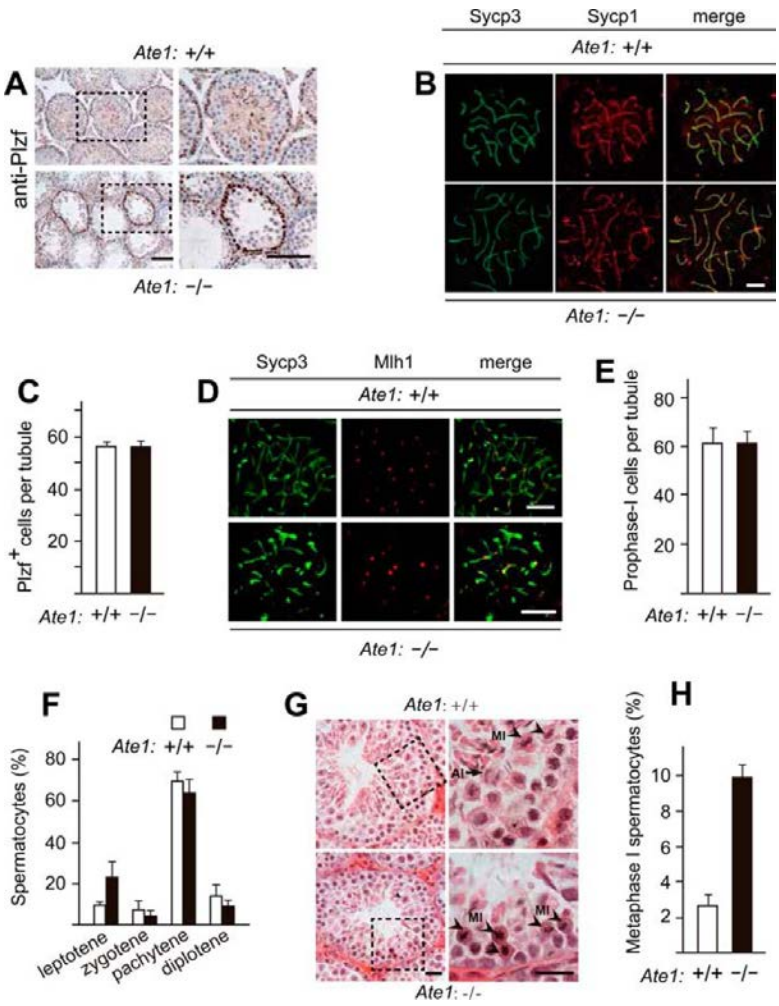


Fig. 5.3. **Knockout of the *Ate1* R-transferase gene does not affect the progression of spermatocytes through meiosis I until their arrest at metaphase.** (A) Similar frequencies of cells ($Plzf^+$ cells) that could be stained with antibody to Plzf, a spermatogonia-specific marker, in testes of *Tnap-Ate1*$^{-/-}$ vs. wild-type males. Scale bars: 50 μm. (B) Sycp3 (green) and Sycp1 (red) proteins, the markers for synaptonemal complex, detected (using corresponding antibodies) on chromosomes of *Tnap-Ate1*$^{-/-}$ vs. wild-type spermatocytes at the pachytene stage of meiosis I. Scale bars, 5 μm. (C) Numbers of $Plzf^+$ cells (presumptive spermatogonia) per section of a seminiferous tubule in *Tnap-Ate1*$^{-/-}$ vs. wild-type males (quantification of results in A). (D) Same as in B but for Sycp3 and Mlh1 (the latter a marker for chromosome crossovers) in *Tnap-Ate1*$^{-/-}$ vs. wild-type males. Scale bars: 5 μm. (E) Numbers of prophase I (Sycp3-positive) cells per section of a seminiferous tubule in *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (F) Percentages of spermatocytes at different prophase stages of meiosis I *Tnap-Ate1*$^{-/-}$ vs. wild-type males. (G) Sections of *Tnap-Ate1*$^{-/-}$ vs. wild-type testes were stained with hematoxylin/eosin. Sub-panels on the right are enlargements of the areas demarcated by dashed rectangles on the left. An arrow and arrowheads indicate anaphase and metaphase cells, respectively. Scale bars, 50 μm. (H) Percentages of spermatocytes at metaphase I of meiosis I (quantification of cell images in G). Standard deviations are indicated in C, E, F, and H (the corresponding assays were carried out in triplicate).

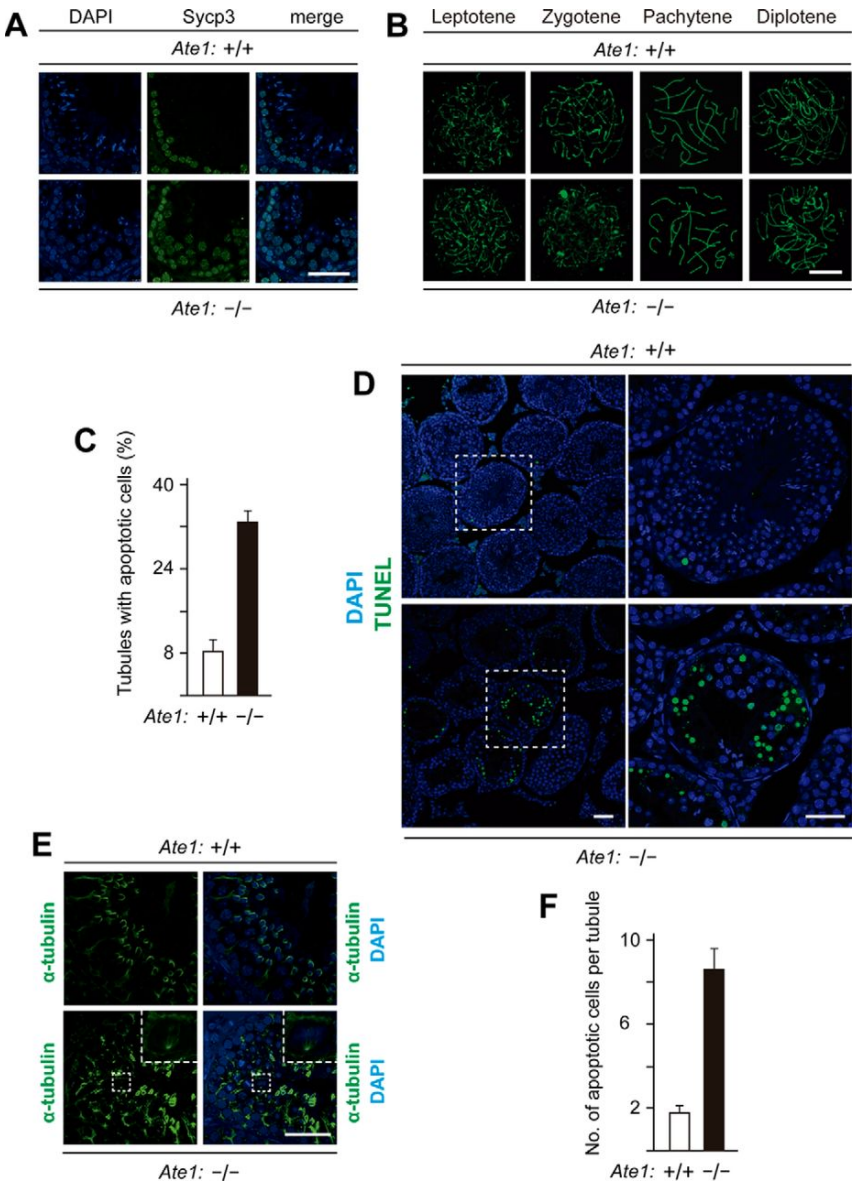


Fig. 5.4. **Increased apoptosis in seminiferous tubules of *Tnap-Ate1*$^{-/-}$ mice.**
(A) Spermatocytes in prophase I of meiosis I, stained with DAPI for DNA and with anti-Sycp1 antibody in sections of seminiferous tubes in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. Scale bar, 50 μm. (B) Stages of meiosis I in spermatocytes of *Tnap-Ate1*$^{-/-}$ vs. wild-type testes, detected by staining chromosome spreads with anti-Sycp3 antibody. Scale bar, 10 μm. (C) Percentages of seminiferous tubules containing apoptotic (TUNEL-positive) cells in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. (D) Representative TUNEL-assay patterns (stained with DAPI as well) in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. Sub-panels on the right are enlargements of areas (indicated by dashed squares) on the left. Scale bars, 50 μm. (E) Staining of testes sections from *Tnap-Ate1*$^{-/-}$ vs. wild-type testes with DAPI for DNA, and with antibody to ±-tubulin for spindle bodies (see the main text). Scale bar, 50 μm. (F) Numbers of apoptotic (TUNEL-positive) cells per a section of seminiferous tubules in *Tnap-Ate1*$^{-/-}$ vs. wild-type testes. Standard deviations are indicated in C and F (the corresponding assays were carried out in triplicate).

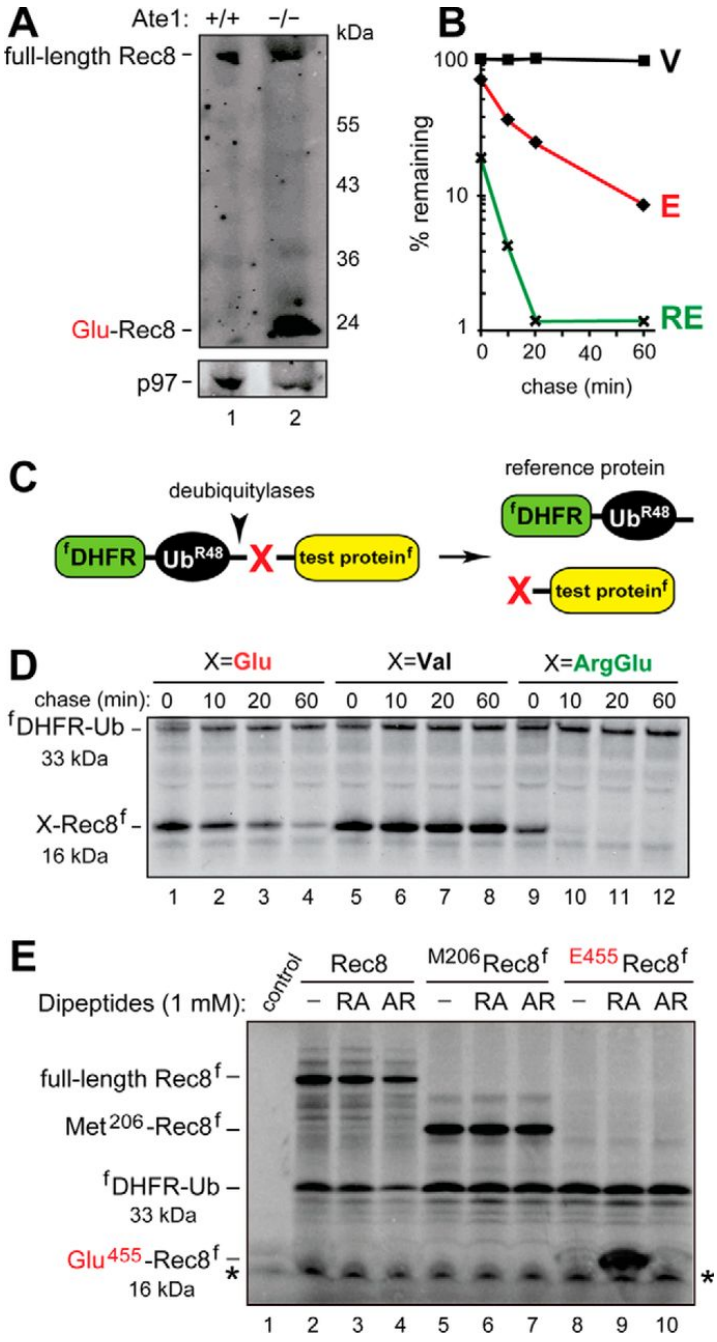


Fig. 5.5. **The separase-generated C-terminal fragment of Rec8 as a short-lived substrate of the Arg/N-end rule pathway.** (A) Immunoblotting of extracts from wild-type (lane 1) vs. *Tnap-Ate1*$^{-/-}$ (lane 2) mouse testes with antibody to a C-terminal region of Rec8. Note the presence of a prominent band, inferred to be the endogenous Glu^{455}-Rec8 fragment, in *Tnap-Ate1*$^{-/-}$ but not wild-type testes. The lower panel shows the results of control immunoblots using antibody to the unrelated p97 protein. (B) Quantification of data in D. (C) The ubiquitin reference technique (URT; see the main text). (D) Lanes 1-4, ^{35}S-pulse-chase of the C-terminally flag-tagged Glu^{455}-Rec8^{f} fragment of mouse Rec8, produced as the URT fusion ^{f}DHFR-Ub^{R48}-Glu^{455}-Rec8^{f} in reticulocyte extract (see the

main text). The bands of Glu^{455}-$Rec8^{f}$ and the reference protein $^{f}DHFR\text{-}Ub^{R48}$ are indicated on the left. Lanes 5-8, same as in A but with the otherwise identical Val^{455}-$Rec8^{f}$, bearing N-terminal Val, which is not targeted by the Arg/N-end rule pathway. Lanes 9-12, same as in but with Arg-Glu^{455}-$Rec8^{f}$ (see the main text). (E) In these assays, the synthesis-deubiquitylation-degradation of URT-based $^{f}DHFR\text{-}Ub^{R48}$-**X**-$Rec8^{f}$ fusions in reticulocyte extract was allowed to proceed for 1 hr, followed by detection of Glu^{455}-$Rec8^{f}$, of other test proteins, and of the $^{f}DHFR\text{-}Ub^{R48}$ reference protein by SDS-PAGE and immunoblotting with anti-flag antibody. Lane 1, control extract with the added vector plasmid (see Experimental Procedures). Lane 2, expression of the $^{f}DHFR\text{-}Ub^{R48}$-$Rec8^{f}$ fusion, which is processed by deubiquitylases in the extract to yield full-length $Rec8^{f}$. Lane 3, same as lane 2, but in the presence of the Arg-Ala (RA) dipeptide at 1 mM. Lane 4, same as lane 3, but in the presence of the Ala-Arg (AR) dipeptide at 1 mM. Lanes 5-7, same as lanes 2-4, but expression of the $^{f}DHFR\text{-}Ub^{R48}$-Met^{206}-$Rec8^{f}$ fusion, which is processed in the extract to yield $^{f}DHFR\text{-}Ub^{R48}$ and Met^{206}-$Rec8^{f}$ (see the main text). Lanes 8-10, same as lanes 5-7, but expression of the $^{f}DHFR\text{-}Ub^{R48}$-Glu^{455}-$Rec8^{f}$ fusion, which is processed in the extract to yield $^{f}DHFR\text{-}Ub^{R48}$ and the Glu^{455}-$Rec8^{f}$ protein, the main product of the cleavage of full-length Rec8 by separase. Note the presence of a prominent protein band, inferred to be the 16 kDa flag-tagged Glu^{455}-$Rec8^{f}$ protein, in the presence of the Arg-Ala dipeptide (lane 9) but neither in the absence of added dipeptide or in the presence of Ala-Arg (lanes 8 and 10; see also the main text). Also indicated, in panels D and E, are the molecular masses of key protein species, the C-terminally flag-tagged Rec8 fragment (16 kDa) and the N-terminally flag-tagged reference protein DHFR-Ub (33 kDa). An asterisk denotes a protein band that crossreacted with anti-flag antibody.

REFERENCES

1. Bachmair, A., Finley, D., and Varshavsky, A. (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234, 179-186
2. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327, 973-977
3. Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A., and Hwang, C. S. (2014) The N-terminal methionine of cellular proteins as a degradation signal. Cell 156, 158-169
4. Varshavsky, A. (2011) The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20, 1298-1345
5. Gibbs, D. J., Bacardit, J., Bachmair, A., and Holdsworth, M. J. (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. Trends Cell Biol. 24, 603-611
6. Tasaki, T. S., Sriram, S. M., Park, K. S., and Kwon, Y. T. (2012) The N-end rule pathway. Annu. Rev. Biochem. 81, 261-289
7. Dougan, D. A., Micevski, D., and Truscott, K. N. (2011) The N-end rule pathway: from recognition by N-recognins to destruction by AAA+ proteases. Biochim. Biophys. Acta 1823, 83-91
8. Varshavsky, A. (2008) Discovery of cellular regulation by protein degradation. J. Biol. Chem. 283, 34469-34489
9. Bachmair, A., and Varshavsky, A. (1989) The degradation signal in a short-lived protein. Cell 56, 1019-1032

10. Inobe, T., and Matouschek, A. (2014) Paradigms of protein degradation by the proteasome. Curr. Op. Struct. Biol. 24, 156-164

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

12. Mogk, A., Schmidt, R., and Bukau, B. (2007) The N-end rule pathway of regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 17, 165-172

13. Rivera-Rivera, I., Román-Hernández, G., Sauer, R. T., and Baker, T. A. (2014) Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. Proc. Natl. Acad. Sci. USA 111, E3853-E3859.

14. Piatkov, K. I., Vu, T. T. M. H., C.-S., and Varshavsky, A. (2015) Formyl-methionine as a degradation signal at the N-termini of bacterial proteins. Microbial Cell 2, 376-393

15. Humbard, M. A., Surkov, S., De Donatis, G. M., Jenkins, L., and Maurizi, M. R. (2013) The N-degradome of Escherichia coli: limited proteolysis in vivo generates a large pool of proteins bearing N-degrons. J. Biol. Chem. 288, 28913-28924

16. Graciet, E., Hu, R. G., Piatkov, K., Rhee, J. H., Schwarz, E. M., and Varshavsky, A. (2006) Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc. Natl. Acad. Sci. USA 103, 3078-3083

17. Shemorry, A., Hwang, C. S., and Varshavsky, A. (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50, 540-551

18. Park, S. E., Kim, J. M., Seok, O. H., Cho, H., Wadas, B., Kim, S. Y., Varshavsky, A., and Hwang, C. S. (2015) Control of mammalian G protein signaling by N-terminal acetylation and the N-end rule pathway. Science 347, 1249-1252

19. Aksnes, H., Hole, K., and Arnesen, T. (2015) Molecular, cellular, and physiological significance of N-terminal acetylation. Int. Rev. Cell. Mol. Biol. 316, 267-305

20. Starheim, K. K., Gevaert, K., and Arnesen, T. (2012) Protein N-terminal acettyltransferases: when the start matters. Trends Biochem. Sci. 37, 152-161

21. Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2012) The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. Proc. Natl. Acad. Sci. USA 109, E1839-E1847

22. Piatkov, K. I., Oh, J.-H., Liu, Y., and Varshavsky, A. (2014) Calpain-generated natural protein fragments as short-lived substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 111, E817-E826

23. Brower, C. S., Piatkov, K. I., and Varshavsky, A. (2013) Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. Mol. Cell 50, 161-171

24. Yamano, K., and Youle, R. J. (2013) PINK1 is degraded through the N-end rule pathway. Autophagy 9, 1758-1769

25. Cha-Molstad, H., Sung, K. S., Hwang, J., Kim, K. A., Yu, J. E., Yoo, Y. D., Jang, J. M., Han, D. H., Molstad, M., Kim, J. G., Lee, Y. J., Zakrzewska, A., Kim, S. H., Kim, S. T., Kim, S. Y., Lee, H. G., Soung, N. K., Ahn, J. S., Ciechanover, A., Kim, B. Y., and Kwon, Y. T. (2015) Amino-terminal arginylation targets endoplasmic reticulum chaperone BiP for autophagy through p62 binding. Nat. Cell Biol. 17, 917-929

26. Wang, H., Piatkov, K. I., Brower, C. S., and Varshavsky, A. (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell 34, 686-695

27. Hwang, C. S., Shemorry, A., and Varshavsky, A. (2010) The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 12, 1177-1185

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

29. Kwon, Y. T., Kashina, A. S., Davydov, I. V., Hu, R.-G., An, J. Y., Seo, J. W., Du, F., and Varshavsky, A. (2002) An essential role of N-terminal arginylation in cardiovascular development. Science 297, 96-99

30. Hu, R.-G., Sheng, J., Xin, Q., Xu, Z., Takahashi, T. T., and Varshavsky, A. (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437, 981-986

31. Hu, R.-G., Brower, C. S., Wang, H., Davydov, I. V., Sheng, J., Zhou, J., Kwon, Y. T., and Varshavsky, A. (2006) Arginyl-transferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem. 281, 32559-32573

32. Saha, S., and Kashina, A. (2011) Posttranslational arginylation as a global biological regulator. Dev. Biol. 358, 1-8

33. Brower, C. S., and Varshavsky, A. (2009) Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE 4, e7757

34. Brower, C. S., Rosen, C. E., Jones, R. H., Wadas, B. C., Piatkov, K. I., and Varshavsky, A. (2014) Liat1, an arginyltransferase-binding protein whose evolution among primates involved changes in the numbers of its 10-residue repeats. Proc. Natl. Acad. Sci. USA 111, E4936-4945

35. Wang, J., Han, X., Saha, S., Xu, T., Rai, R., Zhang, F., Wolf, Y. I., Wolfson, A., Yates, J. R., and Kashina, A. (2011) Arginyltransferase is an ATP-independent self-regulating enzyme that forms distinct functional complexes in vivo. Chem. Biol. 18, 121-130

36. Hu, R.-G., Wang, H., Xia, Z., and Varshavsky, A. (2008) The N-end rule pathway is a sensor of heme. Proc. Natl. Acad. Sci. USA 105, 76-81

37. Varshavsky, A. (2004) Spalog and sequelog: neutral terms for spatial and sequence similarity. Curr. Biol. 14, R181-R183

38. Lee, M. J., Tasaki, T., Moroi, K., An, J. Y., Kimura, S., Davydov, I. V., and Kwon, Y. T. (2005) RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 102, 15030-15035

39. Varshavsky, A. (2012) Augmented generation of protein fragments during wakefulness as the molecular cause of sleep: a hypothesis. Prot. Sci. 21, 1634-1661

40. Piatkov, K. I., Colnaghi, L., Bekes, M., Varshavsky, A., and Huang, T. T. (2012) The auto-generated fragment of the Usp1 deubiquitylase is a physiological substrate of the N-end rule pathway. Mol. Cell 48, 926-933

41. Crawford, E. D., Seaman, J. E., Agard, N., Hsu, G. W., Julien, O., Mahrus, S., Nguyen, H., Shimbo, K., Yoshihara, H. A., Zhuang, M., Chalkley, R. J., and Wells, J. A. (2013) The DegraBase: a database of proteolysis in healthy and apoptotic human cells. Mol. Cell. Proteomics 12, 813-824

42. Singh, V. P., and Gerton, J. L. (2015) Cohesin and human disease: lessons from mouse models. Curr. Op. Cell Biol. 37, 9-17

43. Rankin, S. (2015) Complex elaboration: making sense of meiotic cohesin dynamics. FEBS J. 282, 2426-2443

44. Hirano, T. (2015) Chromosome Dynamics during Mitosis. Cold Spring Harb. Perspect. Biol. 7

45. Haarhuis, J. H., Elbatsh, A. M., and Rowland, B. D. (2014) Cohesin and its regulation: on the logic of X-shaped chromosomes. Dev. Cell 31, 7-18

46. Bloom, K. S. (2014) Centromeric heterochromatin: the primordial segregation machine. Annu. Rev. Genet. 48, 457-484

47. Peters, J. M., and Nishiyama, T. (2012) Sister chromatid cohesion. Cold Spring Harb. Perspect. Biol. 4

48. Nasmyth, K., and Haering, C. H. (2009) Cohesin: its roles and mechanisms. Annu. Rev. Genet. 43, 525-558

49. Camdere, G., Guacci, V., Stricklin, J., and Koshland, D. (2015) The ATPases of cohesin interface with regulators to modulate cohesin-mediated DNA tethering. eLife 4, ??

50. Ocampo-Hafalla, M. T., Katou, Y., Shirahige, K., and Uhlmann, F. (2007) Displacement and re-accumulation of centromeric cohesin during transient pre-anaphase centromere splitting. Chromosoma 116, 531-544

51. Winters, T., McNicoll, F., and Jessberger, R. (2014) Meiotic cohesin STAG3 is required for chromosome axis formation and sister chromatid cohesion. EMBO J. 33, 1256-1270

52. Rudra, S., and Skibbens, R. V. (2013) Cohesin codes - interpreting chromatin architecture and the many facets of cohesin function. J. Cell Sci. 126, 31-41

53. Murayama, Y., and Uhlmann, F. (2015) DNA entry into and exit out of the cohesin ring by an interlocking gate mechanism. Cell 163, 1628-1640

54. Heidinger-Pauli, J. M., Unal, E., Guacci, V., and Koshland, D. (2008) The kleisin subunit of cohesin dictates damage-induced cohesion. Mol. Cell 31, 47-56

55. Dorsett, D., and Strom, L. (2012) The ancient and evolving roles of cohesin in gene expression and DNA repair. Curr. Biol. 22, R240-250

56. Chen, F., Kamradt, M., Mulcahy, M., Byun, Y., Xu, H., McKay, M. J., and Cryns, V. L. (2002) Caspase proteolysis of the cohesin component RAD21 promotes apoptosis. J. Biol. Chem. 277, 16775-16781

57. Pati, D., Zhang, N., and Plon, S. E. (2002) Linking sister chromatid cohesion and apoptosis: role of Rad21. Mol. Cell. Biol. 22, 8267-8277

58. Panigrahi, A. K., Zhang, N., Mao, Q., and Pati, D. (2011) Calpain-1 cleaves Rad21 to promote sister chromatid separation. Mol. Cell. Biol. 31, 4335-4447

59. McNicoll, F., Stevense, M., and Jessberger, R. (2013) Cohesin in gametogenesis. Curr. Top. Dev. Biol. 102, 1-34

60. Wood, A. J., Severson, A. F., and Meyer, B. J. (2010) Condensin and cohesin complexity: the expanding repertoire of functions. Nat. Rev. Genet. 11, 391-404

61. McAleenan, A., Clemente-Blanco, A., Cordon-Preciado, V., Sen, N., Esteras, M., Jarmuz, A., and Aragon, L. (2013) Post-replicative repair involves separase-dependent removal of the kleisin subunit of cohesin. Nature 493, 250-254

62. Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001) Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature 410, 955-960

63. Marston, A. L. (2014) Chromosome segregation in budding yeast: sister chromatid cohesion and related mechanisms. Genetics 196, 31-63

64. Sullivan, M., Hornig, N. C., Porstmann, T., and Uhlmann, F. (2004) Studies on substrate recognition by the budding yeast separase. J. Biol. Chem. 279, 1191-1196

65. Zhang, N., Jiang, Y., Mao, Q., Demeler, B., Tao, Y. J., and Pati, D. (2013) Characterization of the interaction between the cohesin subunits Rad21 and SA1/2. PLoS One 8, e69458

66. Miller, M. P., Amon, A., and Unal, E. (2013) Meiosis I: when chromosomes undergo extreme makeover. Curr. Op. Cell Biol. 25, 687-696

67. Kitajima, T. S., Miyazaki, Y., Yamamoto, M., and Watanabe, Y. (2003) Rec8 cleavage by separase is required for meiotic nuclear divisions in fission yeast. EMBO J. 22, 5643-5653

68. Watanabe, Y., and Nurse, P. (1999) Cohesin Rec8 is required for reductional chromosome segregation at meiosis. Nature 400, 461-464

69. Eijpe, M., Offenberg, H., Jessberger, R., Revenkova, E., and Heyting, C. (2003) Meiotic cohesin REC8 marks the axial elements of rat synaptonemal complexes before cohesins SMC1-beta and SMC3. J. Cell. Biol. 160, 657-670

70. Lee, J., Iwai, T., Yokota, T., and Yamashita, M. (2003) Temporally and spatially selective loss of Rec8 protein from meiotic chromosomes during mammalian meiosis. J. Cell Sci. 116, 2781-2790

71. Lee, J., and Hirano, T. (2011) RAD21L, a novel cohesin subunit implicated in linking homologous chromosomes in mammalian meiosis. J. Cell Biol. 192, 263-276

72. Petronczki, M., Siomos, M. F., and Nasmyth, K. (2003) Un mènage à quatre: the molecular biology of chromosome segregation in meiosis. Cell 112, 423-440

73. Zickler, D., and Kleckner, N. (2015) Recombination, pairing, and synapsis of homologs during meiosis. Cold Spring Harb. Perspect. Biol. 7, a016626

74. Hess, R. A., and Renato de Franca, L. (2008) Spermatogenesis and cycle of the seminiferous epithelium. Adv. Exp. Med. Biol. 636, 1-15

75. Hauf, S., Waizenegger, I. C., and Peters, J.-M. (2001) Cohesin cleavage by separase required for anaphase and cytokinesis in human cells. Science 293, 1320-1323

76. Kudo, N. R., Anger, M., Peters, A. H., Stemmann, O., Theussl, H. C., Helmhart, W., Kudo, H., Heyting, C., and Nasmyth, K. (2009) Role of cleavage by separase of the Rec8 kleisin subunit of cohesin during mammalian meiosis I. J. Cell Sci. 122, 2686-2698

77. Kwon, Y. T., Xia, Z. X., An, J. Y., Tasaki, T., Davydov, I. V., Seo, J. W., Xie, Y., and Varshavsky, A. (2003) Female lethality and apoptosis of spermatocytes in mice lacking the UBR2 ubiquitin ligase of the N-end rule pathway. Mol. Cell. Biol. 23, 8255-8271

78. Lomeli, H., Ramos-Mejia, V., Gertsenstein, M., Lobe, C. G., and Nagy, A. (2000) Targeted insertion of Cre recombinase into the TNAP gene: excision in primordial germ cells. Genesis (New York, N.Y. : 2000) 26, 116-117

79. Wang, H., Wan, H., Li, X., Liu, W., Chen, Q., Wang, Y., Yang, L., Tang, H., Zhang, X., Duan, E., Zhao, X., Gao, F., and Li, W. (2014) Atg7 is required for acrosome biogenesis during spermatogenesis in mice. Cell Res. 24, 852-869

80. Hwang, C. S., and Varshavsky, A. (2008) Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 105, 19188-19193

81. Peters, A. H., Plug, A. W., van Vugt, M. J., and de Boer, P. (1997) A drying-down technique for the spreading of mammalian meiocytes from the male and female germline. Chromosome Res. 5, 66-68

82. Song, Z. H., Yu, H. Y., Wang, P., Mao, G. K., Liu, W. X., Li, M. N., Wang, H. N., Shang, Y. L., Liu, C., Xu, Z. L., Sun, Q. Y., and Li, W. (2015) Germ cell-specific Atg7 knockout results in primary ovarian insufficiency in female mice. Cell death & disease 6, e1589

83. Storlazzi, A., Xu, L., Cao, L., and Kleckner, N. (1995) Crossover and noncrossover recombination during meiosis: timing and pathway relationships. Proc. Natl. Acad. Sci. USA 92, 8512-8526

84. Rossitto, M., Philibert, P., Poulat, F., and Boizet-Bonhoure, B. (2015) Molecular events and signalling pathways of male germ cell differentiation in mouse. Semin. Cell Dev. Biol. 45, 84-93

85. Ohkura, H. (2015) Meiosis: an overview of key differences from mitosis. Cold Spring Harb. Perspect. Biol. 7

86. Tachibana-Konwalski, K., Godwin, J., Borsos, M., Rattani, A., Adams, D. J., and Nasmyth, K. (2013) Spindle assembly checkpoint of oocytes depends on a kinetochore structure determined by cohesin in meiosis I. Curr. Biol. 23, 2534-2539

87. Kouznetsova, A., Benavente, R., Pastink, A., and Hoog, C. (2011) Meiosis in mice without a synaptonemal complex. PLoS One 6, e28255

88. Yin, Y., Lin, C., Kim, S. T., Roig, I., Chen, H., Liu, L., Veith, G. M., Jin, R. U., Keeney, S., Jasin, M., Moley, K., Zhou, P., and Ma, L. (2011) The E3 ubiquitin

ligase Cullin 4A regulates meiotic progression in mouse spermatogenesis. Dev. Biol. 356, 51-62

89. de Boer, E., Dietrich, A. J., Hoog, C., Stam, P., and Heyting, C. (2007) Meiotic interference among MLH1 foci requires neither an intact axial element structure nor full synapsis. J. Cell Sci. 120, 731-736

90. Chapman, D. L., and Wolgemuth, D. J. (1994) Regulation of M-phase promoting factor activity during development of mouse male germ cells. Dev. Biol. 165, 500-506

91. Varshavsky, A. (2005) Ubiquitin fusion technique and related methods. Meth. Enzymol. 399, 777-799

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

93. Choi, W. S., Jeong, B.-C., Joo, Y. J., Lee, M.-R., Kim, J., Eck, M. J., and Song, H. K. (2010) Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat. Struct. Mol. Biol. 17, 1175-1181

94. Matta-Camacho, E., Kozlov, G., Li, F. F., and Gehring, K. (2010) Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat. Struct. Mol. Biol. 17, 1182-1188

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

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

97. Xia, Z., Webster, A., Du, F., Piatkov, K., Ghislain, M., and Varshavsky, A. (2008) Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. J. Biol. Chem. 283, 24011-24028

98. Kwon, Y. T., Xia, Z., Davydov, I. V., Lecker, S. H., and Varshavsky, A. (2001) Construction and analysis of mouse strains lacking the ubiquitin ligase UBR1 (E3-alpha) of the N-end rule pathway. Mol. Cell. Biol. 21, 8007-8021

99. Zenker, M., Mayerle, J., Lerch, M. M., Tagariello, A., Zerres, K., Durie, P. R., Beier, M., Hülskamp, G., Guzman, C., Rehder, H., Beemer, F. A., Hamel, B., Vanlieferinghen, P., Gershoni-Baruch, R., Vieira, M. W., Dumic, M., Auslender, R., Gil-da-Silva-Lopes, V. L., Steinlicht, S., Rauh, R., Shalev, S. A., Thiel, C., Winterpacht, A., Kwon, Y. T., Varshavsky, A., and Reis, A. (2005) Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). Nat. Genet. 37, 1345-1350

CHAPTER 6:

LIAT1, AN ARGINYLTRANSFERASE-BINDING PROTEIN WHOSE EVOLUTION AMONG PRIMATES INVOLVED CHANGES IN THE NUMBERS OF ITS 10-RESIDUE REPEATS

From Brower, C.S., Rosen, C.E., Jones, R.H., Wadas, B.C., Piatkov, K., and Varshavsky, A. (2014) *PNAS* 111(46):E4936-4945

ABSTRACT

The arginyltransferase Ate1 is a component of the N-end rule pathway, which recognizes proteins containing N-terminal degradation signals called N-degrons, polyubiquitylates these proteins and thereby causes their degradation by the proteasome. At least six isoforms of mouse Ate1 are produced through alternative splicing of *Ate1* pre-mRNA. We identified a previously uncharacterized mouse protein, termed Liat1 (ligand of Ate1), that interacts with Ate1 but does not appear to be its arginylation substrate. Liat1 has a higher affinity for the isoforms $Ate1^{1A7A}$ and $Ate1^{1B7A}$. Liat1 stimulated the *in vitro* N-terminal arginylation of a model substrate by Ate1. All examined vertebrate and some invertebrate genomes encode proteins sequelogous (similar in sequence) to mouse Liat1. Sequelogs of Liat1 share a highly conserved ~30-residue region that is shown here to be required for the binding of Liat1 to Ate1. We also identified non-Ate1 proteins that interact with Liat1. In contrast to *Liat1* genes of non-primate mammals, *Liat1* genes of primates are subtelomeric, a location that tends to confer evolutionary instability on a gene. Remarkably, Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, whereas Liat1 proteins of other mammals contain a single copy of this motif. Quantities of these repeats are, in general, different in Liat1 of different primates. For example, there are 1, 4, 13, 13, 17, and 17 repeats in the gibbon, gorilla, orangutan, bonobo, neanderthal, and human Liat1, respectively, suggesting that repeat number changes in this previously uncharacterized protein may contribute to evolution of primates.

INTRODUCTION

The N-end rule pathway recognizes proteins containing N-terminal degradation signals called N-degrons, polyubiquitylates these proteins, and thereby causes their degradation by the proteasome (Fig. 1A, B) (1-9). The main determinant of an N-degron is a destabilizing N-terminal residue of a protein. Recognition components of the N-end rule pathway are called N-recognins. In eukaryotes, N-recognins are E3 ubiquitin (Ub) ligases that can target N-degrons (Fig. 1). Bacteria also contain a (Ub-independent) version of the N-end rule pathway (10, 11).

Regulated degradation of proteins or their fragments by the N-end rule pathway mediates a strikingly broad range of functions, including the sensing of heme, nitric oxide, oxygen, and short peptides; control of protein quality and subunit stoichiometries, including the elimination of misfolded proteins; regulation of G proteins; repression of neurodegeneration; regulation of apoptosis, chromosome cohesion/segregation, transcription, and DNA repair; control of peptide import; regulation of meiosis, autophagy, immunity, fat metabolism, cell migration, actin filaments, cardiovascular development, spermatogenesis, and neurogenesis; the functioning of adult organs, including the brain, muscle and pancreas; and the regulation of many processes in plants (4-9) (Fig. 1A, B; see *Supporting Information* for a supplementary legend and references to this figure).

In eukaryotes, the N-end rule pathway consists of two branches. One branch, called the Ac/N-end rule pathway, targets proteins for degradation through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues (Fig. 1B) (2, 3, 12). Degradation signals and E3 Ub ligases of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-

recognins, respectively. Nt-acetylation of cellular proteins is apparently irreversible, in contrast to acetylation-deacetylation of internal Lys residues. Approximately 90% of human proteins are cotranslationally Nt-acetylated by ribosome-associated Nt-acetylases (13). Posttranslational Nt-acetylation occurs as well. Many, possibly most, Nt-acetylated proteins contain Ac/N-degrons (Fig. 1B) (2-4, 12).

The pathway's other branch, called the Arg/N-end rule pathway, targets specific unacetylated N-terminal residues (Fig. 1A) (3, 14-16). The "primary" destabilizing N-terminal Arg, Lys, His, Leu, Phe, Tyr, Trp, and Ile are directly recognized by N-recognins. The unacetylated N-terminal Met, if it is followed by a bulky hydrophobic ($) residue, also acts as a primary destabilizing residue (Fig. 1A) (3). In contrast, unacetylated N-terminal Asn, Gln, Asp, and Glu (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications, which include N-terminal deamidation (Nt-deamidation) of Asn and Gln and Nt-arginylation of Asp, Glu and oxidized Cys (Fig. 1A) (4, 6, 7, 17).

Nt-arginylation is mediated by the *Ate1*-encoded arginyltransferase (Arg-tRNA-protein transferase; R-transferase), a component of the Arg/N-end rule pathway and a subject of the present study (Fig. 1A, D) (18-23). Alternative splicing of mouse *Ate1* pre-mRNAs yields at least six R-transferase isoforms, which differ in their Nt-arginylation activity (Fig. 1C, D). (18, 21). R-transferases are sequelogous (similar in sequence) (24) throughout most of their ~60 kDa spans from fungi to mammals (4). R-transferase can arginylate not only N-terminal Asp and Glu but also N-terminal Cys, if it has been oxidized to Cys-sulfinate or Cys-sulfonate, through reactions mediated by NO, oxygen, and N-terminal Cys-oxidases (6, 20). The resulting circuits can act as sensors of NO and

oxygen in a cell through reactions that start with a conditional oxidation of N-terminal Cys in proteins such as the Rgs4, Rgs5, and Rgs16 regulators of G proteins in mammals (4, 20, 25) and specific transcriptional regulators in plants (reviewed in (4, 6, 7)).

There are dozens of either identified Nt-arginylated proteins (including natural protein fragments) or proteins that are predicted to be Nt-arginylated. Many, possibly most, of these proteins are conditionally or constitutively short-lived substrates of the Arg/N-end rule pathway (Fig. 1A) ((14, 15, 20, 26) and refs. therein). In contrast, there were, until now, no analyzed protein ligands of R-transferase that did not appear to be its substrates. We describe a previously uncharacterized mouse protein, termed Liat1 (ligand of Ate1) that binds to the mouse Ate1 R-transferase, is apparently not arginylated by it, and has a higher affinity for specific splicing-derived Ate1 isoforms. We also identified proteins other than R-transferase that appear to interact with Liat1. Remarkably, Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, in contrast to a single copy of this motif in Liat1 of other mammals. Quantities of repeats in Liat1 are, in general, different among different primates, suggesting that repeat number changes in this previously uncharacterized protein may contribute to evolution of primates.

RESULTS

Alternative Ate1 Exons and Their Sequelogies (Sequence Similarities). The 59 kDa mouse R-transferase is encoded by *Ate1*. This gene, located on chromosome 7 (7-F3), contains 14 protein-coding exons that encompass, together with introns, ~128 kb of *Ate1* DNA (Fig. 1C, D) (18, 21, 23, 27). Mouse *Ate1* encodes at least six splicing-

derived Ate1 isoforms. Fig. 1D shows designations and exon compositions of the four major isoforms. They are enzymatically active R-transferases whose levels of expression vary among different mouse tissues (21). The splicing of *Ate1* pre-mRNAs involves transcription from two distinct *Ate1* promoters and the alternative utilization of two exon pairs: (i) either exon 1A or 1B; and (ii) either exon 7A or 7B (Fig. 1C, D). The alternative exons 1A and 1B encode two sequelogous (similar in sequence) (24) ~30-residue N-terminal regions of R-transferase (Figs. 1D and S1) (21). The alternative exons 7A and 7B encode two sequelogous 43-residue regions of R-transferase (Figs. 1D and S1).

Two-Hybrid Detection of a Protein That Interacts with Ate1. In a search for mouse proteins that specifically interact with mouse R-transferase without being arginylation substrates, we employed a yeast-based two-hybrid (Y2H) assay, using a mouse testis cDNA library and a Gal4 DNA-binding domain (DBD)-Ate1^{1B7A} fusion as bait. This screen identified 9 independent positive DNA clones encoding different (overlapping) segments of a previously uncharacterized mouse protein (LOC74230; Acc. NP_941039) (Fig. 2A). This 228-residue protein, encoded by a single-copy mouse gene, was termed Liat1 (ligand of Ate1) (Fig. 3A).

To verify and expand these results, we expressed, in *S. cerevisiae* of appropriate genetic backgrounds, Gal4-DBD fusions to three mouse Ate1 isoforms, Ate1^{1B7A}, Ate1^{1A7A}, or Ate1^{1B7B}, and a Gal4 activation domain (AD) fusion to the full-length mouse Liat1. These strains were assayed for their ability to grow on minimal (SD) media lacking Leu, Trp, His, and Ade, thereby indicating the presence or absence of interactions among the examined protein fusions. The results were in agreement with the findings by the

initial two-hybrid screen in that the Ate1^{1B7A} R-transferase isoform interacted with Liat1 (Figs. 1D and 2A).

However, the two-hybrid assay did not detect interactions of Liat1 with the Ate1^{1A7A} and Ate1^{1B7B} isoforms, in contrast to the Ate1^{1B7A} isoform (Figs. 1D and 2A). This result was unexpected because differences among Ate1 isoforms are significant but not large, given sequelogies (24) between the encoded sequences of the Ate1 exons 1A vs. 1B (36% identity) and 7A vs. 7B (30% identity), respectively (Figs. 1D and S1). Although the affinity of Ate1^{1A7A} and Ate1^{1B7B} for Liat1 was too low for detection by two-hybrid assays (in contrast to the affinity of Ate1^{1B7A}) (Figs. 1D and 2A), an immunoprecipitation assay, described below, did detect a complex between Liat1 and the Ate1^{1A7A} isoform. One ramification of these two-hybrid results was a high likelihood that the interaction between Liat1 and Ate1^{1B7A} was specific, given the ability of Liat1 to distinguish, in its binding patterns, among three similar Ate1 isoforms (Figs. 1D, 2A, and S1).

A Region of Liat1 Required for Interaction with Ate1. The yeast two-hybrid assay was also employed, using the mouse Ate1^{1B7A} isoform and either N-terminal/C-terminal truncations or internal deletions of mouse Liat1, to delineate a region of Liat1 that was required for the observed interaction (Figs. 2A and 3B). Liat1$^{113-165}$, a 53-residue internal segment of the 228-residue Liat1 that contained a particularly strongly conserved ~30-residue region (termed the Liat1 domain; see below and Fig. A3.S2) was sufficient for the Liat1-Ate1^{1B7A} interaction, whereas the Liat1 domain alone did not suffice (Fig. 3B). It was also found that only a part of the Liat1 domain was

required for the binding of a C-terminally truncated Liat1 to Ate1^{1B7A}, provided that a region immediately upstream of the Liat1 domain was present as well (Fig. 3B).

Coimmunopreciptation Assays. We also transiently coexpressed, in mouse *Ate1*$^{-/-}$ embryonic fibroblasts (EFs) lacking R-transferase (19), the N-terminally triple ha-tagged mouse Liat1 (^{3ha}Liat1) and one of the four untagged mouse R-transferases, Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, or Ate1^{1B7B} (Figs. 1D and 2B). ^{3ha}Liat1 in cell extracts was immunoprecipitated with anti-ha antibody, followed by SDS-PAGE and immunoblotting with either anti-ha or the previously characterized, affinity-purified antibody to mouse Ate1 (20). The results of these coimmunoprecipitation (co-IP) assays were reproducible in independent assays and showed an efficacious co-IP of ^{3ha}Liat1 with Ate1^{1A7A} (Fig. 2B, lanes 6 and 8). A less efficacious but still detectable co-IP with Ate1^{1B7A} was also observed (Fig. 2B, lane 4 vs. lanes 6 and 8), whereas co-IP with the isoforms Ate1^{1A7B} and Ate1^{1B7B} was negligible (Fig. 2B).

We also carried out co-IP assays with recombinant proteins that had been expressed in *E. coli* using the Ub fusion technique and were purified by affinity chromatography that included the removal of Ub moiety (28, 29). These procedures yielded purified ^{3ha}Liat1 and purified, untagged Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B} or Ate1^{1B7B}. Equal amounts of Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B} or Ate1^{1B7B} were incubated with a ~4-fold molar excess of purified full-length ^{3ha}Liat1, followed by immunoprecipitation with anti-Ate1, SDS-PAGE, and immunoblotting with either anti-ha or anti-Ate1 (Fig. 2C). These “carrier-free” co-IP assays with purified proteins (instead of cell extracts containing these proteins) detected interactions between Liat1 and all four Ate1 isoforms (Fig. 2C). Note, however, that co-IPs of purified ^{3ha}Liat1 with purified Ate1^{1B7A} and

Ate1^{1A7A} were significantly more efficacious than co-IPs with the other two isoforms, Ate1^{1A7B} and Ate1^{1B7B} (Fig. 2C, lanes 1, 3 vs. lanes 2, 4). These results were in qualitative agreement with the findings by both two-hybrid assays and co-IP assays with cell extracts, neither of which could detect Liat1 interactions with the isoforms Ate1^{1A7B} and Ate1^{1B7B} (Fig. 2B, C). In addition, co-IPs with purified Liat1 and purified Ate1^{1B7A} or Ate1^{1A7A} (Fig. 2C) indicated that Liat1-Ate1 interactions did not require other proteins.

We do not understand the cause of the reproducible difference between Ate1^{1B7A} as the only Liat1-interacting Ate1 isoform in two-hybrid assays (Figs. 2A and 3B), in contrast to Ate1^{1A7A} as the apparently highest-affinity Liat1-interacting Ate1 isoform in co-IP assays with cell extracts (Fig. 2B). A plausible possibility is a different pattern of modifications (e.g., phosphorylation) of Ate1^{1B7A} vs. Ate1^{1A7A} that influences the outcomes of two different kinds of binding experiments (protein fusions in yeast-based Y2H assays *in vivo* vs. epitope-tagged ^{3ha}Liat1 and untagged Ate1 in co-IP assays with mammalian cell extracts *in vitro*). The other two isoforms, Ate1^{1A7B} and Ate1^{1B7B}, were negative in regard to interactions with Liat1 in both two-hybrid assays and co-IP assays with cell extracts (Figs. 2B and 3B). The non-interacting (or weakly interacting) Ate1^{1A7B} and Ate1^{1B7B} isoforms both contained the 7B exon, in contrast to the other two Ate1 isoforms (Figs. 1D and S1).

Sequence Features of the Mouse Liat1 Protein. The mouse *Liat1* cDNA (1700016K19Rik; Acc. NM_198637) encodes a 228-residue (25.5 kDa), previously uncharacterized protein (NP_941039) with a deduced pI of 7.45 (Fig. 3A). The *Liat1* gene, located on the mouse chromosome 11-B5, contains a single 2,737 bp intron between two protein-coding exons, and no alternative mouse *Liat1* cDNA isoforms could

be detected in databases (Figs. 3A and 4). Amino acid sequence features of mouse Liat1 include a negatively charged 10-residue region (9 of 10 residues are Glu) and a positively charged 12-residue region (11 of 12 residues are basic, largely Lys) (Fig. 3A). These regions are present in Liat1 of all examined mammals, but Liat1 of other vertebrates can lack either one of the two charged regions or both of them (Fig. 4B).

An aspect of Liat1 that is particularly conserved in evolution is a ~30-residue region termed the Liat1 domain (Figs. 3, 4, and S2-S4). The 32-residue mouse and human Liat1 domains are 96% identical (Figs. S2 and S3). Genomes of all examined vertebrates, and of some invertebrates as well, encode proteins sequelogous to mouse Liat1 (Figs. 4 and S2-S4). All sequelogs of Liat1 share at least the ~30-residue Liat1 domain. In fact, among vertebrates other than mammals and in Liat1-containing invertebrates as well, the Liat1 domain is often the only region that identifies a protein as a sequelog of mammalian Liat1 (Figs. 3, 4, S2-S4).

The extent of conservation of the Liat1 domain among mammals (>95%) and a weaker but still considerable conservation of this domain between, e.g., mammals and invertebrates such as sea anemone or acorn worm (Figs. 4 and S2-S4) indicate a purifying selection that maintained this Ate1-interacting domain (of unknown function) in the course of animal evolution. At the same time, large clades of organisms, including plants and fungi, lack proteins that can be identified as Liat1 through sequelogies (sequence similarities) alone, as distinguished from still possible spalogies (spatial similarities) (24).

Tandem repeats of a 10-residue motif in Liat1 of primates. Yet another feature of mammalian Liat1 is a 10-residue sequence immediately downstream of the Liat1 domain (Figs. 3, 4, and S3). This 10-residue motif is present as a single copy in non-

primate mammalian Liat1 proteins, but in Liat1 of some primates, including humans, this sequence is tandemly repeated (Figs. 4 and S3). For example, there are 4, 13, 13, 17, 17, and 18 repeats in the, gorilla, orangutan, bonobo, neanderthal, human, and baboon Liat1, respectively (Figs. 4 and S3). At the same time, some primates, such as gibbon and bushbaby, contain just one copy of the 10-residue motif, similarly to non-primate mammals (Figs. 4 and S3). The specific sequences of 10-residue repeats in a tandem array are either identical or nearly identical to each other both within a given Liat1 and among Liat1 proteins of different primate species (e.g., Fig. A3.S3).

The probability of mutations that alter the quantity of repeats in a gene can be orders of magnitude higher than the probability of, e.g., missense mutations (see Discussion) (30, 31). As a result, genetic variation that stems from repeat number changes in specific proteins can greatly exceed variation caused by other changes. This difference would be even higher if a gene in question is subtelomeric (see Discussion) (31). *Liat1* genes are subtelomeric in all examined primates, including humans (32), and most (though not all) primate Liat1 proteins contain tandem repeats of the 10-residue motif (Figs. 4 and S3). In contrast, *Liat1* genes of non-primate mammals are not subtelomeric, and all such Liat1 proteins contain one copy of the 10-residue motif (Fig. 4A).

The possibility that differences, among primates, in the quantities of their 10-residue repeats (Fig. 4) signify a role for these repeats in primate evolution is considered in Discussion. In contrast to non-primate *Liat1*, still unresolved aspects of subtelomeric primate *Liat1* genes are complicated enough to preclude definitive conclusions about expression patterns of, e.g., human *Liat1* (called *C17orf97* in databases) until its DNA,

pre-mRNAs and mature mRNAs are extensively characterized. For example, current descriptions of human *Liat1* (*C17orf97*) (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=C17orf97) are consistent with the existence of at least one additional intron in the region of human *Liat1* repeats, in contrast to mouse *Liat1*, which lacks both a second intron and the tandem repeats (Figs. 3A, 4, and S3) (see Discussion).

Expression of Liat1 in Mouse Tissues and in Mouse or Human Cell Lines. A commercial antibody to human C17orf97 (Liat1) was raised in rabbits against a 93-residue internal segment that included the highly conserved 32-residue human Liat1 domain, which is 96% identical to its counterpart in mouse Liat1 (Figs. 3A, S2, and S3). In immunoblots of extracts from mouse tissues or NIH-3T3 mouse cell line, this anti-Liat1 antibody detected largely a single band at the expected ~26 kDa M_r of mouse Liat1 in the heart, kidney, liver, spleen, testis, lung, thymus, pancreas, and brown adipose tissue, and in NIH-3T3 cells as well (Fig. 5B, lanes 1-5, and Fig. 5C, lanes 1-9). Immunoblotting with decreasing amounts of purified recombinant (untagged) mouse Liat1 indicated that this antibody to human Liat1 could detect down to ~50 ng of mouse Liat1 per lane (Fig. 5A). At this (moderate) sensitivity, little if any Liat1 was detected in the total mouse brain, hippocampus, cerebellum, and white adipose tissue (Fig. 5B, lane 6, and Fig. 5C, lanes 1-3). We note that histochemical data in the Human Protein Atlas (HPA023583), produced through the use of the same antibody, suggested the presence of Liat1 (C17orf97) in the human brain (http://www.proteinatlas.org/ENSG00000187624/tissue). *In situ* hybridization of a mouse 1700016K19Rik (*Liat1*) cDNA to sections of mouse brain indicated the presence of *Liat1*

mRNA at least in the cerebellum (http://mouse.brain-map.org/experiment/show?id=69114594).

Immunoblotting with extracts from human HeLa and HEK293T cells detected two Liat1 species, of ~55 kDa and ~50 kDa (Fig. 5C, lanes 10, 11). Human Liat1 is predicted to be 423 residues long (46.4 kDa), owing to 17 tandem repeats, in human Liat1, of a 10-residue sequence, in contrast to a single copy of this sequence in the 228-residue (25.5 kDa) mouse Liat1 and other non-primate Liat1 proteins (Fig. 3, 4, and Fig. 5C, lanes 10, 11). A smaller of two human Liat1 species in Fig. 5C (lanes 10, 11) may be either a cleavage product of the larger human Liat1 or a pre-mRNA splicing-derived Liat1 isoform. Remarkably, this human Liat1 immunoblotting pattern also contained a ~26 kDa band that comigrated with the ~26 kDa band recognized by the same antibody in mouse tissues (Fig. 5C, lanes 10, 11 vs. lanes 4-9). These results are consistent with the existence of human Liat1 isoforms in which quantities of the 10-residue repeat can vary between ~17 and 1, a functionally remarkable possibility that will be followed up through a more detailed understanding of primate *Liat1* genes.

Effect of Liat1 on Ate1-Mediated Nt-Arginylation. The previously characterized *in vitro* Nt-arginylation assay (21) used [^{14}C]-L-Arg, purified mouse Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B} and Ate1^{1B7B} isoforms, an Arg-tRNA-generating system, and either bovine ±-lactalbumin (it bears the Nt-arginylatable N-terminal Glu residue; Fig. A3.S5B) or purified, C-terminally tagged recombinant reporters X-DHFRbt (X=Asp, Cys or Arg-Cys) based on the mouse dihydrofolate reductase (DHFR (the “bt” tag is described in (17)). X-DHFRbt contained the 6-residue sequence KGLAGL immediately after the N-terminal X residue (X=Asp, Cys or Arg-Cys), followed by the DHFRbt

moiety. KGLAGL is the sequence of mouse Rgs4 (a physiological Arg/N-end rule substrate; see Introduction) immediately after its wild-type N-terminal Cys (20).

Initially, identical amounts of the individual purified Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B} and Ate1^{1B7B} isoforms were incubated for 60 min at 37°C with the rest of assay's components, including [^{14}C]-L-Arg and X-DHFRbt (X=Asp, Cys, or Arg-Cys), and either with or without Ate1-lacking extract from mouse *Ate1*$^{-/-}$ embryonic fibroblasts (EFs), followed by SDS-PAGE and autoradiography (Fig. A3.S5A). As expected (21), all Ate1 R-transferase isoforms could Nt-arginylate the N-terminal Asp residue of Asp-DHFRbt, with Ate1^{1B7B} being most active, and with Ate1^{1B7A}, Ate1^{1A7A}, and Ate1^{1A7B} exhibiting, respectively, ~66%, ~65%, and ~12% of the activity of Ate1^{1B7B} (Fig. A3.S5A). No Nt-arginylation of the "pre-arginylated" Arg-Asp-DHFRbt was observed under any conditions (Fig. A3.S5A), as expected, given the specificity of R-transferase (Fig. 1A) (4). There was no detectable Nt-arginylation of Cys-DHFRbt in the absence of extract from *Ate1*$^{-/-}$ EF cells, but this reporter was Nt-arginylated in the presence of extract, owing, presumably, to the oxidation of N-terminal Cys by compounds in cell extracts (7, 20).

The Liat1-supplemented version of this assay (without extract from *Ate1*$^{-/-}$ EF cells) employed X-DHFRbt (X=Asp, Cys). Each of the Ate1^{1A7A}, Ate1^{1B7A}, Ate1^{1A7B}, and Ate1^{1B7B} isoforms was pre-incubated for 30 min at 37°C with a ~4-fold molar excess of purified untagged mouse Liat1 or with buffer alone before their addition to the arginylation assay. The Nt-arginylation of Asp-DHFRbt by Ate1 was reproducibly enhanced in the presence of Liat1, from 1.3-fold to 2.3-fold (Fig. A3.S5C). This effect is unlikely to stem from increased macromolecular crowding upon the addition of Liat1,

because the addition of equal or larger amounts of bovine serum albumin did not alter the efficacy of Nt-arginylation. The unknown mechanistic cause of moderate but reproducible effects of Liat1 on Nt-arginylation is a subject for future studies.

Significantly, we could not detect, despite attempts to do so, the conjugation of ^{14}C-Arg to Liat1 itself in this assay, in either the presence or absence of test proteins such as ±-lactalbumin or X-DHFRbt. (Upon SDS-PAGE, the untagged mouse Liat1 migrates significantly below the band of X-DHFRbt.) Even overexposures of autoradiograms of electrophoretically fractionated proteins after ^{14}C-arginylation in the presence of purified Liat1 did not reveal any significant ^{14}C in the vicinity of the ~26 kDa Liat1 band, indicating that Liat1 is not an Ate1 substrate.

Identification of Liat1-Binding Proteins Other Than Ate1. We also carried out a search for mouse proteins other than Ate1 that interact with mouse Liat1 (Fig. 3A), using a library of mouse brain cDNAs fused to the Gal4 activation domain (AD) vis-á-vis *Liat1* cDNA fused to the Gal4 DNA-binding domain (DBD). This screen identified ~40 different mouse proteins that appeared to bind to mouse Liat1, with the corresponding cDNA isolates having passed standard controls of the Y2H assay (Fig. A3.S6C). In addition, an independent Y2H screen for interactions among protein methyltransferases and proteins that bind to them (including their substrates) identified the human Jmjd6 methyltransferase as a putative ligand of human C17orf97 (33), i.e., the human Liat1 protein (Fig. A3.S6C).

We also searched for protein interactions with Liat1using GST-pulldowns with GST-Liat1 and extracts from mouse EF cells (Fig. A3.S6A). These mass spectrometry (MS)-based assays added the ribosomal proteins S14 and S19 as well as three specific

histones to the Y2H-based list of other Liat1 ligands (Fig. A3.S6). Thus identified putative ligands of Liat1 encompassed a broad range of functional classes, including components of the translation, transcription, and Ub-proteasome systems. Our ongoing co-IP assays that independently verify the binding of Liat1 to its putative ligands (which were detected by Y2H or GST-pulldowns (Fig. A3.S6C)) have recently confirmed that Jmjd6 and the ribosomal protein S14 can be coimmunoprecipitated with mouse Liat1 from a cell extract (Fig. A3.S6B). These and other binding assays with putative ligands of Liat1 will continue to verify the current preliminary list (Fig. A3.S6), thereby making possible systematic analyses of confirmed Liat1 ligands in regard to specific functions of their interactions with Liat1.

DISCUSSION

We identified a 26 kDa mouse protein, termed Liat1 (ligand of Ate1), by detecting its binding to the 59 kDa mouse Ate1 R-transferase, a component of the Arg/N-end rule pathway (Figs. 1-3). The Liat1-Ate1 interaction was shown to require at least a part of the highly conserved ~30-residue region of Liat1 (Fig. 3B). Biological functions of Liat1-Ate1 interactions (Figs. 2 and 3), remain to be understood, in part because Liat1 is a previously uncharacterized protein. (In current databases, human Liat1 is called C17orf97.) We also identified putative Liat1-binding proteins other than Ate1 (Fig. A3.S6), but the biological function of Liat1 remains to be discovered.

Further analyses showed that Liat1 proteins of some primates, from macaques to humans, contain tandem repeats of a 10-residue sequence, in contrast to a single copy of this motif in Liat1 of other mammals, including rodents. Quantities of these repeats are,

in general, different in Liat1 of different primates. For example, there are 4, 13, 13, 17, and 17 repeats in the gorilla, orangutan, bonobo, neanderthal, and human Liat1, respectively (Figs. 4 and S3), suggesting that repeat number changes in Liat1 might play a role in evolution of primates. As evidence, these differences in the quantities of Liat1 repeats do not rise above a correlational argument, and the current disposition is further complicated by our finding of intra-species variability in the number of repeats. For example, the Liat1 proteins encoded by an anonymous human genome (NM_001013672) and by the genome of James D. Watson (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&l=C17orf97) were found to contain, each, 17 tandem repeats, whereas the genome of J. Craig Venter encodes Liat1 that contains 18 repeats (NP_001013694).

Analyses of primate genomes have identified, so far, relatively few alterations that could be demonstrated to contribute to phenotypic differences among these species. The “causative” alterations include variants of developmental enhancer DNA sequences as well as specific alleles of genes that influence circadian rhythms, human brain size, language, other human traits, and differences in body sizes among primates (34-37). On evolutionary timescales, the emergence of anatomical and behavioral features that distinguish humans from great apes was remarkably fast, as the last common ancestor of humans and chimpanzees lived about 6 million years ago.

The efficacy of selection pressure is limited by the extent of relevant genetic variation within a breeding population. We suggest, therefore, that the rapidity of evolution that led to humans may have involved proteins containing tandemly repeated sequences, given frequent changes in the numbers of such repeats. Analogous arguments

have been made both in general and to account for the strikingly rapid (in less than 1,000 years) emergence of modern dog breeds, in response to selection pressures imposed by breeders (30, 38). Morphological differences among dog breeds, and even within a breed, were found to correlate with variations in the numbers of 1-residue or 2-residue repeats in proteins that regulate embryonic development and postnatal growth (30).

Tandem repeats of amino acid sequences that range in size from 1 to more than 100 residues are a feature of many cellular proteins (30, 36, 38, 39). Heritable changes in the number of repeat units are known to underlie a significant fraction of phenotypic variability both among different species and within a species (36, 38). The previously demonstrated genetic instability of repeats stems from several causes, including unequal crossover, DNA replication slippage, and double-strand break repair (30, 31). In an evolving species, genetic variation that results from repeat number changes in specific proteins can greatly exceed variation that is caused, for example, by missense mutations. This difference would be even higher if a gene in question is subtelomeric, because proximity to a telomere tends to increase the frequency of recombination-mediated changes in the quantity of repeats in a gene (31).

Remarkably, *Liat1* genes are subtelomeric in all examined primates, including humans, and most primate Liat1 proteins contain tandem repeats of a 10-residue motif (Figs. 4 and S3). Potentially telling exceptions include the gibbon and bushbaby Liat1 proteins, which contain a single copy of the 10-residue motif that is repeated in other examined primates (Fig. 4B) (37). In contrast to primate *Liat1* genes, their counterparts in other mammals, including the mouse, are not subtelomeric (Fig. 4A), and all predicted non-primate Liat1 proteins in databases contain one copy of the 10-residue motif. The

function of Liat1 repeats (Fig. 4) and their possible relevance to anatomic and phenotypic evolution of primates, remain to be addressed.

One subtelomeric gene encoding a protein that contains tandem repeats is *Drd4* (40). A human dopamine receptor encoded by *Drd4* contains varying numbers of tandem repeats of a 16-residue sequence that forms the third cytosolic loop of the receptor. Both the quantity of 16-residue repeats and specific sequences within each repeat are often altered among individual humans, with repeat numbers varying between 2 and 10 (40). Analogous polymorphisms of these repeats were also observed in Drd4 of other primates (40). Human Drdr4 receptors with different repeat numbers were reported to differ in functional properties, and were also differentially regulated in diseases such as schizophrenia. Nevertheless, there is still no definitive evidence that the observed frequencies of *Drd4*-encoded isoforms containing different quantities of repeats had been caused, at least in part, by positive selection, as distinguished from a quasi-neutral drift (41).

In sum, a major unanswered question is whether changes in repeat quantities among Liat1 proteins of different primates stemmed, at least in part, from specific selection pressures. The alternative scenario is that most variation in Liat1 repeats (Figs. 4 and S3) may be the result of genetic drift and unselected fixations of altered repeat quantities, owing to a small effective population size of an evolving species, including occasional population bottlenecks. A smaller population is characterized by correspondingly weaker forces of natural selection. Near-neutral evolution of proteins under such conditions is discussed by Lynch (42). Because small changes in Liat1 repeat numbers may have modest phenotypic effects, a near-neutral drift would be likely to at

least contribute to evolution of repeats in Liat1. If so, it is the selection-based, adaptation-centered hypothesis about a functional role of repeat number changes in Liat1 that must be experimentally verified vis-á-vis the competing null hypothesis, in which these changes would be caused by a near-neutral genetic drift.

The understanding of both Liat1 itself and the functional significance of its binding to specific isoforms of the Ate1 R-transferase would be advanced by answers to at least the following questions. What are the composition and functions of *in vivo* complexes that contain Ate1, Liat1, and other macromolecular components? (Gel filtration of mouse cell extracts indicates that Ate1 and Liat1 can be parts of specific but not necessarily identical protein complexes (data not shown).) What is the biological function of Liat1? How does this function relate to the known role of the Ate1 R-transferase in the Arg/N-end rule pathway (Fig. 1A)? Does a divergent Liat1 protein, e.g., the one of sea anemone (it is identifiable as Liat1 solely through its ~30-residue Liat1 domain; Figs. S2 and S4), specifically bind to the sea anemone Ate1? (The binding of mouse Liat1 to mouse Ate1 requires the Liat1 domain and apparently does not involve the 10-residue motif that forms repeats in Liat1 of primates; Figs. 2, 3B, 4, and S2-S4.)

Furthermore: do repeats of the 10-residue motif in, for example, human Liat1 (Figs. 4 and S3) interact with each other? Do these repeats specifically bind to any human protein? If they do, does the binding of Liat1 repeats to that protein also involve other regions of human Liat1? Is there a counterpart of a human repeat-binding protein in, for example, mouse cells? (Mouse Liat1 contains one copy of the sequence that forms repeats in human Liat1.) As mentioned in Results, it would also be essential to understand, in detail, a primate (e.g., the human or macaque) *Liat1* gene, given its more

complex organization (including the presence of Liat1 isoforms) than the structure of repeat-lacking *Liat1* genes of non-primate mammals (Figs. 3A and 4).

The understanding of Liat1 would also benefit from constructing and characterizing mouse strains that either lack Liat1 or contain its counterpart in which the 10-residue motif had been amplified to yield primate-like tandem repeats. It would also be informative to determine, with a primate such as, for example, macaque, the phenotypic effects of strongly increased or strongly decreased quantities of 10-residue repeats in Liat1 (Fig. 4). (Gene-specific alterations of the macaque genome through the CRISPR technology have already been achieved (43).) Given the preferential binding of Liat1 to specific isoforms of the Ate1 R-transferase (Fig. 2), further studies of Liat1 may also advance the functional understanding of Ate1 isoforms. In addition, the obscurity, until the present study, of tandem repeats in primate Liat1, and still incomplete descriptions of subtelomeric loci among the sequenced metazoan genomes (owing to complexities of dealing with telomere-proximal microsatellite DNA repeats) suggest that it may be informative to further explore subtelomeric regions of the 23 human chromosomes for the presence of other uncharacterized genes that might encode repeats analogous to those of Liat1.

EXPERIMENTAL PROCEDURES

Yeast Two-Hybrid (Y2H) Assays. They were carried with *S. cerevisiae*, using the BD Matchmaker kit (BD Biosciences, Palo Alto, CA), the plasmid pCB132, which expressed the $Gal4^{DBD}$-$ATE1^{1B7A}$ fusion (Table S1), and a mouse testis cDNA pACT library (Clontech).

Construction and Expression of Recombinant Proteins in BL21 (DE3) *E. coli*. The untagged mouse Ate1^{1B7A}, Ate1^{1B7B}, Ate1^{1A7A}, and Ate1^{1A7B}, the untagged mouse Liat, and ^{3ha}Liat1 were expressed as Ub fusions in *E. coli*, followed by the removal of Ub and purification of recombinant proteins by Mono-S chromatography (see *SI Experimental Procedures*).

***In Vitro* Arginylation Assay.** The Nt-arginylation assay was performed essentially as described previously (21).

Tissue Extracts and Immunoblotting. Extract preparation and immunoblotting were carried out essentially as described previously (3, 12, 21).

GST pulldown assay and immunoprecipitations. Mouse EF cells were transiently transformed with a plasmid expressing either glutathione transferase (GST) or a GST-Liat1 fusion, and GST-pulldown assays were carried out as described in *SI Experimental Procedures*. Immunoprecipitations with anti-flag or anti-ha antibodies were performed as previously described (3, 12).

Additional information regarding experimental procedures is given in *SI Experimental Procedures*.

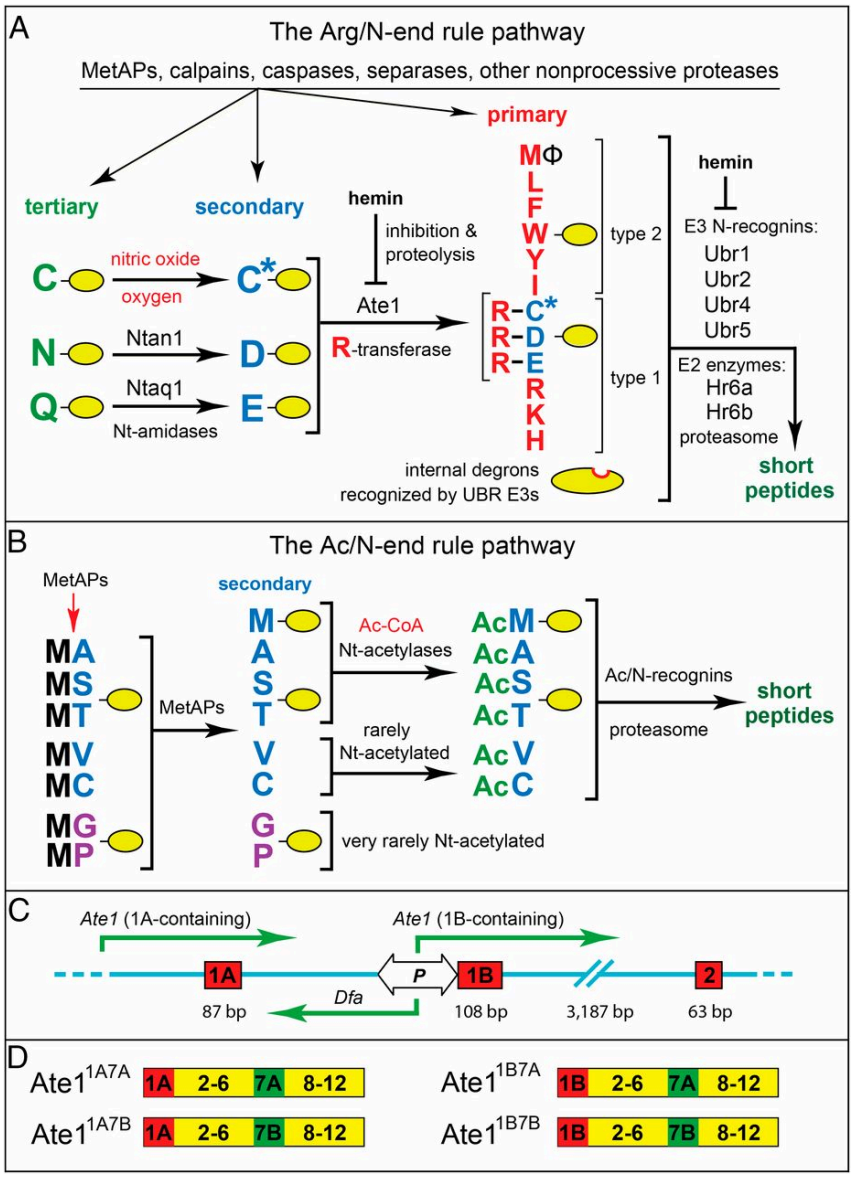


Fig. 6.1. The mammalian N-end rule pathway and the Ate1 arginyltransferase (R-transferase). See Introduction for a summary of pathway's mechanistic aspects and biological functions. N-terminal residues are denoted by single-letter abbreviations. A yellow oval denotes the rest of a protein substrate. (*A*) The Arg/N-end rule pathway. It recognizes proteins through their unacetylated N-terminal residues and contains the N-terminal arginylation (Nt-arginylation) branch, mediated by the arginyltransferase (R-transferase) Ate1 and by the Nt-amidases Ntan1 and Ntaq1 that act upstream of Ate1 (4-6). (*B*) The Ac/N-end rule pathway. It recognizes proteins through their $N^{\pm}$-terminally acetylated (Nt-acetylated) residues (2, 3, 12). Red arrow on the left indicates the removal of the N-terminal Met residue by Met-aminopeptidases (MetAPs). N-terminal Met is

retained if a residue at position 2 is larger than Val. (*C*) The bidirectional $_{Dfa}P_{Ate1}$ promoter upstream of exon 1B of the mouse *Ate1* gene (4, 21, 27). (*D*) Four major mouse Ate1 R-transferase isoforms and their designations (4, 21). See *SI Text* for a detailed supplementary legend and supplementary references to this figure.

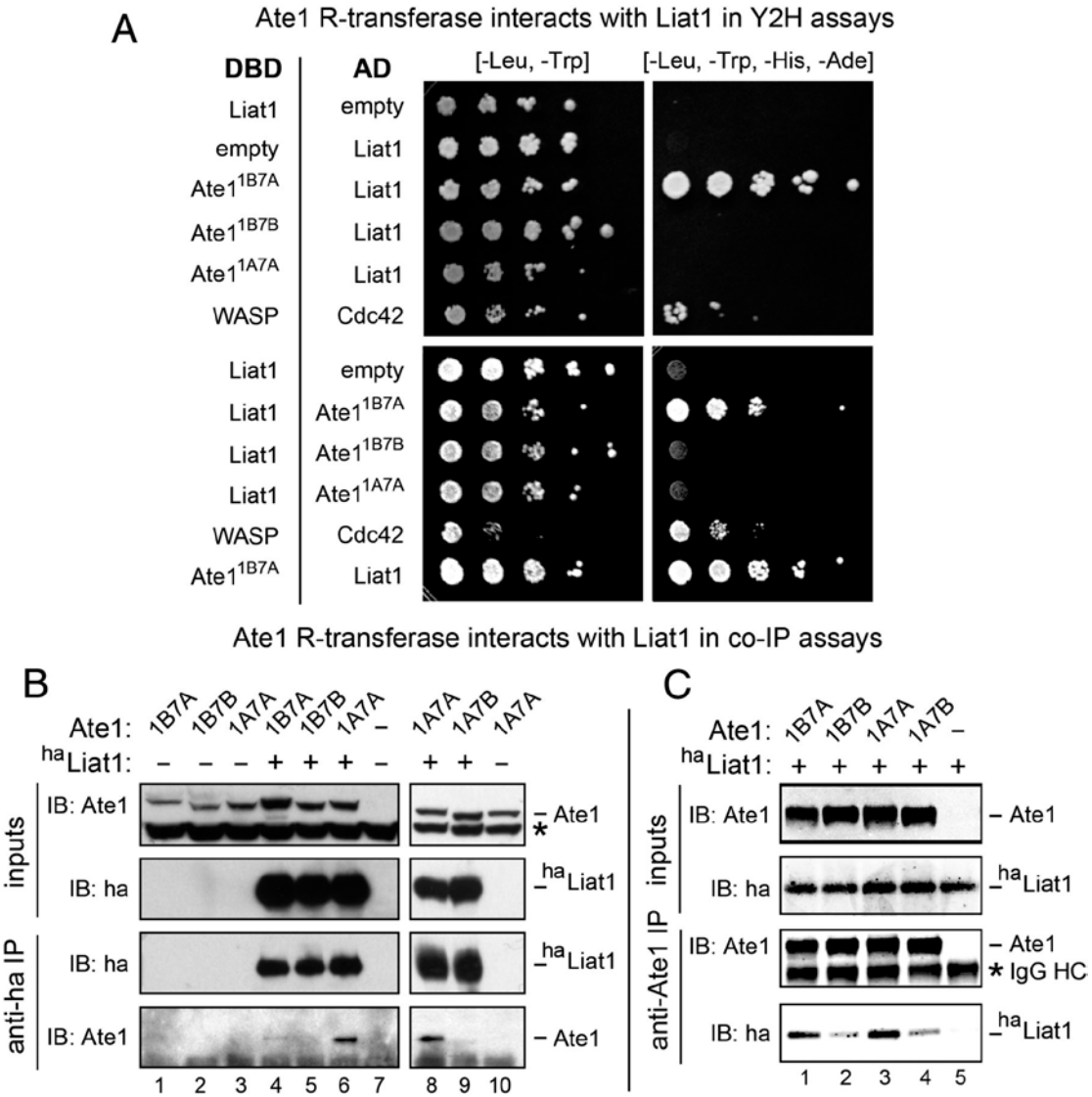


Fig. 6.2. Detection and characterization of interactions between the mouse Ate1 R-transferase and a previously uncharacterized mouse protein termed Liat1. (*A*) Detection of Ate1-Liat1 interactions using two-hybrid (Y2H) assay. A Gal4 DNA-binding domain (DBD)-Ate1^{1B7A} fusion and other DBD-based fusions (e.g., a control DBD fusion containing the WASP protein) vs. coexpressed Gal4 activation domain (AD) fusions to full-length Liat1 and to other proteins, including specific Ate1 isoforms as well as the control (WASP-interacting) Cdc42 protein. The ability to grow on minimal media lacking Leu, Trp, His and Ade was used to detect interactions among the examined protein fusions. See *SI Experimental Procedures*. (*B*) Detection of Ate1-Liat1 interactions using coimmunoprecipitation (co-IP) assays with mouse cell extracts. Plasmids expressing ^{3ha}Liat1 and/or specific Ate1 isoforms were transiently expressed in mouse *Ate1*$^{-/-}$ EF cells, followed by immunoprecipitations with anti-ha antibody, SDS-PAGE of immunoprecipitates, and immunoblotting with both anti-ha and anti-Ate1 antibodies. (*C*) Same as in *B* but using purified mouse ^{3ha}Liat1 and purified mouse Ate1 isoforms, with immunoprecipitation by anti-Ate1 antibody.

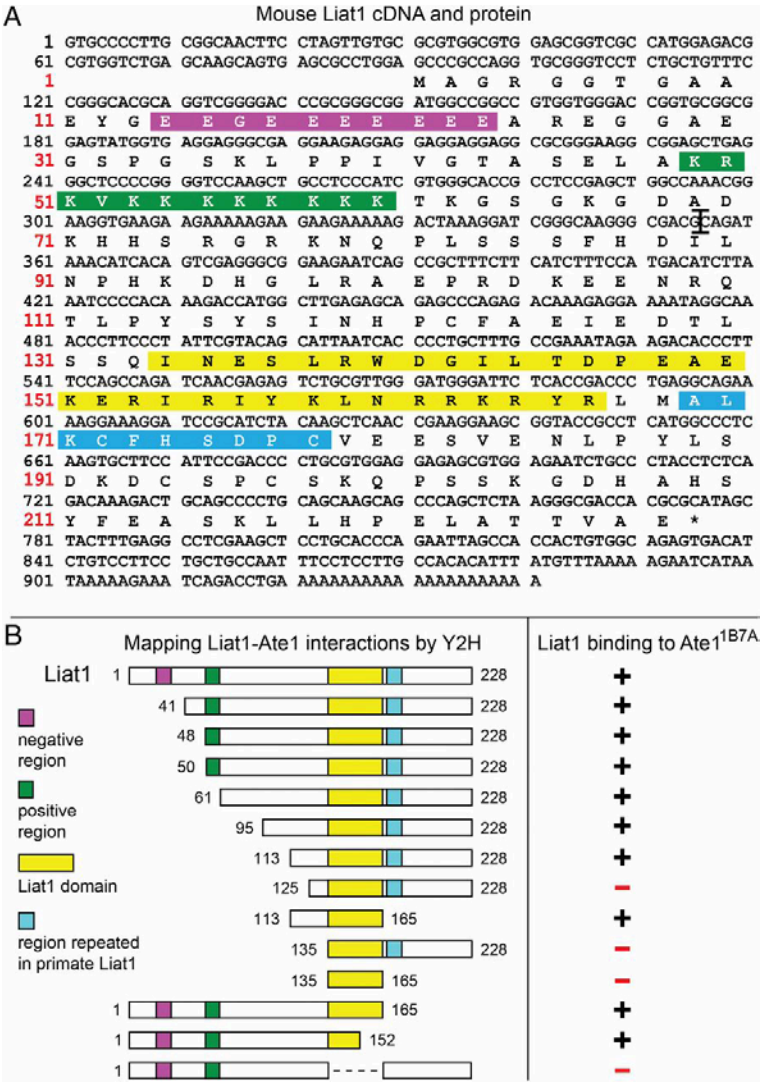


Fig. 6.3. Mouse Liat1 and the mapping of its Ate1-binding region. (*A*) Nucleotide sequence of mouse *Liat1* cDNA and amino acid sequence features of mouse Liat1 (NP_941039). A large black "**I**" after the Lys-rich region denotes the position of a single 2,737 bp intron between two protein-coding exons in the *Liat1* genomic DNA. The purple, green, yellow, and blue rectangles denote, respectively, the negatively charged region of mouse Liat1, its positively charged region, its particularly highly conserved ~30-residue domain, termed the Liat1 domain, and a 10-residue region that becomes tandemly repeated in Liat1 of some primates, including humans, but is a single-copy sequence in Liat1 of other mammals, including mouse Liat1. The same color scheme is used to denote these regions of Liat1 in other figures of this paper. Black and red numbers on the left indicate nucleotide and amino acid residue numbers, respectively. See also the main text and Figs. S2-S4. *(B)* Mapping Liat1-Ate1 interactions using two-hybrid (Y2H) assays. The Gal4 DNA-binding domain (DBD)-Ate1^{1B7A} fusion was examined, using Y2H, for interactions with the coexpressed Gal4 activation domain (AD) fusion to either full-length Liat1 or to the indicated AD fusions encoding Liat1 fragments. The results of this assay are summarized on the right, and the color coding of Liat1 domains (the same as in *A*) is indicated on the left.

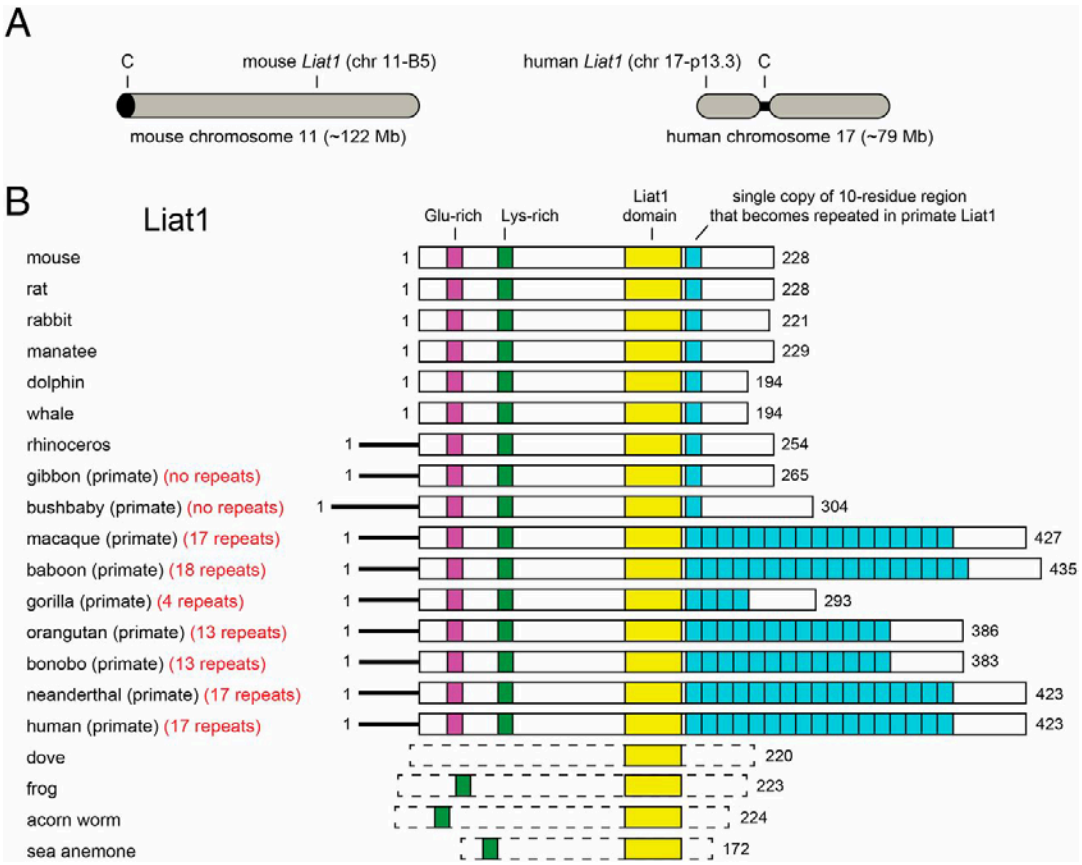


Fig. 6.4. Chromosomal locations of *Liat1* genes and evolution of Liat1 proteins. (*A*) *Liat1* genes are located away from telomeres in both mice (chromosome 11-B5) and other non-primate mammals. In contrast, *Liat1* genes are subtelomeric in all examined primates (e.g., human chromosome 17-p13.3), indicating a translocation of *Liat1* to a subtelomeric site in an ancestor of modern primates, from gibbons and bushbabies to macaques and humans. (*B*) Evolution of Liat1 proteins. The colors of rectangles (the same as in Fig. 3A) denote, respectively, the negatively charged region of Liat1 (purple), its positively charged region (green), its particularly highly conserved ~30-residue domain (the Liat1 domain) (yellow), and a 10-residue segment that becomes tandemly repeated in Liat1 of some primates (blue). In all examined Liat1 proteins that contain these regions they are present in the same (indicated) order. While each of these regions, in a given Liat1, is unambiguously recognizable upon inspection of its amino acid sequence, the "linear" distances between these regions are not identical even among mammals, and are particularly variable among non-mammalian Liat1 proteins (Figs. S3 and S4). This aspect of Liat1 is largely bypassed in these diagrams, whose chief aim is to highlight the presence or absence of specific regions and the emergence of tandem 10-residue repeats in Liat1 of some primates, including humans and great apes. Dashed and shifted rectangles in Liat1 of non-mammalian vertebrates and invertebrates signify a non-conservation of distances between specific domains of Liat1 in the indicated organisms, in comparison to a significant (though incomplete) conservation of inter-region distances among mammalian Liat1 proteins. See also the main text and Figs. S2-S4. For the accession numbers of specific Liat1 proteins and for Latin names of the cited animal species, see the legend to Fig. A3.S2.

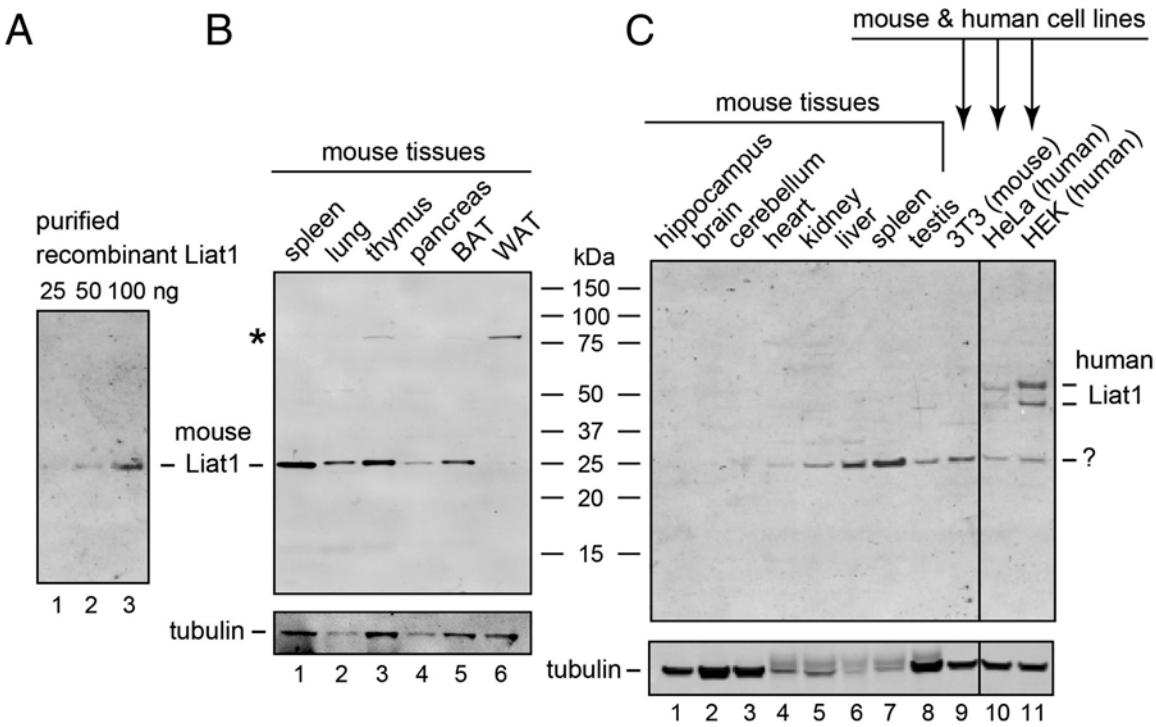


Fig. 6.5. Analyses of Liat1 proteins by immunoblotting. (*A*) Purified recombinant (untagged) mouse Liat1 was fractionated by SDS-PAGE, following by immunoblotting with an antibody to a highly conserved region of human Liat1 (C17orf97) (see the main text). Lanes 1-3: 25, 50, and 100 ng of mouse Liat1 per lane, respectively. (*B*) Same as in *A* but with extracts from the indicated mouse tissues that had been fractionated by SDS-PAGE. An asterisk indicates a larger, apparently unrelated (crossreacting) protein species in extracts from thymus and white adipose tissue (WAT). The results of immunoblotting with anti-tubulin antibody are shown as well. (*C*) Lanes 1-8, same as in *B*, with extracts from indicated mouse tissues. Lane 9, same as in lanes 1-8 but an extract from mouse NIH-3T3 cells. Lanes 10, 11, same as lane 8 but extracts from human HeLa and HEK293T cells. Larger human Liat1 proteins, corresponding, presumably, to species with multiple tandem repeats (see Fig. 4 and the main text) are indicated on the right. A question mark denotes a putative Liat1 species from human cell lines that comigrates with the much smaller (26 kDa) mouse Liat1 and may contain just one copy of the 10-residue motif that is tandemly repeated in other species of human Liat1 (see the main text).

REFERENCES

1. Bachmair A, Finley D, & Varshavsky A (1986) In vivo half-life of a protein is a function of its amino-terminal residue. Science 234:179-186.
2. Hwang C-S, Shemorry A, & Varshavsky A (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327:973-977.
3. Kim H-K, et al. (2014) The N-terminal methionine of cellular proteins as a degradation signal. Cell 156:158-169.
4. Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20:1298-1345.
5. Kim JM & Hwang CS (2014) Crosstalk between the Arg/N-end and Ac/N-end rule. Cell Cycle 13:1366-1367.
6. Tasaki TS, Sriram SM, Park KS, & Kwon YT (2012) The N-end rule pathway. Annu. Rev. Biochem. 81:261-289.
7. Gibbs DJ, Bacardit J, Bachmair A, & Holdsworth MJ (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. Trends Cell Biol. 24:603-611.
8. Dougan DA, Micevski D, & Truscott KN (2011) The N-end rule pathway: from recognition by N-recognins to destruction by AAA+ proteases. Biochim. Biophys. Acta 1823:83-91.
9. Varshavsky A (2008) Discovery of cellular regulation by protein degradation. J. Biol. Chem. 283:34469-34489.
10. Tobias JW, Shrader TE, Rocap G, & Varshavsky A (1991) The N-end rule in bacteria. Science 254:1374-1377.

11. Rivera-Rivera I, Román-Hernández G, Sauer RT, & Baker TA (2014) Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. Proc. Natl. Acad. Sci. USA 111:E3853-E3859.

12. Shemorry A, Hwang C-S, & Varshavsky A (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50:540-551.

13. Starheim KK, Gevaert K, & Arnesen T (2012) Protein N-terminal acettyltransferases: when the start matters. Trends Biochem. Sci. 37:152-161.

14. Brower CS, Piatkov KI, & Varshavsky A (2013) Neurodegeneration-associated protein fragments as short-lived substrates of the N-end rule pathway. Mol. Cell 50:161-171.

15. Piatkov KI, Oh J-H, Liu Y, & Varshavsky A (2014) Calpain-generated natural protein pragments as short-lived substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 111:E817-E826.

16. Yamano K & Youle RJ (2013) PINK1 is degraded through the N-end rule pathway. Autophagy 9:1758-1769.

17. Wang H, Piatkov KI, Brower CS, & Varshavsky A (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell 34:686-695.

18. Kwon YT, Kashina AS, & Varshavsky A (1999) Alternative splicing results in differential expression, activity, and localization of the two forms of arginyl-tRNA-protein transferase, a component of the N-end rule pathway. Mol. Cell. Biol. 19:182-193.

19. Kwon YT, et al. (2002) An essential role of N-terminal arginylation in cardiovascular development. Science 297:96-99.

20. Hu R-G, et al. (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437:981-986.

21. Hu R-G, et al. (2006) Arginyl-transferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem.:32559-32573.

22. Saha S & Kashina A (2011) Posttranslational arginylation as a global biological regulator. Dev. Biol. 358:1-8.

23. Brower CS & Varshavsky A (2009) Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE 4:e7757.

24. Varshavsky A (2004) Spalog and sequelog: neutral terms for spatial and sequence similarity. Curr. Biol. 14:R181-R183.

25. Lee MJ, et al. (2005) RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 102:15030-15035.

26. Varshavsky A (2012) Augmented generation of protein fragments during wakefulness as the molecular cause of sleep: a hypothesis. Prot. Sci. 21:1634-1661.

27. Brower CS, Veiga L, Jones RH, & Varshavsky A (2010) Mouse Dfa Is a Repressor of TATA-Box Promoters and Interacts with the Abt1 Activator of Basal Transcription. J. Biol. Chem. 285:17218-17234.

28. Catanzariti A-M, Soboleva TA, Jans DA, Board PG, & Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. Protein Sci. 13:1331-1339.

29. Varshavsky A (2005) Ubiquitin fusion technique and related methods. Meth. Enzymol. 399:777-799.

30. Fondon JWr & Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. Proc. Natl. Acad. Sci. USA 101:18058-18063.

31. Brown CA, Murray AW, & Verstrepen KJ (2010) Rapid expansion and functional divergence of subtelomeric gene families in yeasts. Curr Biol. 20:895-903.

32. Zody MC, et al. (2006) DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. Nature 440:1045-1049.

33. Weimann M, et al. (2013) A Y2H-seq approach defines the human protein methyltransferase interactome. Nat. Methods 10:339-342.

34. Enard W (2011) FOXP2 and the role of cortico-basal ganglia circuits in speech and language evolution. Curr. Opin. Neurobiol. 21:415-424.

35. Capra JA, Erwin GD, McKinsey G, Rubenstein JLR, & Pollard KS (2013) Many human accelerated regions are developmental enhancers. Phil. Trans. R. Soc. B 368:20130025.

36. Sabino FC, et al. (2014) Evolutionary History of the PER3 Variable Number of Tandem Repeats (VNTR): Idiosyncratic Aspect of Primate Molecular Circadian Clock. PLoS One 9:e107198.

37. Carbone L & al. e (2014) Gibbon genome and the fast karyotype evolution of small apes. Nature 513:195-201.

38. Gemayel R, Vinces MD, Legendre M, & Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. 44:445-477.

39. Verstrepen KJ, Jansen A, Lewitter F, & Fink GR (2005) Intragenic tandem repeats generate functional variability. Nat. Genet. 37:986-990.

40. Livak KJ, Rogers J, & Lighter JB (1995) Variability of dopamine D4 receptor (DRD4) gene sequence within and among nonhuman primate species. Proc. Natl. Acad. Sci. USA 92:427-431.

41. Hattori E, et al. (2009) Variable number of tandem repeat polymorphisms of DRD4: re-evaluation of selection hypothesis and analysis of association with schizophrenia. Eur. J. Hum. Genet. 17:793-801.

42. Lynch M (2007) The Origins of Genome Architecture (Sinauer Associates, Inc., Sunderland, MA).

43. Niu Y, et al. (2014) Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell 156:836-843.

APPENDIX 1:

SUPPLEMENTAL MATERIAL FOR CHAPTER 2

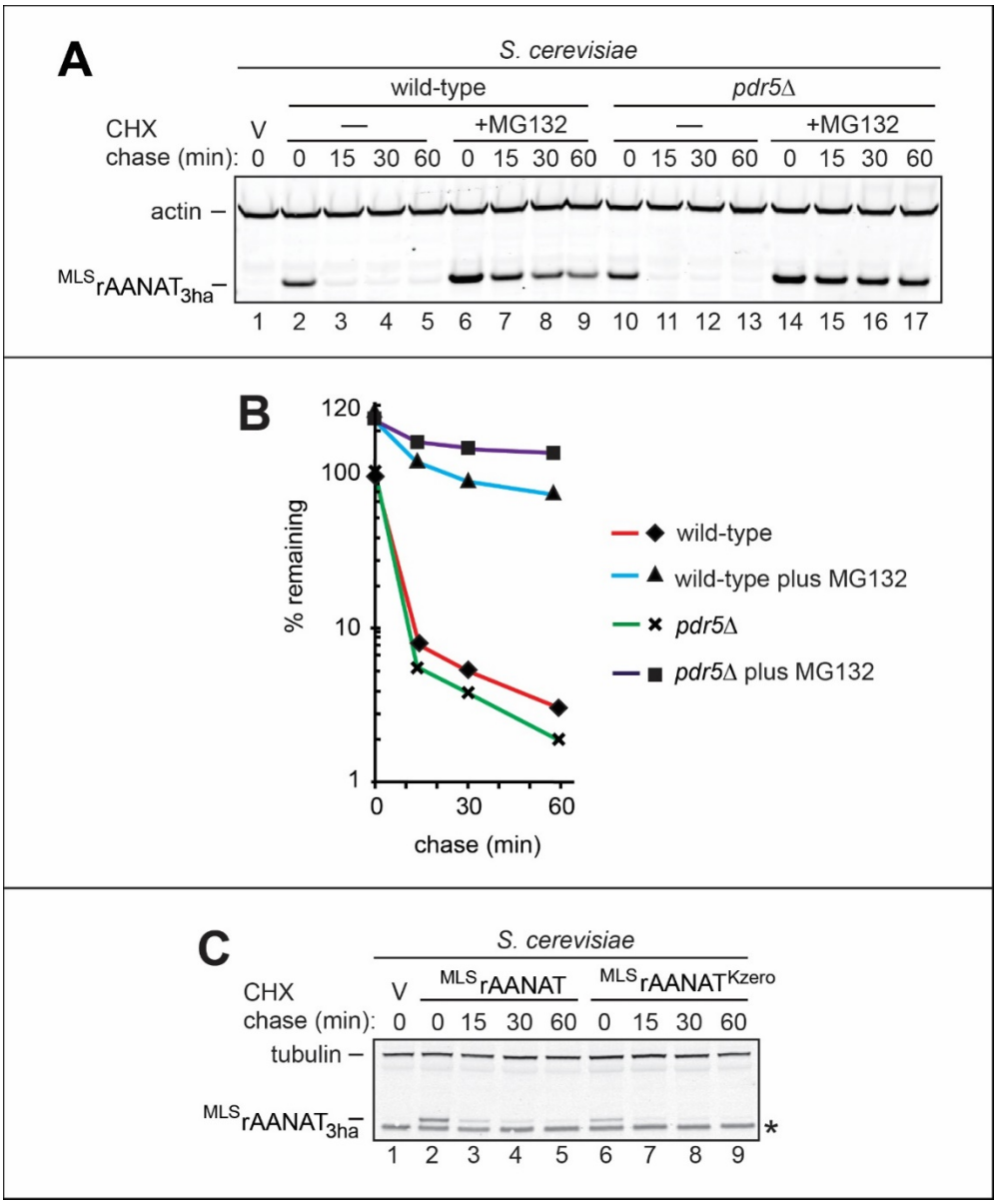


Fig A1.S1. Degradation assays with a proteasome inhibitor and a mutant of rat AANAT in *S. cerevisiae*. *A*, Lane 1, wild-type *S. cerevisiae* were transformed with vector (V) alone (control). Lanes 2-5, CHX-chase was performed at 30°C for the indicated times in wild-type *S. cerevisiae* with the wild-type rat ^{MLS}rAANAT$_{3ha}$. Lanes 6-9, same as in lanes 2-5 but the CHX-chase was in the presence of the MG132 proteasome inhibitor (see Experimental Procedures). Lanes 10-13, same as in lanes 2-5 but in *pdr5*" S. cerevisiae lacking the efflux pump Pdr5. Lanes 14-17, same as in lanes 10-13 but in the presence of MG132. The bands of actin and ^{MLS}rAANAT$_{3ha}$ are indicated on the left. *B*, quantification of data in *A*. C, CHX-chases were performed in wild-type *S. cerevisiae* with the wild-type rat ^{MLS}rAANAT$_{3ha}$ (lanes 2-5) and with its Lys-8-to-Arg mutant ^{MLS}rAANAAT$^{K8R}_{3ha}$ (see Results). The bands of tubulin and AANAT are indicated on the left. An asterisk on the right denotes a band of protein that crossreacted with anti-ha antibody (the band is also present in lane 1, the vector-only control).

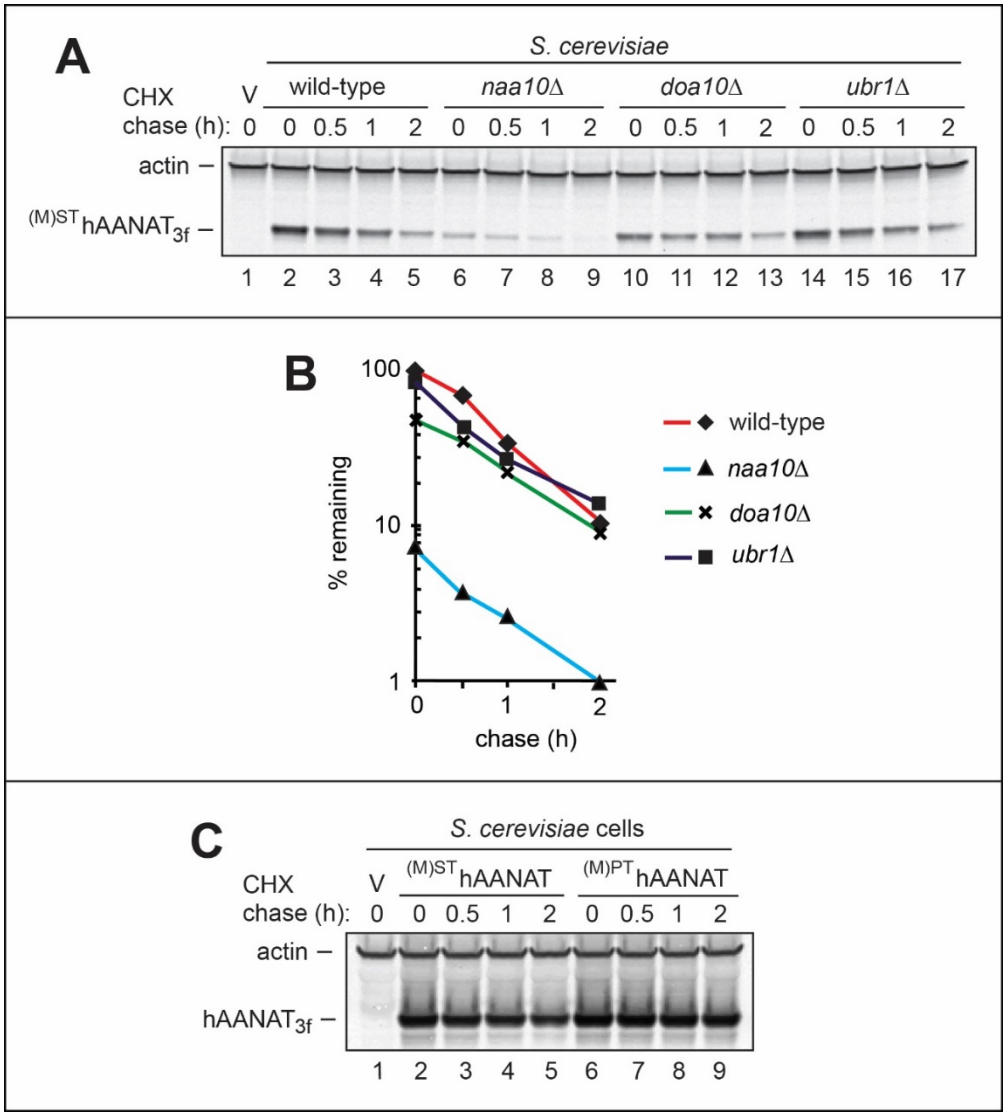


Fig. A1.S2. Degradation assays with wild-type and mutant human AANATs in *S. cerevisiae*. *A*, Lane 1, wild-type *S. cerevisiae* were transformed with vector (V) alone (control). Lanes 2-5, CHX-chase was performed at 30°C for the indicated times in wild-type *S. cerevisiae* with the wild-type human $^{(M)ST}$hAANAT$_{3f}$. Lanes 6-9, same as in lanes 2-5 but with the mutant human $^{(M)PT}$hAANAT$_{3f}$. The bands of actin and AANAT are indicated on the left. *B*, quantification of data in *A*.

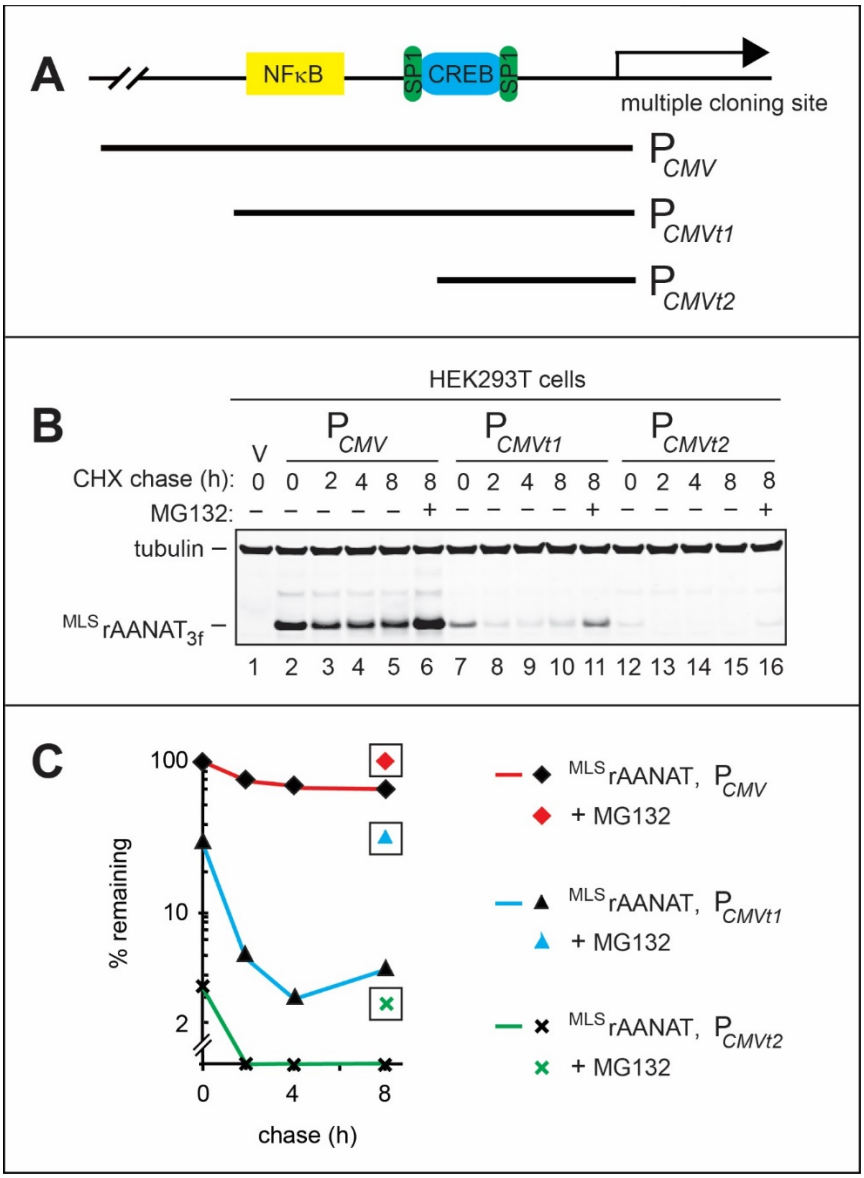


Fig. A1.S3. Construction and testing of weakened derivatives of the P_{CMV} promoter in HEK293T cells. *A*, a simplified diagram of "wild-type" (unmodified) P_{CMV} in the cloning vector, with schematically illustrated binding sites for the NF^{o} B, SP1 and CREB transcriptional regulators. Two 5'-terminal truncations of wild-type P_{CMV}, denoted as P_{CMVt1} and P_{CMVt2}, are also indicated. *B*, Lane 1, HEK293T cells were transformed with vector (V) alone (control). CHX-chases were performed at 37°C for the indicated times in HEK293T cells with the wild-type rat ^{MLS}rAANAT$_{3f}$ expressed either from the unmodified P_{CMV} promoter (lanes 2-5), or from the P_{CMVt1} promoter (lanes 7-10), or from the P_{CMVt2} promoter (lanes 12-15). Lanes 6, 10, and 16, same as lanes 5, 9, and 15, except that the MG132 proteasome inhibitor was present during a 8-h CHX-chase, in each case (see Experimental Procedures and Results). The bands of tubulin and ^{MLS}rAANAT$_{3f}$ are indicated on the left. *C*, quantification of data in *B*. Symbols shown separately and framed in black squares correspond to altered levels of test proteins in the presence of MG132.

Supplementary Table S1. *S. cerevisiae* strains used in this study.

Strains	Relevant genotypes	Sources
BY4742	*MAT± his3-1 leu2-0 lys2-0 ura3-0 can1-100*	Open Biosystems.
BY15470	*naa30"::kanMX6* in BY4742	Open Biosystems
BY10976	*naa10"::KanMX6* in BY4742	Open Biosystems
BY17299	*doa10":KanMX6 in BY4742*	Open Biosystems
CHY345	*ubr1"::LEU2* in BY4742	(Kim et al., 2014)
CHY346	*ubr1"::LEU2 doa10":KanMX6* in BY4742	This study
BWY29	*naa10"::kanMX6 ubr1"::LEU2* in BY4742	This study
CHY349	*naa30"::kanMX6 ubr1"::LEU2* in BY4742	(1)
JOY487	*naa30"::natNT2* in BY4742	This study

Supplementary Table S2. Plasmids used in this study.

Plasmid	Description	Source or Reference
pcDNA3		Invitrogen
pCISII-rAANAT	rAANAT in pCISII	(2)
pRS313Cup1	pRS313 containing the P_{CUP1} promoter	Varshavsky Lab Collection
pRS416Gal1	pRS416 containing the P_{GAL1} promoter	Varshavsky Lab Collection
pBW105	^{MLS}rAANAT$_{3f}$ in pRS416 with P*Gal1*	This study
pBW106	$^{(M)SI}$rAANAT$_{3f}$ in pRS416 with P*Gal1*	This study
pBW107	$^{(M)SMLS}$rAANAT$_{3f}$ in pRS416 with P*Gal1*	This study
pBW135	^{MLS}rAANAT$_{3ha}$ in pRS313 with P*Cup1*	This study
pBW173	pcDNA3(P_{CMVt1})	This study
pBW206	^{MLS}rAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW209	$^{(M)PLS}$rAANAT$_{3f}$ in pRS416 with P*Gal1*	This study
pBW211	$^{(M)SMLS}$rAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW212	$^{(M)SI}$rAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW213	$^{(M)PLS}$rAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW394	^{MLS}rAANAAT$^{K8R}_{3ha}$ in pRS416 with P*Gal1*	This study
pBW395	$^{(M)ST}$hAANAT$_{3f}$ in pRS416 with P_{Gal1}	This study
pBW419	$^{(M)PT}$hAANAT$_{3f}$ in pRS416 with P_{Gal1}	This study
pBW466	$^{(M)ST}$hAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW467	$^{(M)PT}$hAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW477	^{MLS}rAANAT$_{3ha}$ in pcDNA3(P_{CMVt1})	This study
pBW478	^{MLS}rAANAAT$^{Kzero}_{3ha}$ in pcDNA3(P_{CMVt1})	This study
pBW481	$^{(M)ST}$hAANAT$_{3f}$ in pcDNA3(P_{CMVt1})	This study
pBW482	$^{(M)ST}$hAANAAT$^{Kzero}_{3ha}$ in pcDNA3(P_{CMVt1})	This study

Supplementary Table S3. PCR primers used in this study.

Primer	Sequence
BW239	5'-CCTTGAATTCATACCCATGTTGAGCATCCA-3'
BW240	5'-CCTTGAATTCATACCCATGAGCATCCA-3'
BW241	5'-CCTTGAATTCATACCCATGAGCATGTTGAGCA-3'
BW243	5'-CCTTTCCTTGTAGTCGGATCCACCGCAGCCACTGTT-3'
BW244	5'-CCTTAAGCTTACTTGTCATCGTCGTCCTTGTAGTCGG-3'
BW335	5'-TTGAATTCATGTTGAGCATCCACCCC-3'
BW338	5'-GTATGGGTAGCAGCCACTGTTCC-3'
BW339	5'-CAGTGGCTGCTACCCATACGATG-3'
BW340	5'-TTCTCGAGTCAAGCGTAATCTGGAACGTC-3'
BW352	5'-AATTGAATTCACCATGTTGAGCATCC-3'
BW353	5'-AATTGAATTCACCATGAGCATCCACC-3'
BW354	5'-AATTGAATTCACCATGAGCATGTTGA-3'
BW355	5'-AATTCTCGAGTTACTTGTCATCGTCG-3'
BW375	5'-TTAATCTAGATGCATGCTCGAGC-3'
BW376	5'-CGTTAATCGCGACCAAAATCAACGGGACTTTC-3'
BW412	5'-TTGAATTCATGCCCTTGAGCATCCACCCC-3'
BW705	5'-TTGAATTCATACCCATGTCCACGCAGAGCACCCACCCCCTGAAACCTGA GGCCCCACGTCTGCCACCTGGGATC-3'
BW711	5'-CTTATAGTCACCAGAACCAGAGCAGCCGCTGTTCCTGCGCAGG-3'
BW712	5'-CAGGAACAGCGGCTGCTCTGGTTCTGGTGACTATAAGGATG-3'
BW771	5'-TTGAATTCATACCCATGCCAACGCAGAGCACCCACCCCCTGAAACCTG AGGCCCCACGTCTGCCACCTGGGATC-3'
BW784	5'-CTGAGTAAGTCTCTCTCTGTCCCAAAGTGAACC-3'
BW785	5'-GGTTCACTTTGGGACAGAGAGAGACTTACTCAG-3'
BW786	5'-CAGCAGGACGGAGCCTCTGCCCTGCTGCCGGAAG-3'
BW787	5'-CTTCCGGCAGCAGGGCAGAGGCTCCGTCCTGCTG-3'
BW788	5'-GGCCTGGAAACCAAATCTCTCATAGAAGGGCAC-3'
BW789	5'-GTGCCCTTCTATGAGAGATTTGGTTTCCAGGCC-3'
BW866	5'-TTGAATTCACCATGTCCACGCAGAGCACCCACCCCC-3'
BW867	5'-TTGAATTCACCATGTCCACGCAGAGCACCCACCCCCTGAGACCTGA GGC-3'
BW868	5'-TTGAATTCACCATGCCAACGCAGAGCACCCACCCC-3'
BW884	5'-TTAAGAATTCACCATGTCCACGCAGAGCACC-3'
BW885	5'-GTCAGGAACATCGTATGGGTAGCAGCCGCTGTTCCTGCGCAGG-3'

SUPPLEMENTAL REFERENCES

1. Kim, H. K., Kim, R. R., Oh, J. H., Cho, H., Varshavsky, A., and Hwang, C. S. (2014) The N-terminal methionine of cellular proteins as a degradation signal. *Cell* **156**, 158-169
2. Huang, Z., Deng, J., and Borjigin, J. (2005) A novel H28Y mutation in LEC rats leads to decreased NAT protein stability in vivo and in vitro. *J. Pineal Res.* **39**, 84-90

APPENDIX 2:

SUPPLEMENTAL MATERIAL FOR CHAPTER 3

MATERIALS AND METHODS

Reagents and antibodies

Cycloheximide, dithiobis(succinimidylpropionate) (DSP; Lomant's reagent), MG132 and complete protease inhibitor cocktail tablet were from Sigma, Thermo Scientific, Calbiochem and Roche, respectively. Antibodies to the following antigens were used for immunoblotting and/or immunoprecipitation: anti-flag M2 (Sigma, F1804 or Stratagen, 200471), anti-c-Myc-9E10 (Sigma, M5546), anti-hemaggutinin (ha) tag (Sigma, H9658 or H6958), anti-tubulin (Sigma, T5168), anti-Teb4 (Bethyl Laboratories, A304-171A), anti-ubiquitin (Stressgen, SPA-203), anti-Naa20 (Nat5) (Santa Cruz Biotechnology, sc-100645), anti-Naa60 (Nat15) (Abcam, ab103800), anti-phospho-p44/42 Erk1/2 (Cell Signaling Technologies, Ab9101), anti-p44/42 Erk1/2 (Cell Signaling Technologies, Ab9102) and anti-Rgs2 (Abgent, AT3628a). Secondary antibodies for immunoblotting were HRP-conjugated goat anti-rabbit (Bio-Rad, 170-6515) or anti-mouse (Bio-Rad, 170-6516), and either ECL Prime Western Blotting Detection System (GE Healthcare) or Clarity Western ECL substrate (Bio-Rad) were used, according to manufacturers' protocols. The following reagents were used for immunoprecipitation and GST-pulldown assays: anti-flag M2 affinity gel (Sigma, A2220), Dynabeads Protein A and G (Life Technologies), and Glutathione Hicap (Qiagen). For determining the state of $N^{\pm}$-terminal acetylation (Nt-acetylation) of Rgs2, human Rgs2 (purified from the HEK293 cell line) was purchased from Origene (TP302781) and analyzed using mass spectrometric (MS) techniques (see below).

Yeast strains, media, and genetic techniques

The *S. cerevisiae* strains used in this study are described in Table S1. Standard techniques (*27*, *28*) were employed for strain construction and transformation. *S. cerevisiae* media included YPD medium (1% yeast extract, 2% peptone, 2% glucose; only most relevant components are cited); SD medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose); and synthetic complete (SC) medium (0.17% yeast nitrogen base, 0.5% ammonium sulfate, 2% glucose), plus a drop-out mixture of compounds required by a given auxotrophic strain. The *S. cerevisiae* strain CHY367 (*naa20Δ::natNT2*) was constructed using PCR-mediated gene targeting (*29*) and the natNT2 module (*30*) in the strain JD53 (see Table S1).

Cell culture, transfection and RNA interference

Human HeLa cell line was grown in Dulbecco's Modified Eagle Medium (DMEM) (Ronza) supplemented with 10% fetal bovine serum (FBS) (HyClone) and penicillin (100 U/ml)/streptomycin (100 μg/ml) (HyClone) in a 5% CO_2 incubator at 37°C. Transfections with DNA plasmids or siRNA oligoribonucleotides were carried out using Lipofectamine-2000 (Life Technologies) according to the manufacturer's instructions. For RNA interference (RNAi), HeLa cells were seeded on day 1 at 2×10^5 cells per well in 6-well cell culture plates (SPL Life Sciences, Korea) containing DMEM and 10% FCS. After incubations for 24 hr, cells were transfected with 100 nM of either human *Teb4* siRNA, or *Naa20* siRNA, or *Naa60* siRNA, or non-targeting (control) siRNA (Santa Cruz Biotechnology, sc-91789, sc-37007, sc-62662, and sc-93037, respectively), using Lipofectamin-2000 according to the manufacturer's instructions.

DNA constructs and primers

E. coli DH5± strain was used for cloning and maintaining plasmids (Table S1). Solg Pfu-X DNA polymerase (SolGent, Korea) was employed for PCR. The plasmids and PCR primers used in this study are described in Tables S2 and S3, respectively. To express human **MX**-Rgs2 (**X**=Gln, Leu, Arg) in mammalian cells, we used the pCH766, pCH767 and pCH768 plasmids, which expressed, respectively, **MQ**-Rgs2$_{ha}$, **ML**-Rgs2$_{ha}$ and **MR**-Rgs2$_{ha}$ from the P_{CMV} promoter and were a gift from Dr. R. Neubig (Univ. of Michigan Medical School, East Lansing, MI) (*11*). To construct pCH821, which expressed **PQ**-Rgs2$_{ha}$ (**MPQ**-Rgs2$_{ha}$), the **MPQ**-RGS2$_{ha}$-encoding DNA fragment was produced by PCR from the plasmid pCH766 (Table S2) using the primer pair OCH1091/OCH1100. This fragment was digested with *Kpn*I/*Xho*I and ligated into *Kpn*I/*Xho*I-cut pCH766.

To construct the low copy (*CEN*-based) pCH5050 and pCH5051 plasmids, which expressed **MQ**-Rgs2$_{ha}$ and **ML**-Rgs2$_{ha}$ in *S. cerevisiae* from the P_{CUP1} promoter, the relevant *Rgs2* DNA fragments were amplified from pCH766 using the primer pairs OCH5070/OCH5073 and OCH5071/OCH5073, respectively. The resulting PCR products were digested with *Bam*HI/*Sac*I and ligated into *Bam*HI/*Sac*I-cut pCH692, a pRS316-based low copy plasmid containing the P_{CUP1} promoter.

To construct pCH879 and pCH881, which expressed, from the P_{CMV} promoter in mammalian cells, the wild-type human Teb4 and its catalytically-inactive mutant Teb4^{C9A} (both of them were C-terminally tagged with triple flag plus His_6), a DNA sequence encoding the above “double” epitope was produced by annealing the complementary primers OCH1265/OCH1266 (Table S3). The resulting DNA, encoding triple flag plus

His_6, was digested with *Hin*dIII/*Afl*II and ligated into *Hin*dIII/*Afl*II-cut pCH759 and pCH760, which encoded, respectively Teb4 and $Teb4^{C9A}$ in the pcDNA3.1-myc-His(-) A vector (*24*). The plasmid pCH3056, which encoded full-length **MQ**-Rgs2 whose C-terminus was fused to GST, was constructed by producing, at first, a DNA fragment from pCH766 (Table S2), using PCR and the primer pair OCH3006/OCH3013. This fragment was digested with *Nde*I/*Bam*HI. Furthermore, a DNA fragment was produced by PCR from pGEX4T-3 and the primer pair OCH3015/OCH1514, followed by digestion of the resulting fragment with *Bam*HI/*Xho*I. The above two fragments were ligated into *Nde*I/*Xho*I-cut pET30a, yielding the plasmid pCH3056.

To construct the plasmid pCH3080, which expressed, from the IPTG-inducible *T7lac* promoter in *E. coli*, a fusion between the N-terminal 10 residues of **MQ**-Rgs2 and the GST moiety, a DNA fragment was produced by PCR using pCH3056 (Table S2) and the primer pair OCH3095/OCH1514 (Table S3). The resulting fragment was digested with *Nde*I/*Xho*I and ligated into *Nde*I/*Xho*I-cut pET30a (Table S2).

To construct pCH3132 that expressed the human muscarinic acetylcholine receptor M3 from the P_{CMV} promoter, DNA sequence encoding M3 was PCR-amplified from a human HEK293T cDNA library using primer pairs OCH3178/OCH3179. The resulting PCR fragment was digested with *Bam*HI/*Xba*I and ligated into *Bam*HI/*Xba*I-cut pCH61. To clone pCH5064 that expressed the human guanidine-nucleotide binding protein G(q) subunit alpha (G±q) from the P_{CMV} promoter, DNA sequence encoding G±q was also PCR-amplified from HEK293T cDNA library using primer pairs OCH5090/OCH5091. The resulting DNA fragment was digested with *Kpn*I/*Xho*I and ligated into *Kpn*I/*Xho*I-cut pCH61, yielding pCH5064. Construction details for other

plasmids are available upon request. All final constructs were verified by DNA sequencing.

Cycloheximide-chase assays and immunoblotting

For immunoblotting, HeLa cells were lysed by incubating them on ice for 20 min in the TLB lysis buffer (1% Triton X100, 50 mM NaCl, 2 mM Na-EDTA, 50 mM Tris-HCl, pH 7.4), containing complete protease inhibitor mixture (Roche). After precipitating cell debris by centrifugation at 11,200*g* for 10 min at 4°C, total protein concentrations in the supernatants were determined using Bradford or DC protein assay (Bio-Rad). Identical total amounts of protein in each sample were heated at 95°C for 10 min or at 37°C for 15 min (the samples for immunoblotting of Teb4 were incubated, instead, at 37°C for 20 min) in 1x SDS-sample buffer, followed by Tris-glycine-SDS-10% PAGE or Tris-glycine-4-20% PAGE. For immunoblotting analyses of Teb4, protein samples were fractionated by Tris-glycine-SDS-6% PAGE or by Tris-glycine-SDS-8% PAGE. Fractionated proteins were electroblotted to Immobilon-P membranes (Millipore) using mini-transblot cell (Bio-Rad), and were analyzed by immunoblotting with indicated antibodies, followed by secondary antibodies described in the first section above. Immunoblots were detected using Clarity Western ECL substrate (Bio-Rad) or ECL Plus Western blotting Detection System (GE Healthcare).

Cychloheximide (CHX)-chase assays with HeLa cells were set up on day 1 at 2×10^5 cells/well in 6- or 12-well cell culture plates (SPL Life Sciences, Korea) containing DMEM and 10% FCS. After 24-hr incubation at 37°C, the cells were treated with CHX at the final concentration of 0.1 mg/ml, harvested by centrifugation at each time point, and lysed in RIPA buffer (0.1% NP40, 0.15 M NaCl, 5 mM Na-EDTA, 1 mM

dithiothreitol (DTT), 50 mM HEPES, pH 7.5, plus the *"complete protease inhibitor cocktail"* (Roche)). Equal amounts of total protein were fractionated by Tris-glycine SDS-4-20 % PAGE, followed by immunoblotting with indicated antibodies. In particular, immunoblotting with antibody to α-tubulin was used as a loading control for all immunoblotting experiments.

CHX-chases with *S. cerevisiae* cells were performed as previously described (*13, 15*), with slight modifications. *S. cerevisiae* were grown to A_{600} of 0.8-1.0 at 30°C in selective media appropriate for a plasmid(s) carried by a strain. Cells were harvested by centrifugation at 11,200*g* for 1 min and thereafter resuspended in fresh YPD to a final concentration of A_{600} of 1.0, followed by the treatment with CHX at the final concentration of 0.2 mg/ml. At indicated times, cell samples (corresponding to 1 ml of cell suspension at A_{600} of 1) were harvested by centrifugation for 1 min at 11,200*g*, and were resuspended in 0.8 ml of 0.2 M NaOH and incubated for 20 min on ice, followed by centrifugation for 1 min at 11,200*g*. Pelleted cells were resuspended in 80 μl of HU buffer (8 M urea, 5% SDS, 1 mM Na-EDTA, 0.1 M DTT, 0.005% bromophenol blue, 0.2 M Tris-HCl, pH 6.8) containing 1x protease inhibitor cocktail *"for use with fungal and yeast extracts"* (Sigma), and heated for 10 min at 70°C. After centrifugation for 5 min at 11,200*g*, 10 μl of supernatant was subjected to SDS-10% PAGE, followed by immunoblotting with anti-ha (1:2,000) and anti-tubulin (1:2,000) antibodies. Quantification of immunoblotting data was carried out using GelQuantNET (http://biochemlabsolutions.com/GelQuantNET.html).

Cycloheximide-chase assays with co-expressed $Rgs2_{ha}$ and $G\pm q_{ha2}$

of a chase by rapid scraping into 0.1 ml of TSD buffer (1% SDS, 5mM DTT, 50 mM Tris-HCl, pH7.4) and snap-freezing in liquid nitrogen. Samples were then heated at 95°C for 10 min and diluted with 10 volumes of TNN buffer (0.5% NP40, 0.25M Na Cl, 5 mM Na-EDTA, 50 mM Tris-HCl, pH 7.4) containing the *"complete protease-inhibitor mixture"* (Roche). The amounts of ^{35}S in the 10% CCl_3COOH-insoluble fraction of each sample were then measured by liquid scintillation counting. For immunoprecipitation, the samples were adjusted to contain equal amounts of total ^{35}S and then added to 20 µl of anti-ha agarose (Sigma). Suspensions of beads were incubated by rocking at 4°C for 2 hr, washed 3 times in TNN buffer, and once in 10 mM Tris-HCl, pH 8.5. Final samples were eluted by 30 µl of SDS-PAGE sample buffer, including the heating at 95°C for 5 min, followed by clarification through centrifugation at 11,200*g* for 1 min. 10 µl of each of the resulting samples were separated by SDS-4-20% PAGE (Bio-Rad, followed by autoradiography using Typhoon 9400 PhosphorImager (GE Healthcare) and quantification of autoradiograms with ImageQuant (GE Healthcare).

RNA isolation and quantitative RT-qPCR

Total RNA was isolated from HeLa cells using an RNeasy Mini kit (Qiagen), according to the manufacturer's instructions. 2 µg of total RNA was reverse-transcribed into cDNA with oligo-dT as a primer by using Promega cDNA reverse transcription kit. Quantitative real-time PCR was performed in 20-µl reactions containing 10 ng of the cDNA sample, 50 nM forward and reverse primers, and the SYBR Green Master Mix (Applied Biosystems). Primers for human *Rgs2* and ² -actin genes were designed using GenScript Real-time PCR Primer Design tool. Reaction mixtures were incubated at 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. No template

control sample was included, to enable detection of RNA and/or DNA contamination. Quantification of relative *Rgs2* mRNA expression levels was carried out using the threshold-cycle difference method (ref. (*31*) and *Applied Biosystems StepOne and StepOnePlus Real-Time PCR Systems Getting Started Guide for Relative Standard Curve and Comparative CT experiments*), with ² -actin as a normalization control. The *Rgs2* and β-actin gene primer pairs for RT-qPCR were OCH5082/OCH5083 and OCH5084/OCH5085, respectively (Table S3).

In vivo ubiquitylation assay

HeLa cells were transfected with pcDNA3.1 (control vector), or CH776 (a pcDNA3.1-based plasmid expressing **MQ**-$Rgs2_{ha}$), or pCH879 (expressing $Teb4_{3f}$) or pCH881 (expressing $Teb4_{3f}^{C9A}$) (Table S2). After 24 hr, transfected cells were treated with 10•µM MG132 for 4•hr and thereafter lysed with RIPA buffer. Supernatants were immunoprecipitated with a monoclonal anti-ha antibody (Sigma), and were washed three times in RIPA buffer. Immunoprecipitates were fractionated by SDS-8% PAGE, followed by immunoblotting with anti-ubiquitin and/or rabbit polyclonal anti-ha antibodies.

UQDl0

plus *"complete protease inhibitor cocktail"* (Roche)) and incubating on ice for 20 min. The extracts were centrifuged at 11,200*g* for 20 min at 4°C and total protein concentrations in the supernatants were determined using Bradford assay (Bio-Rad). Samples containing 1 mg of total protein were incubated with anti-ha or anti-flag bound IgG Dynabeads (Life Technologies) at 4°C for 16 hr and the magnetic beads were washed 3 times in buffer A. Captured proteins were eluted in 15 μl of 0.1 M glycine-HCl (pH 2.8), followed by incubation in SDS-sample buffer containing 50 mM DTT for 30 min at 37°C, SDS-10% PAGE and immunoblotting with anti-ha or anti-flag as described above.

Purification of MQ-Rgs2$^{3-10}$-[GST] and GST-pulldown assays

To produce and purify the conditionally N$^{\pm}$-terminally acetylated (Nt-acetylated) **MQ**-Rgs2$^{3-10}$-[GST] fusion, we employed an *E.coli*-based expression system. "Wild-type" *E. coli* very rarely Nt-acetylates heterologous eukaryotic proteins. In the resulting novel system, **MQ**-Rgs2$^{3-10}$-[GST] was expressed in *E. coli* either by itself or together with the cognate (Nt-acetylating the Met-Gln N-terminal sequence) NatB Nt-acetylase complex from the fission yeast *Schizosaccharomyces pombe* (*32*). The plasmid pCH3080, expressing **MQ**-Rgs2$^{3-10}$-[GST], was transformed into *E. coli* BL21(DE3) that carried either a vector plasmid (pACYCDuet (pCH15)) or pCH3025 (*32*), which expressed the *Naa20* and *Naa25* subunits of the *S. pombe* NatB Nt-acetylase in *E. coli*.

Overnight cultures (5 ml) of transformed *E. coli* were inoculated into LB medium (500 ml) containing chloramphenicol (34 μg/ml) and kanamycin (50 μg/ml), followed by growth at 37°C to A_{600} of ~0.7. Expression of **MQ**-Rgs2$^{3-10}$-[GST] was then induced with 1 mM isopropyl β-D-thiogalactoside (IPTG) at 30°C for 4 hr. Cells were harvested by

centrifugation and frozen at -80°C. Cell pellets were thawed and resuspended in STE buffer (0.1 M NaCl, 1 mM Na-EDTA, 10 mM Tris-HCl, pH 8.0) containing 1 mM DTT, 1mM phenylmethylsulfonyl fluoride (PMSF) and 1 mg/ml of chicken egg white lysozyme (Sigma). Cell suspensions were incubated on ice for 20 min and thereafter treated by sonication for 1 min 3 times, at 1-min intervals, followed by the addition of Triton X100 to the final concentration of 1%. After centrifugation at 11,200g for 20 min at 4°C, the supernatants (~25 ml) were incubated with 1 ml of Glutathione Hicap agarose (50% slurry; Qiagen) at 4°C for 2 hr. The beads were washed twice in 25 ml of the STE buffer. Carrier-bound GST or **MQ**-Rgs2$^{3-10}$-[GST] were eluted with 1 ml of STE buffer containing 10 mM glutathione (GSH), followed by overnight dialysis against storage buffer (10% glycerol, 0.15 M NaCl, 10 mM β-mercaptoethanol, 50 mM HEPES, pH 7.5). The state of Nt-acetylation of the resulting **MQ**-Rgs2$^{3-10}$-[GST] fusion proteins was determined by MALD-TOF mass spectrometry.

For GST-pulldown assays, 50 μg of either GST, or **MQ**-Rgs2$^{3-10}$-[GST] (expressed in *E. coli* without *S. pombe* NatB), or (presumably) Nt-acetylated **MQ**-Rgs2$^{3-10}$-[GST] (coexpressed in *E. coli* with *S. pombe* NatB) were incubated with 50 μl of the Glutathione Hicap agarose beads in GST-loading buffer (10% glycerol, 0.5 M NaCl, 1% NP40, 1 mM Na-EDTA, 50 mM Tris-HCl, pH 8.0) at 4°C for 1 hr. The beads were washed once with 0.5 ml of the loading buffer and once with 0.5 ml GST-binding buffer (10% glycerol, 0.05% NP40, 50 mM NaCl, 50 mM Na-HEPES, pH 7.8). The beads in 0.2 ml-GST binding buffer were thereafter incubated at 4°C for 2 hr with 0.25 ml extracts from HeLa cell (2 mg of total protein) that expressed C-terminally triple flag-tagged (plus His_6-tagged) Teb4. The beads were washed 3 times in 0.5 ml of GST-binding buffer,

followed by elution of bound proteins by the addition of 15 μl of 2xSDS-sample buffer and incubation for 30 min at 37°C. The resulting samples were fractionated by SDS-8% PAGE and analyzed by immunoblotting with anti-flag antibody. The blotting membrane was also stained with Coomassie Brilliant Blue to verify the approximate equality of total protein loads vis-à-vis GST, **MQ**-Rgs2$^{3-10}$-[GST] and Ac-**MQ**-Rgs2$^{3-10}$-[GST].

Analysis of endogenous MQ-Rgs2 from human cells by mass spectrometry

Purified human Rgs2 (~ 0.4 μg; see the first section of Materials and Methods) was treated with 8 M Urea in 25 mM NH_4HCO_3 (pH 8.0) and thereafter reduced with 9.5 mM DTT in 25 mM NH_4HCO_3 (pH 8.0) at 25°C for 1 hr. Rgs2 was then alkylated with iodoacetamide (32 mM) in NH_4HCO_3 (pH 8.0) at 25°C for 1 hr in the dark, followed by quenching of excess unreacted iodoacetamide with DTT (40 mM) at 25°C for 10 min. The sample was diluted 10-fold with 25 mM NH_4HCO_3 (pH 8.0), and 1 ml of the resulting sample was digested with of trypsin (1.2 μg/ml) overnight at 37°C. Nt-acetylated peptides in the digested sample were analyzed by nanoflow liquid chromatography-tandem mass spectrometry (LC-MS/MS) using the LTQ Orbitrap Velos (Thermo Scientific) hybrid mass spectrometer. Nt-acetylation sites were assigned by manual inspection of MS/MS spectra and also using MaxQuant software (*33*).

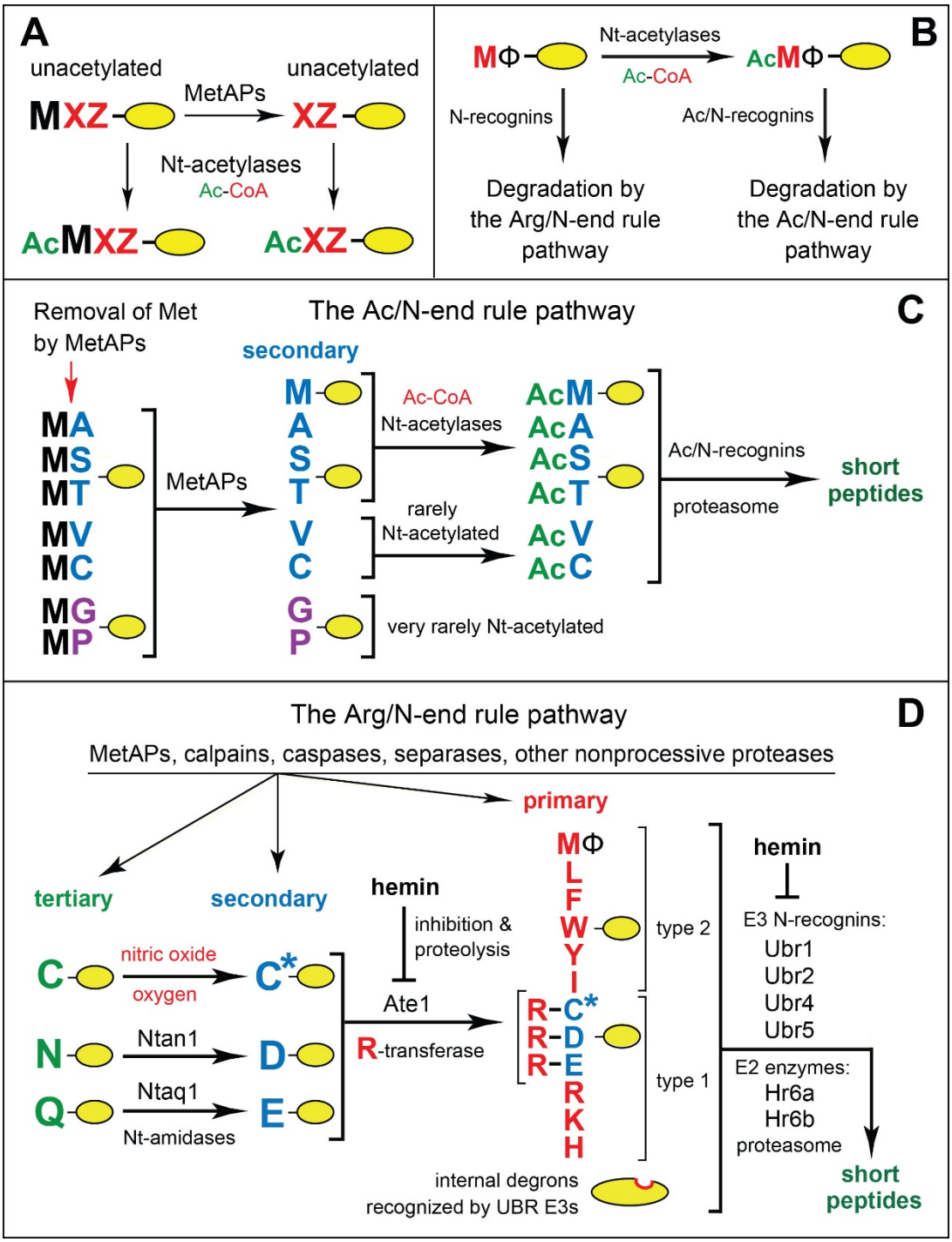


Fig. A2.S1. Conditional removal of N-terminal methionine from nascent proteins, their N-terminal acetylation, and the N-end rule pathway.

(**A**) N-terminal processing of nascent proteins by $N^{\pm}$-terminal acetylases (Nt-acetylases) and Met-aminopeptidases (MetAPs). "**Ac**" denotes the $N^{\pm}$-terminal acetyl moiety. **M**, methionine (Met) residue. **X** and **Z**, single-letter abbreviations for any amino acid residue. Yellow ovals denote the rest of a protein molecule. MetAPs

cotranslationally remove N-terminal Met from a nascent protein unless the side chain of a residue at position 2 is bulkier than Val, in which case Met is retained (*12, 34*).

(**B**) Functional complementarity between the Arg/N-end rule pathway and the Ac/N-end rule pathway for proteins bearing the N-terminal **M$** motif, i.e., unacetylated N-terminal Met followed by a bulky hydrophobic ($) residue. This motif is targeted by the Arg/N-end rule pathway (*16*). Because the Met-$ motif is also a substrate of the NatC Nt-acetylase (Fig. A3.S2), Met-$ proteins are usually Nt-acetylated at least in part, thereby acquiring the previously characterized Ac/N-degrons (*13, 16*). Thus, *both* Nt-acetylated and unacetylated versions of the same Met-$ protein can be destroyed, either by the Ac/N-end rule pathway or by the "complementary" Arg/N-end rule pathway (*16*).

(C, D) The mammalian N-end rule pathway. The main determinant of an N-degron is a destabilizing N-terminal residue of a protein. Recognition components of the N-end rule pathway, called N-recognins, are E3 ubiquitin ligases that can recognize N-degrons.

Regulated degradation of proteins or their fragments by the N-end rule pathway mediates a strikingly broad range of functions, including the sensing of heme, nitric oxide, oxygen, and short peptides; control of protein quality and subunit stoichiometries, including the elimination of misfolded proteins; regulation of signaling by G proteins; repression of neurodegeneration; regulation of apoptosis, chromosome cohesion/segregation, transcription, and DNA repair; control of peptide import; regulation of meiosis, autophagy, immunity, fat metabolism, cell migration, actin filaments, cardiovascular development, spermatogenesis, and neurogenesis; the functioning of adult organs, including the brain, muscle and pancreas; and the regulation of many processes in plants (references (*12-14, 16, 18, 20, 35-88*) and references therein). In eukaryotes, the N-end rule pathway consists of two branches, the Ac/N-end rule pathway and the Arg/N-end rule pathway.

(**C**) The Ac/N-end rule pathway. This diagram illustrates the mammalian Ac/N-end rule pathway through extrapolation from its *S. cerevisiae* version. As described in the main text, the Ac/N-end rule pathway has been identified only in *S. cerevisiae* (*13, 15, 16*), i.e., the presence of this pathway in mammals and other multicellular eukaryotes was conjectural, until the present work. Red arrow on the left indicates the removal of N-terminal Met by Met-aminopeptidases (MetAPs). N-terminal Met is retained if a residue at position 2 is nonpermissive (too large) for MetAPs. If the retained N-terminal Met or N-terminal Ala, Ser, Thr are followed by acetylation-permissive residues, the above N-terminal residues are $N^{\pm}$-terminally acetylated (Nt-acetylated) by ribosome-associated Nt-acetylases (*21, 89, 90*). Nt-terminal Val and Cys are Nt-acetylated relatively rarely, while N-terminal Pro and Gly are almost never Nt-acetylated. (N-terminal Gly is often N-myristoylated.) N-degrons and N-recognins of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively (*12*). The term "secondary" refers to Nt-acetylation of a destabilizing N-terminal residue before a protein can be recognized by a cognate Ac/N-recognin. As described in the main text, the human Teb4 E3 ubiquitin ligase was identified, in the present study, as an Ac/N-recognin of the mammalian Ac/N-end rule pathway. Natural Ac/N-degrons are regulated at least in part by their reversible steric shielding in protein complexes (*15, 16*).

(**D**) The Arg/N-end rule pathway. The prefix "Arg" in the pathway's name refers to N-terminal arginylation (Nt-arginylation) of N-end rule substrates by the Ate1

arginyltransferase (R-transferase), a significant feature of this proteolytic system. The Arg/N-end rule pathway targets specific unacetylated N-terminal residues. In the yeast *S. cerevisiae*, the Arg/N-end rule pathway is mediated by the Ubr1 N-recognin, a 225 kDa RING-type E3 ubiquitin ligase and a part of the targeting apparatus comprising a complex of the Ubr1-Rad6 and Ufd4-Ubc4/5 holoenzymes (*12, 14, 18*). In multicellular eukaryotes several functionally overlapping E3 ubiquitin ligases (Ubr1, Ubr2, Ubr4, Ubr5) function as N-recognins of this pathway. An N-recognin binds to the "primary" destabilizing N-terminal residues Arg, Lys, His, Leu, Phe, Tyr, Trp and Ile. In contrast, the N-terminal Asn, Gln, Asp, and Glu residues (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications. These modifications include the Nt-deamidation of N-terminal Asn and Gln by the Ntan1 and Ntaq1 Nt-amidases, respectively, and the Nt-arginylation of N-terminal Asp and Glu by the Ate1 R-transferase, which can also Nt-arginylate oxidized Cys, either Cys-sulfinate or Cys-sulfonate. They can form in animal and plant cells through oxidation of Cys by nitric oxide (NO) and oxygen, and also by N-terminal Cys-oxidases (*19, 44, 52, 59, 61, 91-93*). In addition to its type-1 and type-2 binding sites that recognize, respectively, the basic and bulky hydrophobic unacetylated N-terminal residues, an N-recognin such as Ubr1 contains other substrate-binding sites as well. These sites recognize substrates that are targeted through their internal (non-N-terminal) degrons, as indicated on the diagram (*12*). Hemin (Fe^{3+}-heme) binds to the Ate1 R-transferase, inhibits its Nt-arginylation activity and accelerates its in vivo degradation. Hemin also binds to Ubr1 (and apparently to other N-recognins as well) and alters its functional properties, in ways that remain to be understood (*43*). As shown in the diagram, the unacetylated Met of the N-terminal **M$** motif is recognized by both yeast and mammalian Ubr1; this capability greatly expands the substrate range of the Arg/N-end rule pathway (*16*).

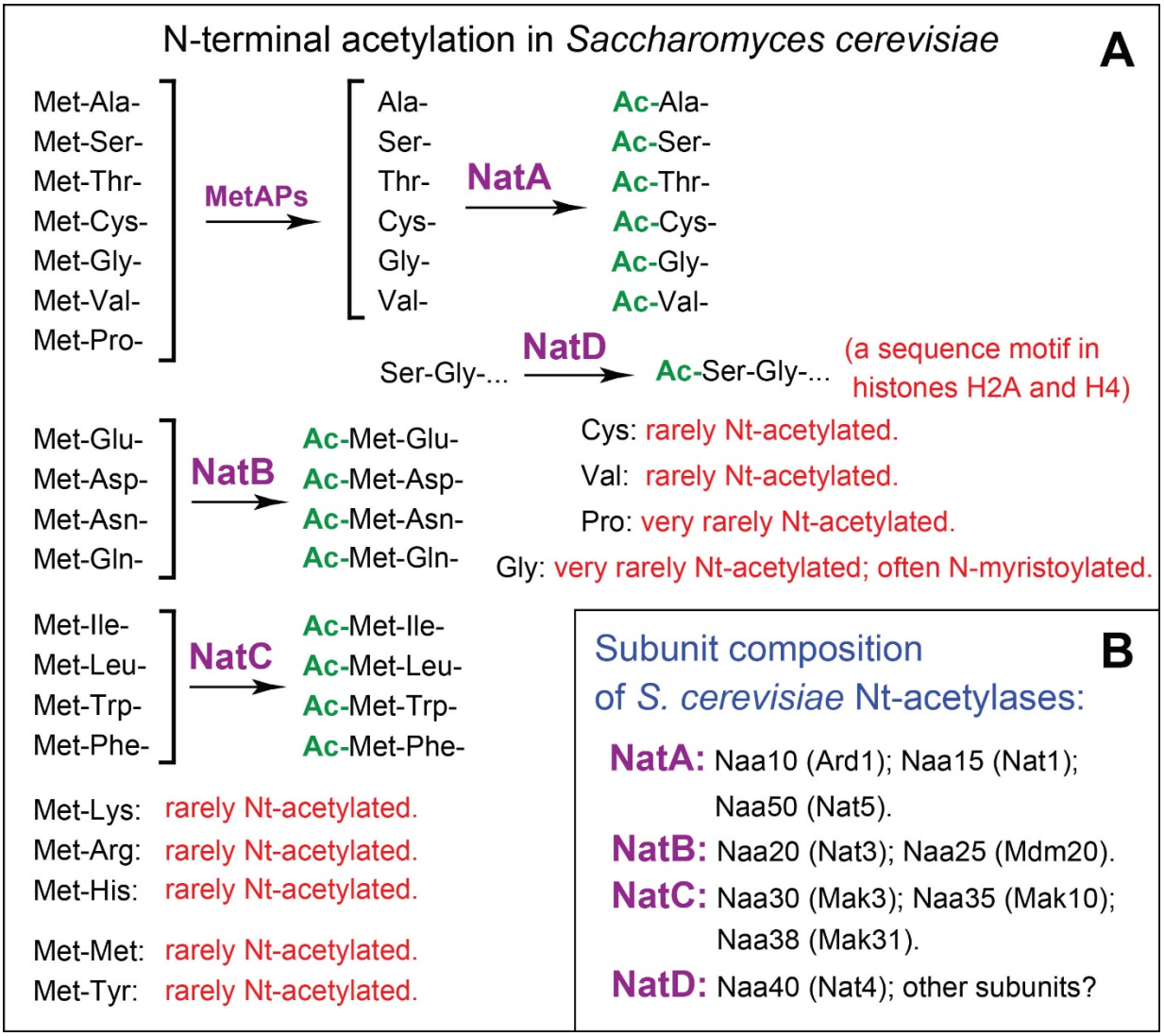


Fig. A2.S2. Specificities and subunit compositions of N$^{\pm}$-terminal acetylases (Nt-acetylases).

(**A**) Substrate specificities of *S. cerevisiae* Nt-acetylases. "**Ac**" denotes the N$^{\pm}$-terminal acetyl moiety. The bulk of Nt-acetylases are associated with ribosomes (*94, 95*). The specificities of mammalian Nt-acetylases are similar to those of their yeast counterparts, but an individual mammalian genome encodes more than ten Nt-acetylases, in contrast to four in *S. cerevisiae*. Some mammalian Nt-acetylases, such as NatF (its catalytic subunit is called Naa60 (see fig. A3.S4H, I)) (*22*), can Nt-acetylate N-terminal motifs that include Met-Lys or Met-Arg, which are rarely if ever Nt-acetylated in *S. cerevisiae*. This compilation of Nt-acetylases and their specificities is derived from data in the literature ((*21, 89, 90, 96-108*) and references therein).

(**B**) Subunits of *S. cerevisiae* Nt-acetylases. This paper employs the revised nomenclature for subunits of Nt-acetylases (*109*) and cites older names of these subunits in parentheses.

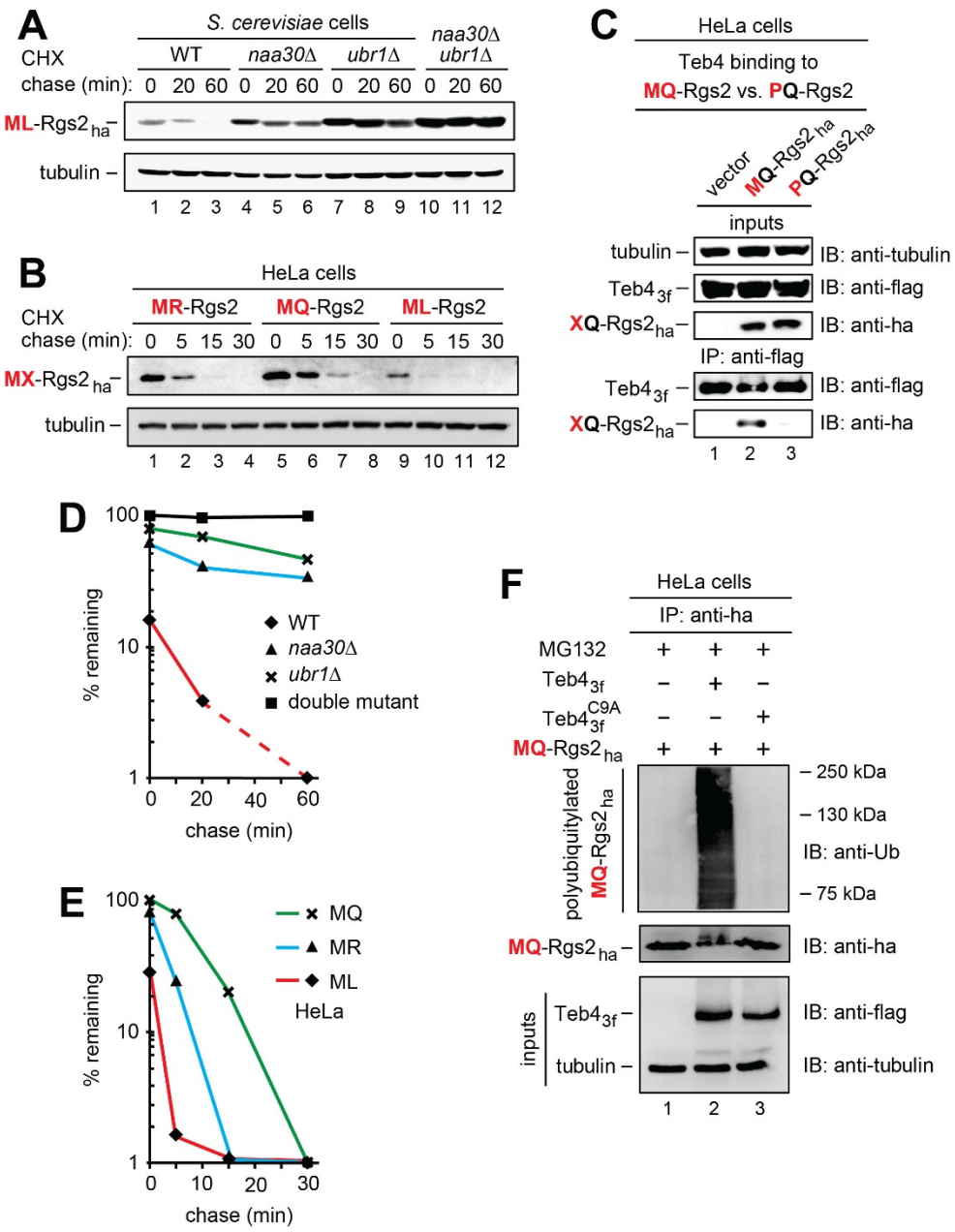


Fig. A2.S3. The targeting of MX-Rgs2 proteins by the yeast and human Ac/N-end rule and Arg/N-end rule pathways.

(**A**) CHX chases with the hypertension-associated mutant human **ML**-Rgs2$_{ha}$ in wild-type, *naa30"*, *ubr1"*, and *naa30" ubr1" S. cerevisiae*.

(**B**) As in **A** but in human HeLa cells with wild-type human **MQ**-Rgs2$_{ha}$ and hypertension-associated mutants **MR**-Rgs2$_{ha}$ and **ML**-Rgs2$_{ha}$.

(**C**) HeLa cells expressing **MQ**-Rgs2$_{ha}$ or **PQ**-Rgs2$_{ha}$ together with Teb4$_{3f}$ were treated with a cell-penetrating crosslinker, followed by immunoprecipitations with anti-flag, reversal of crosslinks, SDS-PAGE, and immunoblotting with anti-ha and anti-

flag (see also Fig. 2H and Materials and Methods). Note the coimmunoprecipitation of **MQ**-Rgs2$_{ha}$ with Teb4$_{3f}$, and the absence of coimmunoprecipitation of **PQ**-Rgs2$_{ha}$.

(**D**) Quantification of data in **A**. 100% corresponded to the zero-time level of **ML**-Rgs2$_{ha}$ in *naa30" ubr1" S. cerevisiae*.

(**E**) Quantification of data in **B**. 100% corresponded to the zero time level of **PQ**-Rgs2$_{ha}$ in HeLa cells.

In every set of degradation curves shown in this figure and in the main Figs. 1 and 3, 100% at zero time was chosen to correspond to the relative initial level of the most slowly degraded test protein in a given set. Panel **A** (quantified in **D**) and panel **B** (quantified in **E**) show the results of CHX-chases that were independently carried out counterparts of CHX-chases that are shown in the main Fig. 1**B** (quantified in **D**) and 1**F** (quantified in **E**), respectively. These comparisons illustrate qualitative reproducibility of independently generated data (particularly the robustly reproduced order of metabolic stabilities of specific test proteins in specific genetic backgrounds), which differed in relatively minor aspects of corresponding degradation curves. In our extensive experience with repeated CHX-chases and ^{35}S-pulse-chases in either *S. cerevisiae* or HeLa cells, a strictly quantitative reproducibility of independently produced and measured chase-degradation patterns remains the aim to be reached.

(**F**) Extracts from HeLa cells expressing **MQ**-Rgs2$_{ha}$ in the presence or absence of coexpressed Teb4$_{3f}$ or Teb4$_{3f}^{C9A}$ and in the presence of the MG132 proteasome inhibitor, as indicated. **MQ**-Rgs2$_{ha}$ was immunoprecipitated from extracts using anti-ha antibody, followed by SDS-PAGE and immunoblotting with anti-ubiquitin antibody to detect polyubiquitin chains linked to **MQ**-Rgs2, and with anti-flag, anti-ha and anti-tubulin antibodies as well, as shown.

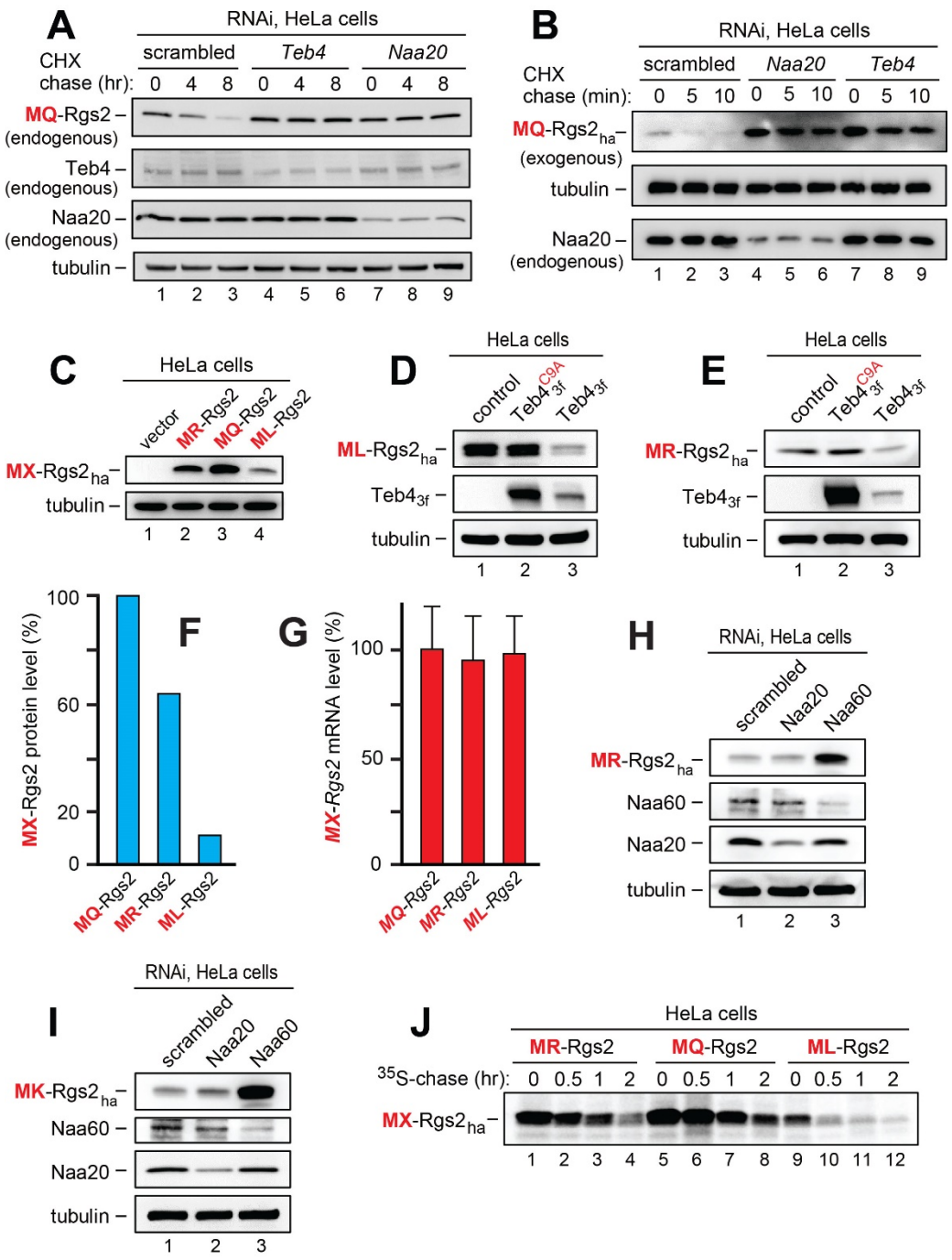


Fig. A2.S4. Degradation assays with MX-Rgs2 proteins in HeLa cells subjected to RNAi of specific Nt-acetylases and Teb4.

(**A**) CHX-chases with endogenous (not overexpressed) **MQ**-Rgs2 in HeLa cells subjected to RNAi-mediated knockdowns of endogenous Teb4 or Naa20 (NatB).

(**B**) Same as in **A** but CHX-chases with transiently expressed (exogenous) **MQ**-Rgs2$_{ha}$.

(**C**) Relative steady-state levels of exogenous wild-type **MQ**-Rgs2$_{ha}$ and its **MR**-Rgs2$_{ha}$ and **ML**-Rgs2$_{ha}$ mutants in HeLa cells transfected with equal amounts of **MX**-Rgs2$_{ha}$ plasmids.

(**D**) **ML**-Rgs2$_{ha}$ expressed in HeLa cells in the presence of the coexpressed wild-type Teb4$_{3f}$ or $\mathrm{Teb4}_{3f}^{C9A}$. Note a strong decrease of **ML**-Rgs2$_{ha}$ in the presence of wild-type Teb4$_{3f}$ but not in the presence of its inactive missense mutant $\mathrm{Teb4}_{3f}^{C9A}$. Note also a much lower steady-state level of active Teb4$_{3f}$ (in comparison to the inactive $\mathrm{Teb4}_{3f}^{C9A}$), owing to the previously described self-targeting of active Teb4 for degradation (*24*).

(**E**) Same as in **D** but with **MR**-Rgs2$_{ha}$.

(**F**) Quantification of data in **C** (relative levels of the three **MX**-Rgs2$_{ha}$ proteins).

(**G**) Same as in **F** but quantification, using qRT-PCR, of the relative levels of mRNAs encoding the three **MX**-Rgs2$_{ha}$ proteins. Standard errors of means (SEMs), with qRT-PCR assays in triplicate for each sample, are indicated as well.

(**H**) Steady-state levels of transiently expressed (exogenous) **MR**-Rgs2$_{ha}$ in HeLa cells subjected to RNAi-mediated knockdowns of either endogenous Naa60, the cognate (for **MR**-Rgs2$_{ha}$) NatF Nt-acetylase, or endogenous Naa20, the non-cognate (for **MR**-Rgs2$_{ha}$) NatB Nt-acetylase. Note the increase in **MR**-Rgs2$_{ha}$ in response to RNAi of Naa60 but not of Naa20.

(**I**) Same as in **H** but with **MK**-Rgs2$_{ha}$, a derivative of wild-type **MQ**-Rgs2$_{ha}$ not encountered, thus far, among human patients, that contained another basic residue at position 2.

(**J**) ^{35}S-pulse-chases with **MX**-Rgs2$_{ha}$ (**X**=R, Q, L) in HeLa cells.

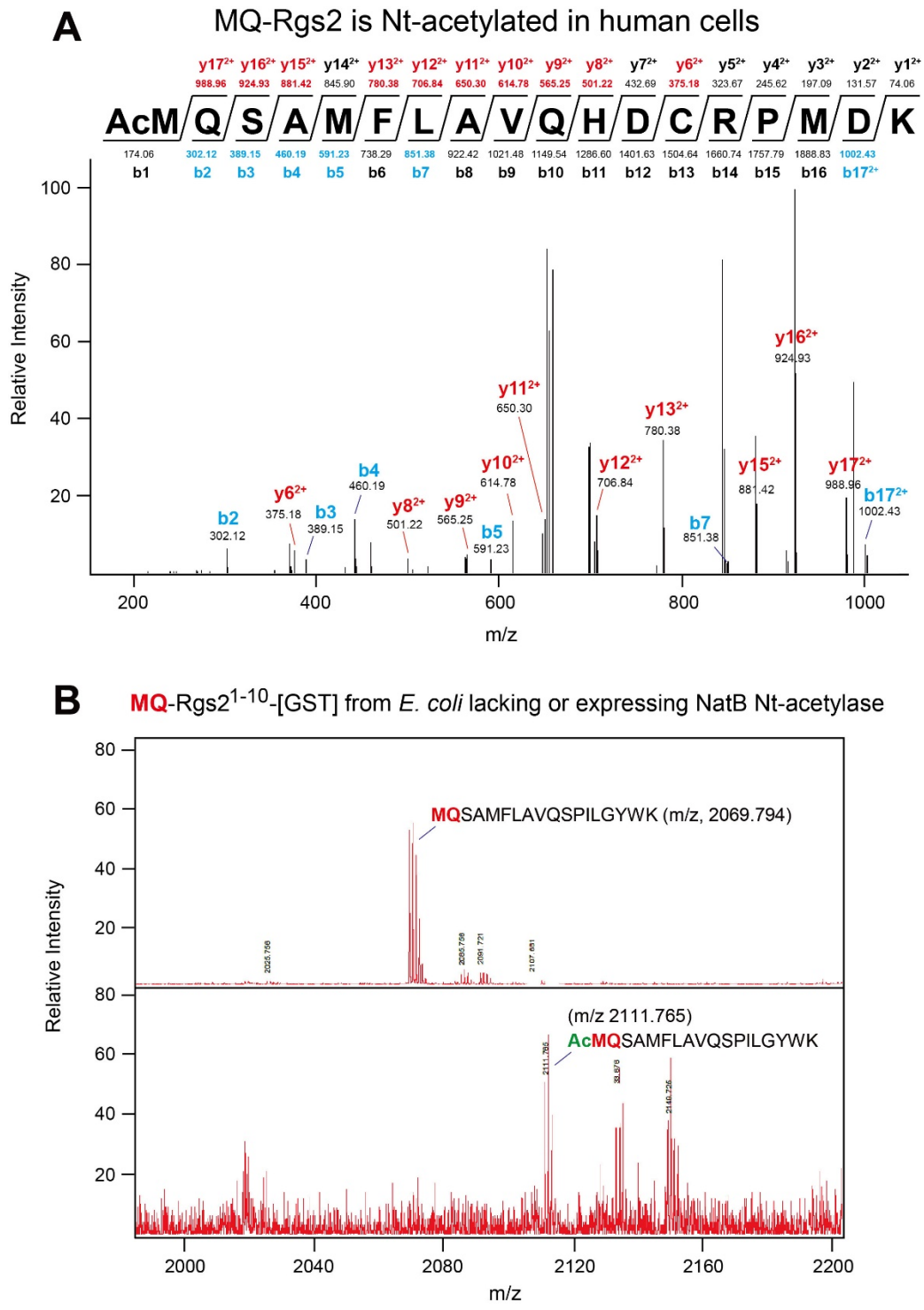


Fig. A2.S5. Mass spectrometric analyses of Nt-acetylation of MQ-Rgs2 and MQ-Rgs2$^{1\text{-}10}$-[GST].

(**A**) Wild-type **MQ**-Rgs2$_{ha}$ purified from human HEK293 cells is Nt-acetylated. (**B**) Non-Nt-acetylated (upper panel) and Nt-acetylated **MQ**-Rgs2$^{1\text{-}10}$-[GST] fusion (lower panel) purified from *E. coli* that either lacked or expressed, respectively, the *Schizosaccharomyces pombe* NatB Nt-acetylase (see Materials and Methods).

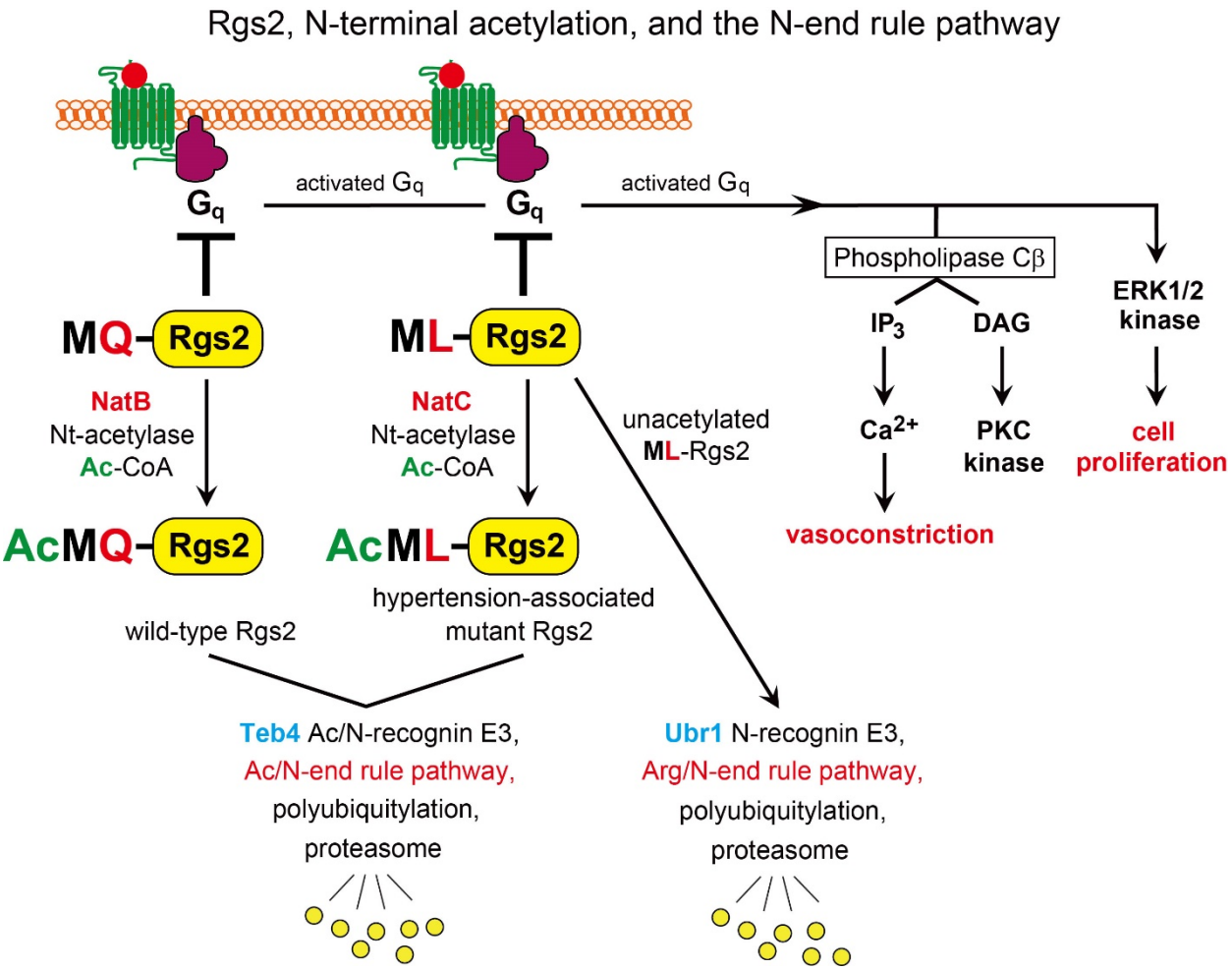


Fig. A2.S6. Rgs2, Teb4, N-terminal acetylation, and the N-end rule pathway. Rgs2 and Teb4 are, respectively, a substrate and Ac/N-recognin of the mammalian Ac/N-end pathway. Only some Rgs2-regulated processes (including the $G\pm_q$-mediated activation of Erk1/2 kinase) are shown. Both wild-type **MQ**-Rgs2 and its hypertension-associated mutant **MR**-Rgs2 (the latter is not shown) are Nt-acetylated and conditionally destroyed by the Teb4-mediated Ac/N-end rule pathway. In contrast, another hypertension-associated mutant, **ML**-Rgs2, is targeted by both the Ac/N-end rule pathway and the Arg/N-end rule pathway (see the main text) (fig. A3.S1C, D). Some abbreviations: IP_3, inositol trisphosphate. DAG, diacyglycerol.

RGS proteins as either identified or possible N-end rule substrates.

Identified RGS N-end rule substrates:

RGS proteins	N-termini	Degradation by the Arg/N-end rule pathway	Degradation by the Ac/N-end rule pathway
MQ-Rgs2, wild-type **1**	**MQ**SAMF-	No	Yes (Ac**MQ**SAMF-)
MR-Rgs2, mutant (hypertension-associated) **2**	**MR**SAMF-	No	Yes (Ac**MR**SAMF-)
ML-Rgs2, mutant (hypertension-associated) **3**	**ML**SAMF-	Yes (**ML**SAMF-) unacetylated	Yes (Ac**ML**SAMF-)
CK-Rgs4, wild-type **4**	M(Y)**CK**GLA-	Yes (Arg**C*K**GLAG-) Cys: oxidized, arginylated	Possibly, to a minor extent, if Cys is also Nt-acetylated
CK-Rgs5, wild-type **5**	M(Y)**CK**GLA-	Yes (Arg**C*K**GLAA-) Cys: oxidized, arginylated	Possibly, to a minor extent, if Cys is also Nt-acetylated
CK-Rgs16, wild-type **6**	M(Y)**CR**TLA-	Yes (Arg**C*R**TLAA-) Cys: oxidized, arginylated	Possibly, to a minor extent, if Cys is also Nt-acetylated
Possible (predicted) RGS N-end rule substrates:			
MR-Rgs1, wild-type **7**	**MR**AAAI-	No	Yes (Ac**MR**AAAI-)
AQ-Rgs6, wild-type **8**	M(Y)**AQ**GSG-	No	Yes (Ac**AQ**GSG-)
AQ-Rgs7, wild-type **9**	M(Y)**AQ**GNN-	No	Yes (Ac**AQ**GNN-)
AA-Rgs8, wild-type **10**	M(Y)**AA**LLM-	Unclear	Yes (Ac**AA**LLM-)
MF-Rgs12, wild-type **11**	**MF**RAGE-	Yes (**MF**RAGE-) unacetylated	Yes (Ac**MF**RAGE-)
SR-Rgs13, wild-type **12**	M(Y)**SR**RNC-	No	Yes (Ac**SR**RNC-)
MR-Rgs17, wild-type **13**	**MR**KRQQ	No	Yes (Ac**MR**KRQQ-)

Fig. A2.S7. Mammalian RGS proteins as identified or predicted N-end rule substrates. Human RGS proteins as identified or possible (predicted) substrates of either the Ac/N-end rule pathway, or the Arg/N-end rule pathway, or both pathways at the same time, depending on the Nt-acetylation status of a specific RGS protein. See also the main text.

Table S1. *S. cerevisiae* strains used in this study.

Strains	Relevant genotypes	Sources
JD53	*MATα trp1- 63 ura3-52 his3- 200 leu2-3112. lys2-801*	(*110*)
BY4742	*MATα his3-1 leu2-0 lys2-0 ura3-0 can1-100*	Open Biosystems, Inc.
CHY223	*doa10Δ::kanMX6* in JD53	(*13*)
CHY272	*naa30Δ::kanMX4* in BY4742	Open Biosystems, Inc.
CHY345	*ubr1Δ::LEU2* in BY4742	(*16*)
CHY349	*naa30Δ::kanMX4 ubr1Δ::LEU2* in BY4742	(*16*)
CHY367	*naa20Δ::natNT2* in JD53	This study

Table S2. Plasmids used in this study.

Plasmids	Descriptions	Sources
pCH12	pET30a	Novagen, Inc.
pCH15	pACYCDuet	Novagen, Inc.
pCH61	pcDNA3.1(+) with ha_2	Hwang's lab collection
pCH692	p316CUP1	Hwang's lab collection
pCH759	Teb4 in pcDNA3.1 myc/His (-)	(*24*)
pCH760	$Teb4^{C9A}$ in pcDNA3.1 myc/His(-)	(*24*)
pCH766	MQ-$Rgs2_{ha}$ in pcDNA3(+)	(*11*)
pCH767	ML-$Rgs2_{ha}$ in pcDNA3(+)	(*11*)
pCH768	MR-$Rgs2_{ha}$ in pcDNA3(+)	(*11*)
pCH821	MPQ-$Rgs2_{ha}$ in pcDNA3(+)	This study
pCH879	$Teb4_{f3h}$ in pcDNA3.1	This study
pCH881	$Teb4^{C9A}{}_{f3h}$ in pcDNA3.1	This study
pCH3002	MQ-Rgs2-GST in pACYCDuet	This study
pCH3015	pET30a	Qiagen, Inc.
pCH3025	*S. pombe Naa20*$^+$*-NAA25*$^+$ in pACYCDuet	(*32*)
pCH3056	MQ-Rgs2-GST in pET30a	This study
pCH3080	MQ-$Rgs2^{1\text{-}10}$-GST in pET30a	This study
pCH3132	$M3_{ha2}$ in pcDNA3.1(+)	This study
pCH5050	MQ-$Rgs2_{ha}$ in pRS316, with P_{CUP1}	This study
pCH5051	ML-$Rgs2_{ha}$ in pRS316, with P_{CUP1}	This study
pCH5064	G±q_{ha2} in pcDNA3.1(+)	This study

Table S3. Some PCR primers used in this study.

Name	Primer sequences
OCH1091	5’-CTTGGTACCATG CCA CAAAGTGCTATGTTCTTGGCTGTT-3’
OCH1100	5’-AACCTCGAGCTAAGCGTAATCTGGAACATCGTATGGGTAAGAACCTGTAGCATGAGGCTCTGTGGTGAT-3’
OCH1265	5’-GGTAAGCTTGACTATAAAGACCATGACGGTGATTATAAAGATCATGATATCGATTACAAG-3’
OCH1266	5’-AATCTTAAGTTAATGATGGTGATGGTGATGAGAACCCTTGTCATCGTCATCCTTGTAATC-3’
OCH1514	5'-TATCTCGAGTCATTTTGGAGGATGGTCGCCACCACC -3'
OCH3006	5’-TCCCATATGCAAAGTGCTATGTTCTTGGCTGTT-3’
OCH3013	5’-CGCGGATCCTGTAGCATGAGGCTCTGTGGTGAT-3’
OCH3015	5’-CGCGGATCCATGTCCCCTATACTAGGTTATTGG-3’
OCH3095	5'-TCCCATATGCAAAGTGCTATGTTCTTGGCTGTTCAATCCCCTATACTAGGTTATTGG-3'
OCH3178	5’-ACAGGATCCATGACCTTGCACAATAACAGTAC-3’
OCH3179	5’-ACATCTAGACAAGGCCTGCTCGGGTGCGCGCTTG-3’
OCH5070	5’-CCCGGATCCATGCAAAGTGCTATGTTCTTGGCT-3’
OCH5071	5’-CCCGGATCCATGTTAAGTGCTATGTTCTTGGCTGTTCAA-3’
OCH5073	5’-CCCGAGCTCCTAAGCGTAATCTGGAACATCGTATGGGTAGCTAGCTGTAGCATGAGG CTCTGTGGTGAT-3’
OCH5090	5’-GTCGGTACCATGACTCTGGAGTCCATCATGGCG-3’
OCH5091	5’-GCTCTCGAGGACCAGATTGTACTCCTTCAGGTT-3’

REFERENCES

1. R. J. Lefkowitz, A brief history of G-protein coupled receptors. Angew. Chem. 52, 6366-6378 (2013).
2. B. Kobilka, The structural basis of G-protein-coupled receptor signaling. Angew. Chem. 52, 6380-6388 (2013).
3. B. Sjögren, L. L. Blazer, R. R. Neubig, Regulators of G protein signaling proteins as targets for drug discovery. Prog. Mol. Biol. Transl. Sci. 91, 81-119 (2010).
4. A. J. Kimple, D. E. Bosch, P. M. Giguère, D. P. Siderovski, Regulators of G-protein signaling and their G± substrates: promises and challenges in their use as drug discovery targets. Pharmacol. Rev. 63, 728-749 (2011).
5. P. Chidiac, A. J. Sobiesiak, K. N. Lee, R. Gros, C. H. Nguyen, The eIF2B-interacting domain of RGS2 protects against GPCR agonist-induced hypertrophy in neonatal rat cardiomyocytes. Cell Signal. 26, 1226-1234 (2014).
6. M. Matsuo, S. L. Coon, D. C. Klein, RGS2 is a feedback inhibitor of melatonin production in the pineal gland. FEBS Lett. 587, 1392-1398 (2013).
7. M. R. Nance et al., Structural and functional analysis of the regulator of G protein signaling 2-(G±)q complex. Structure 21, 438-448 (2013).
8. K. M. Tang et al., Regulator of G-protein signaling-2 mediates vascular smooth muscle relaxation and blood pressure. Nat. Med. 12, 1506-1512 (2003).
9. S. P. Heximer et al., Hypertension and prolonged vasoconstrictor signaling in Rgs2-deficient mice. J. Clin. Investig. 111, 1259 (2003).

10. J. Yang et al., Genetic variations of regulator of G-protein signaling 2 in hypertensive patients and in the general population. J. Hypertens. 8, 1497-1505 (2005).

11. J. Bodenstein, R. K. Sunahara, R. R. Neubig, N-terminal residues control proteasomal degradation of Rgs2, Rgs4, and Rgs5 in human embryonic kidney 293 cells. Mol. Pharmacol. 71, 1040-1050 (2007).

12. A. Varshavsky, The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20, 1298-1345 (2011).

13. C.-S. Hwang, A. Shemorry, A. Varshavsky, N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327, 973-977 (2010).

14. C.-S. Hwang, A. Shemorry, A. Varshavsky, The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 12, 1177-1185 (2010).

15. A. Shemorry, C.-S. Hwang, A. Varshavsky, Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50, 540-551 (2013).

16. H.-K. Kim et al., The N-terminal methionine of cellular proteins as a degradation signal. Cell 156, 158-169 (2014).

17. J. M. Kim, C. S. Hwang, Crosstalk between the Arg/N-end and Ac/N-end rule. Cell Cycle 13, 1366-1367 (2014).

18. T. S. Tasaki, S. M. Sriram, K. S. Park, Y. T. Kwon, The N-end rule pathway. Annu. Rev. Biochem. 81, 261-289 (2012).

19. D. J. Gibbs, J. Bacardit, A. Bachmair, M. J. Holdsworth, The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. Trends Cell Biol. 24, 603-611 (2014).

20. A. Mogk, R. Schmidt, B. Bukau, The N-end rule pathway of regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 17, 165-172 (2007).

21. K. K. Starheim, K. Gevaert, T. Arnesen, Protein N-terminal acettyltransferases: when the start matters. Trends Biochem. Sci. 37, 152-161 (2012).

22. P. Van Damme et al., NatF contrutes to an evolutionary shift in protein N-terminal acetylation and is is important for normal chromosome segregation. PLoS Genet. 7, e1002169 (2011).

23. S. G. Kreft, L. H.-C. Wang, M. Hochstrasser, Membrane topology of the yeast endoplasmic reticulum-localized ubiquitin ligase Doa10 and comparison with its human ortholog TEB4 (MARCH-VI). J. Biol. Chem. 281, 4646-4653 (2006).

24. G. Hassink et al., TEB4 is a C4HC3 RING-finger containing ubiquitin ligase of the endoplasmic reticulum. Biochem. J. 388, 647-655 (2005).

25. J. K. Monda et al., Structural conservation of distinctive N-terminal acetylation-dependent interactions across a family of mammalian NEDD8 ligation enzymes. Structure 21, 1-12 (2012).

26. N. Zelcer et al., The E3 ubiquitin ligase MARCH6 degrades squalene monooxygenase and affects 3-hydroxy-3-methyl-glutaryl coenzyme A reductase and the cholesterol synthesis pathway. Mol. Cell. Biol. 34, 1262-1270 (2014).

27. F. M. Ausubel et al., Current Protocols in Molecular Biology., (Wiley-Interscience, New York, 2010).

28. F. Sherman, Getting started with yeast. Meth. Enzymol. 194, 3-21 (1991).

29. A. Baudin, O. Ozier-Kalogeropolous, A. Denoel, A. Lacroute, C. Cullin, A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. Nucl. Ac. Res. 21, 3329-3330 (1993).

30. C. Janke et al., A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast 21, 947-962 (2004).

31. T. D. Schmittgen, K. J. Livak, Analyzing real-time PCR data by the comparative CT method. Nature Protocols 5, 1101-1108 (2008).

32. M. Johnson, A. T. Coulton, M. A. Geeves, D. P. Mulvihill, Targeted amino-terminal acetylation of recombinant proteins in E. coli. . PLoS One 5, e15801 (2010).

33. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. 26, 1367-1372.

34. Q. Xiao, F. Zhang, B. A. Nacev, J. O. Liu, D. Pei, Protein N-terminal processing: substrate specificity of Escherichia coli and human methionine aminopeptidases. Biochemistry 49, 5588-5599 (2010).

35. F. Eisele, D. H. Wolf, Degradation of misfolded proteins in the cytoplasm by the ubiquitin ligase Ubr1. FEBS Lett. 582, 4143-4146 (2008).

36. J. W. Heck, S. K. Cheung, R. Y. Hampton, Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. Proc. Natl. Acad. Sci. USA 107, 1106-1111 (2010).

37. E. Graciet, F. Wellmer, The plant N-end rule pathway: structure and functions. Trends Plant Sci. 15, 447-453 (2010).

38. D. A. Dougan, D. Micevski, K. N. Truscott, The N-end rule pathway: from recognition by N-recognins to destruction by AAA+ proteases. Biochim. Biophys. Acta 1823, 83-91 (2011).

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

40. C.-S. Hwang, A. Shemorry, A. Varshavsky, Two proteolytic pathways regulate DNA repair by co-targeting the Mgt1 alkylguanine transferase. Proc. Natl. Acad. Sci. USA 106, 2142-2147 (2009).

41. C.-S. Hwang, A. Varshavsky, Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 105, 19188-19193 (2008).

42. Z. Xia et al., Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. J. Biol. Chem. 283, 24011-24028 (2008).

43. R.-G. Hu, H. Wang, Z. Xia, A. Varshavsky, The N-end rule pathway is a sensor of heme. Proc. Natl. Acad. Sci. USA 105, 76-81 (2008).

44. R.-G. Hu et al., The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437, 981-986 (2005).

45. C.-S. Hwang et al., Ubiquitin ligases of the N-end rule pathway: assessment of mutations in UBR1 that cause the Johanson-Blizzard syndrome. PLoS One 6, e24925 (2011).

46. H. Wang, K. I. Piatkov, C. S. Brower, A. Varshavsky, Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell 34, 686-695 (2009).

47. C. S. Brower, A. Varshavsky, Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE 4, e7757 (2009).

48. M. Zenker 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. 37, 1345-1350 (2005).

49. R. Prasad, S. Kawaguchi, D. T. W. Ng, A nucleus-based quality control mechanism for cytosolic proteins. Mol. Biol. Cell 21, 2117-2127 (2010).

50. S. Kurosaka et al., Arginylation-dependent neural crest cell migration is essential for mouse development. PLoS Genet. 6, e1000878 (2010).

51. F. Zhang, S. Saha, S. A. Shabalina, A. Kashina, Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. Science 329, 1534-1537 (2010).

52. M. J. Lee et al., RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 102, 15030-15035 (2005).

53. K. I. Piatkov, C. S. Brower, A. Varshavsky, The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. Proc. Natl. Acad. Sci. USA 109, E1839-E1847 (2012).

54. K. I. Piatkov, L. Colnaghi, M. Bekes, A. Varshavsky, T. T. Huang, The auto-generated fragment of the Usp1 deubiquitylase is a physiological substrate of the N-end rule pathway. Mol. Cell 48, 926-933 (2012).

55. Y. T. Kwon et al., An essential role of N-terminal arginylation in cardiovascular development. Science 297, 96-99 (2002).

56. M. J. Lee et al., Characterization of arginylation branch of the N-end rule pathway in G-protein-mediated proliferation and signaling of cardiomyocytes. J. Biol. Chem. 287, 24043-24052 (2012).

57. W. S. Choi et al., Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat. Struct. Mol. Biol. 17, 1175-1181 (2010).

58. E. Matta-Camacho, G. Kozlov, F. F. Li, K. Gehring, Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat. Struct. Mol. Biol. 17, 1182-1188 (2010).

59. F. Licausi et al., Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature 479, 419-422 (2011).

60. R. Sasidharan, A. Mustroph, Plant oxygen sensing is mediated by the N-end rule pathway: a milestone in plant anaerobiosis. Plant Cell 23, 4173-4183 (2011).

61. D. J. Gibbs et al., Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature 479, 415-418 (2011).

62. H. Rao, F. Uhlmann, K. Nasmyth, A. Varshavsky, Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature 410, 955-960 (2001).

63. S. T. Kim et al., The N-end rule proteolytic system in autophagy. Autophagy 9, 1100-1103 (2013).

64. G. Zhang, R. K. Lin, Y. T. Kwon, Y. P. Li, Signaling mechanism of tumor cell-induced up-regulation of E3 ubiquitin ligase UBR2. FASEB J. 27, 2893-2901 (2013).

65. T. Tasaki et al., UBR box N-recognin-4 (UBR4), an N-recognin of the N-end rule pathway, and its role in yolk sac vascular development and autophagy. Proc. Natl. Acad. Sci. USA 110, 3800-3805 (2013).

66. H. Fujiwara, N. Tanaka, I. Yamashita, K. Kitamura, Essential role of Ubr11, but not Ubr1, as an N-end rule ubiquitin ligase in Schizosaccharomyces pombe. Yeast 30, 1-11 (2013).

67. K. Kitamura, H. Fujiwara, The type-2 N-end rule peptide recognition activity of Ubr11 ubiquitin ligase is required for the expression of peptide transporters. FEBS Lett. 587, 214-219 (2013).

68. J. Y. An et al., UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. Proc. Natl. Acad. Sci. USA 107, 1912-1917 (2010).

69. J. Y. An et al., UBR2 of the N-end rule pathway is required for chromosome stability via histone ubiquitylation in spermatocytes and somatic cells. PLoS One 7, e37414 (2012).

70. R. Sultana, M. A. Theodoraki, A. J. Caplan, UBR1 promotes protein kinase quality control and sensitizes cells to Hsp90 inhibition. Exp. Cell Res. 18, 53-60 (2012).

71. P. Lee, M. E. Sowa, S. P. Gygi, J. W. Harper, Alternative ubiquitin activation/conjugation cascades interact with N-end rule ubiquitin ligases to control degradation of RGS proteins. Mol. Cell 43, 392-405 (2011).

72. G. Román-Hernández, J. Y. Hou, R. A. Grant, R. T. Sauer, T. A. Baker, The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. Mol. Cell 43, 217-228 (2011).

73. F. Yang et al., The ubiquitin ligase Ubr2, a recognition E3 component of the N-end rule pathway, stabilizes Tex.19.1 during spermatogenesis. PLoS One 5, e14017 (2010).

74. R. L. Ninnis, S. K. Spall, G. H. Talbo, K. N. Truscott, D. A. Dougan, Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. EMBO J. 28, 1732-1744 (2009).

75. R. Schmidt, R. Zahn, B. Bukau, A. Mogk, ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. Mol. Microbiol. 72, 506-517 (2009).

76. T. J. Holman et al., The N-end rule pathway promotes seed germination and establishment through removal of ABA sensitivity in Arabidopsis. Proc. Natl. Acad. Sci. USA 106, 4549-4554 (2009).

77. H. Cai, M. Hauser, F. Naider, J. M. Becker, Differential regulation and substrate preferences in two peptide transporters of Saccharomyces cerevisiae. Eukaryot. Cell 6, 1805-1813 (2007).

78. P. Schnupf, J. Zhou, A. Varshavsky, P. D.A., Listeriolysin O secreted by Listeria monocytogenes into the host cell cytosol is degraded by the N-end rule pathway. Inflammation & Immunity 75, 5135-5147 (2007).

79. R.-G. Hu et al., Arginyl-transferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem., 32559-32573 (2006).

80. E. Graciet et al., Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc. Natl. Acad. Sci. USA 103, 3078-3083 (2006).

81. S. Kurosaka et al., Arginylation regulates myofibrils to maintain heart function and prevent dilated cardiomyopathy. J. Mol. Cell. Cardiol. 53, 333-341 (2012).

82. J. Wang et al., Arginyltransferase is an ATP-independent self-regulating enzyme that forms distinct functional complexes in vivo. Chem. Biol. 18, 121-130 (2011).

83. M. B. Decca et al., Post-translational arginylation of calreticulin: a new isospecies of calreticulin component of stress granules. J. Biol. Chem. 282, 8337-8345 (2007).

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

85. G. Boso, T. Tasaki, Y. T. Kwon, N. V. Somia, The N-end rule and retroviral infection: no effect on integrase. Virology J. 10, 233 (2013).

86. S. Saha, A. Kashina, Posttranslational arginylation as a global biological regulator. Dev. Biol. 358, 1-8 (2011).

87. C. Belzil et al., A Ca2+-dependent mechanism of neuronal survival mediated by the microtubule-associated protein p600. J. Biol. Chem. 288, 24452-24464 (2013).

88. K. Yamano, R. J. Youle, PINK1 is degraded through the N-end rule pathway. Autophagy 9, 1758-1769 (2013)10.4161/auto.24633).

89. T. Arnesen et al., Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast to humans. Proc. Natl. Acad. Sci. USA 106, 8157-8162 (2009).

90. P. Van Damme et al., N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proc. Natl. Acad. Sci. USA 109, 12449-12454 (2012).

91. D. J. Gibbs et al., Nitric oxide sensing in plants is mediated by proteolytic control of group VII ERF transcription factors. Mol. Cell 53, 369-379 (2014).

92. F. Licausi, C. Pucciariello, P. Perata, New role for an old rule: N-end rule-mediated degradation of ethylene responsive factor proteins governs low oxygen response in plants. J. Integr. Plant Biol. 55, 31-39 (2013).

93. D. A. Weits et al., Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nat Commun. 5, 3425 (2014).

94. M. Gautsch et al., The yeast N-alpha-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides. Mol. Cell. Biol. 23, 7403-7414 (2003).

95. B. Polevoda, S. Brown, T. S. Cardillo, S. Rigby, F. Sherman, Yeast N(alpha)-terminal acetyltransferases are associated with ribosomes. J. Cell. Biochem. 103, 492-508 (2008).

96. B. Polevoda, F. Sherman, N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins. J. Mol. Biol. 325, 595-622 (2003).

97. A. O. Helbig et al., Profiling of N-acetylated protein termini provides in-depth insights into the N-terminal nature of the proteome. Mol Cell Proteom. 9, 928-939 (2010).

98. J. R. Mullen et al., Identification and characterization of genes and mutants for an N-terminal acetyltransferase from yeast. EMBO J. 8, 2067-2075 (1989).

99. E. C. Park, J. W. Szostak, ARD1 and NAT1 proteins form a complex that has N-terminal acetyltransferase activity. EMBO J. 11, 2087-2093 (1992).

100. S. Goetze et al., Identification and functional characterization of N-terminally acetylated proteins in Drosophila melanogaster. PLoS Biol. 7, e1000236 (2009).

101. T. Arnesen, Toward a functional understanding of protein N-terminal acetylation. PLoS Biol. 9, e1001074 (2011).

102. K. K. Starheim, D. Gromyko, R. Velde, J. E. Varhaug, T. Arnesen, Composition and biological significance of the human N-alpha terminal acetyltransferases. BMC Proceedings 3 (Suppl 6), S3 (2009).

103. K. Hole et al., The human N-alpha-acetyltransferase 40 (hNaa40p/hNatD) is conserved from yeast and N-terminally acetylates histones H2A and H4. PLoS One 6, e24713 (2011).

104. K. Helsens et al., Bioinformatics analysis of a Saccharomyces cerevisiae N-terminal proteome provides evidence of alternative translation initiation and post-translational N-terminal acetylation. J. Proteome Res. 10, 3578-3589 (2011).

105. B. Polevoda, J. Hoskins, F. Sherman, Properties of Nat4, an N-alpha-acetyltransferase of Saccharomyces cerevisiae that modifues N-termini of histones H2A and H4. Mol. Cell. Biol. 29, 2913-2924 (2009).

106. W. V. Bienvenut et al., Comparative large scale characterization of plant versus mammalian proteins reveals similar and idiosyncratic N-±-acetylation features. Mol. Cell. Proteomics 111, M111.015131 (2012).

107. P. Van Damme et al., Proteome-derived peptide libraries allow detailed analysis of the substrate specificities of N(alpha)-acetyltransferases and point to hNaa10p as the post-translational actin N(alpha)-acetyltransferase. Mol. Cell. Proteomics 10, M110.004580 (2011).

108. M. Perrot, A. Massoni, H. Boucherie, Sequence requirements for Nalpha-terminal acetylation of yeast proteins by NatA. Yeast 25, 513-527 (2008).

109. B. Polevoda, T. Arnesen, F. Sherman, A synopsis of eukaryotic N-terminal acetyltransferases: nomenclature, subunits and substrates. BMC Proceedings 3, S2 (2009).

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

APPENDIX 3:

SUPPLEMENTAL MATERIAL FOR CHAPTER 6

Detailed (extended) legend to Fig. 1 of the main text:

Fig. 1. The mammalian N-end rule pathway and the Ate1 arginyltransferase (R-transferase). The main determinant of an N-degron is a destabilizing N-terminal residue of a protein. N-terminal residues are indicated by single-letter abbreviations for amino acids. A yellow oval denotes the rest of a protein substrate. Recognition components of the N-end rule pathway are called N-recognins. In eukaryotes, the N-end rule pathway consists of two branches, the Arg/N-end rule pathway (*A*) and the Ac/N-end rule pathway (*B*) (refs. (1-60) and refs. therein).

(*A*) The Arg/N-end rule pathway targets specific unacetylated N-terminal residues. In the yeast *S. cerevisiae*, the Arg/N-end rule pathway is mediated by the Ubr1 N-recognin, a 225 kDa RING-type E3 Ub ligase and a part of the targeting apparatus comprising a complex of the Ubr1-Rad6 and Ufd4-Ubc4/5 holoenzymes (1, 5, 10). In multicellular eukaryotes several functionally overlapping E3 Ub ligases (Ubr1, Ubr2, Ubr4, and Ubr5) function as N-recognins of this pathway. An N-recognin binds to the "primary" destabilizing N-terminal residues Arg, Lys, His, Leu, Phe, Tyr, Trp and Ile. In contrast, the N-terminal Asn, Gln, Asp, and Glu residues (as well as Cys, under some metabolic conditions) are destabilizing owing to their preliminary enzymatic modifications. These modifications include the Nt-deamidation of N-terminal Asn and Gln by the Ntan1 and Ntaq1 Nt-amidases, respectively, and the Nt-arginylation of N-terminal Asp and Glu by the Ate1 R-transferase, which can also Nt-arginylate oxidized Cys, either Cys-sulfinate or Cys-sulfonate. These derivatives can form in animal and plant cells through oxidation of N-terminal Cys by nitric oxide (NO) and oxygen, and also by N-terminal Cys-oxidases (15, 23, 30, 32, 61-64). In addition to its type-1 and

type-2 binding sites that recognize, respectively, the basic and bulky hydrophobic unacetylated N-terminal residues, an N-recognin such as Ubr1 contains other substrate-binding sites as well. These sites recognize substrates that are targeted through their internal (non-N-terminal) degrons, as indicated on the diagram (1). Hemin (Fe^{3+}-heme) binds to the Ate1 R-transferase, inhibits its Nt-arginylation activity and accelerates its *in vivo* degradation. Hemin also binds to Ubr1 and alters its functional properties, in ways that remain to be understood (14). As shown in the diagram, the unacetylated Met of the N-terminal Met-hydrophobic residue (M$) motif was found to be recognized by both yeast and mammalian Ubr1, thereby greatly expanding the substrate range of the Arg/N-end rule pathway (60).

(*B*) The Ac/N-end rule pathway. This diagram illustrates the mammalian Ac/N-end rule pathway through extrapolation from its *S. cerevisiae* version. Red arrow on the left indicates the removal of N-terminal Met by Met-aminopeptidases (MetAPs). N-terminal Met is retained if a residue at position 2 is nonpermissive (too large) for MetAPs. If the retained N-terminal Met or N-terminal Ala, Ser, Thr are followed by acetylation-permissive residues, the above N-terminal residues are $N^{\pm}$-terminally acetylated (Nt-acetylated) by ribosome-associated Nt-acetylases (65-67). Nt-terminal Val and Cys are Nt-acetylated relatively rarely, while N-terminal Pro and Gly are almost never Nt-acetylated. (N-terminal Gly is often N-myristoylated.) N-degrons and N-recognins of the Ac/N-end rule pathway are called Ac/N-degrons and Ac/N-recognins, respectively (1). The term “secondary” refers to Nt-acetylation of a destabilizing N-terminal residue before a protein can be recognized by a cognate Ac/N-recognin. As described in the main text, the human Teb4 E3 ubiquitin (Ub) ligase was identified, in the

present study, as the first Ac/N-recognin of the previously unexamined mammalian Ac/N-end rule pathway. Natural Ac/N-degrons are regulated by their reversible steric shielding in protein complexes (60, 68).

(*C*) The bidirectional $_{Dfa}P_{Ate1}$ promoter upstream of exon 1B of mouse *Ate1* (1, 50, 69).

(*D*) Four major mouse Ate1 R-transferase isoforms and their designations (1, 50).

SI EXPERIMENTAL PROCEDURES

Yeast Two-Hybrid (Y2H) Assays. Two-hybrid (Y2H) assays were carried out using the BD Matchmaker kit (BD Biosciences, Palo Alto, CA). Briefly, *S. cerevisiae* AH109 (*MAT**a** trp1-901 leu2-3, 112 ura3-52 his3-200 gal4• gal80, LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3 GAL2$_{UAS}$-GAL2$_{TATA}$-ADE2 URA3::MEL1$_{UAS}$-MEL1$_{TATA}$-lacZ, MEL1*) was transformed with pCB132, which expressed the Gal4^{DBD}-ATE1^{1B7A} fusion protein (Table S1). The resulting strain (AH109-Ate1^{1B7A}) were mated with *S. cerevisiae* Y187 (*MAT± ura3-52 his3-200 ade2-101 trp1-90 leu2-3, 112, gal4• gal80• met- URA3::GAL1$_{UAS}$-GAL1$_{TATA}$-lacZ MEL1*) that had been pre-transformed with a mouse testis cDNA pACT library (Clontech). The resulting diploids were incubated on plates with synthetic-complete (SC) medium containing 5 mM 3-aminotriazole (Sigma) and lacking His, Trp and Leu. [Trp$^+$ Leu$^+$ His$^+$] isolates (~1.5 × 10^6 cells) were further assayed for their growth on His-lacking, Ade-lacking media, and also for their ² -galactosidase (² gal) activity. Colonies that were both [Trp$^+$ Leu$^+$ His$^+$ Ade$^+$] and exhibited ² gal activity were then grown in Leu-lacking, Trp-containing media to facilitate the loss of pCB132. The pA6 plasmid was recovered from one strongly positive colony (called “A6”). The

sequencing of pA6 insert identified full-length mouse Liat1 as a potential Ate1^{1B7A}-binding protein.

Vector-swapping experiments to verify detected interactions were then carried out by excising (using *Bam*HI and *Sal*I) the *Ate1^{1B7A}* open reading frame (ORF) from pCB132, and by excising (using *Nde*I and *Xho*I) the Liat1 ORF from the pA6 plasmid. The *Ate1^{1B7A}* ORF was then subcloned into pACT2 (Table S1) to yield pCB135, the *Liat1* ORF was subcloned into pAS2 to yield pCB239, and two-hybrid assays were repeated as described above.

Additional two-hybrid assays with Liat1 and its derivatives were carried out by cotransformation of *S. cerevisiae* AH109 with plasmids expressing Gal4^{DBD}-Liat1 (pCB239) and either Gal4^{AD}-Ate1^{1B7B} (pCB165) or Gal4^{AD}-Ate1^{1A7A} (pCB167), as well as with pCB432 (expressing Gal4^{AD}-Liat1) and either pCB164 (expressing Gal4^{DBD}-Ate1^{1B7B}) or pCB166 (expressing Gal4^{DBD}-Ate1^{1A7A}). The mapping of Ate1-Liat1 interactions was carried out by cotransformation of *S. cerevisiae* AH109 with pCB132 (expressing Gal4^{DBD}-ATE1^{1B7A}) and plasmids that expressed various (indicated in Fig. 3B) segments of mouse Liat1 fused to the Gal4^{AD} (the plasmids pCB432, pCB433, pCB434, pCB435, pCB436, pCB437, pCB438, pCB439, pCER-020, pCER-21, pCER-022) (Table S1). Positive two-hybrid controls employed plasmids expressing Gal4^{DBD}-WASP (pCB142) and Gal4^{AD}-Cdc42 (pCB141) (Table S1).

A separate two-hybrid screen for Liat1-binding proteins was carried out by mating *S. cerevisiae* AH109 (pre-transformed with pCB239 (Table S1)) to *S. cerevisiae* Y187 (pre-transformed with a mouse brain cDNA library (Clontech)), and by carrying out two-hybrid assays as described above.

***In Vitro* Arginylation Assay.** The N-terminal arginylation (Nt-arginylation) assay was performed as described previously (50), with slight modifications. Briefly, the untagged mouse Liat1, $Ate1^{1A7A}$, $Ate1^{1A7B}$, $Ate1^{1B7A}$, and $Ate1^{1B7B}$ were expressed as described below, using the Ub fusion/deubiquitylase-based pH10UE expression system (71). The Nt-arginylation reporter X-DHFRbt (X=Asp, Cys or Arg-Cys) was prepared by expressing His_{10}-Ub-X-DHFRbt (X = Asp, Cys, or Arg-Cys) using the pH10UE expression/purification system and pre-cleaved with the Usp2cc deubiquitylase (71) as described above, except that the cleaved reporter protein was not purified further before Nt-arginylation assays. For Nt-arginylation, purified Liat1 and specific Ate1 isoforms were mixed (as indicated in the main text and the legend to Fig. A3.S5) together with 5.8 µM ^{14}C-Arg (Perkin Elmer NEC267E050UC; 346 mCi/mmol)), *E. coli* tRNA (0.6 mg/ml), *E. coli* aminoacyl-tRNA synthetases (800 U/ml) (Sigma) in 5 mM ATP, 0.15 M KCl, 10 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 50 mM Tris-HCl (pH 8.0), and preincubated for 15 min at 37°C. A sample of (deubiquitylase-cleaved) reporter (10 µg) was then added, followed by incubation for 60 min at 37°C. Reactions were stopped by the addition of a concentrated SDS-sample buffer to the final concentration of 1x and heating at 95°C for 3 min, followed by SDS-4-12% PAGE, electroblotting to Immobilon-P membrane (Millipore, Billerica, MA), and ^{14}C-autoradiography, with quantification using PhosphorImager.

Tissue Extracts and Immunoblotting. Animal care and related procedures were carried out according to relevant NIH guidelines, and were approved (Protocol #1328) by the Institutional Animal Care and Use Committee, the Office of Laboratory Animal Research (OLAR) at the California Institute of Technology. Specific mouse tissues were

harvested and lysed in 0.05% NP40, 10% glycerol, 0.15 M NaCl, 2 mM EDTA, 1 mM dithiothreitol (DTT), 1 mM phenylmethylsulfonyl fluoride (PMSF), 50 mM HEPES (pH 7.5) containing freshly dissolved *"Complete EDTA-Free Protease Inhibitors"* (Roche), using the MP FastPrep-24 instrument and Lysing Matrix D (MP Biomedicals), with 2 or 3 runs at 6.5 m/s for 25 sec each, and with 5-min incubations on ice between the runs. The lysates were centrifuged at 10,000*g* for 20 min at 4°C. The supernatants were fractionated by SDS-12.5% PAGE, transferred to Immobilon-P PVDF membranes, and analyzed by immunoblotting with a rabbit polyclonal anti-C17orf97 antibody (Sigma; HPA023583). Immunoblots were visualized using the Odyssey (Li-Cor) system, with donkey anti-rabbit IgG conjugated to an infrared dye (Li-Cor; #926-32213) as secondary antibody.

Construction and Expression of Recombinant Proteins in BL21 (DE3) *E. coli*. Untagged mouse *Liat1* and *Ate1* cDNAs were amplified by PCR using primers flanked by the *Sac*II and *Hind*III sites. The resulting DNA fragments were subcloned into *Sac*II/*Hind*III-cut pH10UE, yielding the plasmids pCER-006 (Liat1), pCB407(Ate1^{1B7A}), pCB408 (Ate1^{1B7B}), pCB409 (Ate1^{1A7A}), and pCB410 (Ate1^{1A7B}), which expressed His_{10}-Ub-tagged fusions of the Ub fusion technique (71) (Table S1). A cDNA encoding the N-terminally triply ha-tagged mouse Liat1 (^{3ha}Liat1) was constructed using a multi-step PCR protocol. In particular, *Liat1* ORF was amplified from the plasmid pA6 (Table S1). The plasmid pCB403, expressing the N-terminally triple ha-tagged mouse Liat1 (^{3ha}Liat1), was constructed by subcloning the above ^{3ha}Liat1-encoding DNA fragment into *Sac*II/*Hind*III-cut pH10UE, yielding pCB403 (Table S1).

Either 2-liter cultures (for Ate1 production) or 4-liter cultures (for Liat1 production) of *E. coli* BL21(DE3) expressing one of the above-cited plasmids were grown at 37°C to an A_{600} of ~0.6 in Luria Broth (LB) containing ampicillin (50 µg/ml). The cultures were placed on ice for 45 min, followed by the addition of isopropyl-² 0G0 wklr j dodf wrvlgh (IPTG) to the final concentration of 0.25 mM and incubation, with shaking, for 6 hr at room temperature (RT). Cells were harvested by centrifugation at 2,000g for 10 min at 4°C. Cell pellets were resuspended in either 15 ml or 30 ml of phosphate-buffered saline containing 30% glycerol, 0.3 M NaCl, 12 mM imidazole, 20 mM ²-mercaptoethanol, and lysozyme at 1 mg/ml, followed by freezing in liquid nitrogen. Samples were then thawed slowly on ice and centrifuged at 27,000g for 30 min at 4°C. N-terminally His_{10}-tagged Ub fusions of specific Ate1 isoforms, of untagged Liat1 and of ^{3ha}Liat1 in the soluble fraction were purified by Ni-NTA chromatography, with elution of resin-bound proteins by 0.3 M imidazole. The resulting samples were dialyzed overnight at 4°C against 5% glycerol, 0.3 M NaCl, 2 mM ²-mercaptoethanol, and 50 mM Na_2HPO4-NaH_2PO4 (pH 8.0), containing EDTA-free protease inhibitor cocktail (Roche). The N-terminal His_{10}-tagged Ub moiety was then cleaved off by incubating dialyzed samples for 1 hr at 37°C with the purified Usp2cc deubiquitylase (71) (added at 1:10 molar ratio). The resulting samples were dialyzed against 5% glycerol, 0.15 M NaCl, 10 mM ²-mercaptoethanol, 50 mM Tris (pH 7.4), followed by purification of (untagged) Ate1 and either Liat1 or ^{3ha}Liat1 using Mono-S chromatography (Pharmacia Biotech: Id #9723125), with the elution gradient between 0.15 M and 1 M NaCl in the above buffer. Peak fractions were pooled and dialyzed

overnight at 4°C against 30% glycerol 0.15 M NaCl, 10 mM ² -mercaptoethanol, 50 mM Tris (pH 7.4).

Coimmunoprecipitation of Mouse Liat1 and Ate1 Isoforms. As indicated in the legend to Fig. 2, mouse *Ate1*$^{-/-}$ embryonic fibroblasts (EFs) were transfected with either pCB245, or pCB246, or pCB247, or pCB248 (expressing specific Ate1 isoforms) in the presence or absence of pCB179, expressing mouse ^{3ha}Liat1 (Table S1). Transfections were carried out using BioT reagent according to the manufacturer's protocol. 48 hrs after transfection, cell extracts were prepared in lysis buffer (10% glycerol, 0.5% Triton X-100, 0.15 M NaCl, 40 mM HEPES (pH7.6) containing *"complete EDTA-free protease inhibitors"* (Roche). Centrifugation-clarified extracts were incubated with anti HA-EZ View beads (Sigma E6799) for 3 hrs at 4°C, with rotation. Beads were then washed 3 times in lysis buffer. Immunoprecipitated proteins were eluted in 2xSDS-PAGE sample buffer, followed by SDS-12% PAGE, electroblotting of proteins onto PVDF membrane, and the detection of ^{3ha}Liat1 and Ate1 using anti-ha and anti-Ate1, respectively.

For "carrier-free" coimmunoprecipitation (co-IP) of purified recombinant ^{3ha}Liat1 and Ate1 isoforms, 1 µg of purified Ate1 isoforms were incubated for 1 hr at 4ºC in the presence or absence of 2 µg of purified Liat1 (~4-fold molar excess) in 10% glycerol, 0.5% Triton-X100, 0.15 M NaCl, 40 mM HEPES (pH 7.6) containing *"complete EDTA-free protease inhibitors"* (Roche). The mixture was then incubated for one more 1 hr in the presence of anti-Ate1 antibody (0.4 µg) and for 3 more hrs in the presence of 10 µl (settled bead volume) of Protein A-agarose (Replicon; IAP300; Lot#RN040321). The beads were then washed 3 times in the above buffer. The proteins were eluted in 2xSDS-

PAGE sample buffer and heated at 95ºC for 5 min, followed by SDS-12% PAGE and immunoblotting as described above.

Coimmunoprecipitation assays with Liat1 and putative Liat1 ligands. A cDNA encoding the N-terminally triply ha-tagged mouse Liat1 (^{3ha}Liat1) was constructed as briefly described in a section above. The final ORF, encoding mouse ^{3ha}Liat1, was subcloned into *Nhe*I/*Xho*I-cut pcDNA3.1 (Invitrogen), yielding pCB179 (Table S1). Untagged mouse *Ate1* cDNAs were amplified by PCR using primers flanked by the *Bam*HI and *Xho*I sites. The resulting DNA fragments were subcloned into *Bam*HI/*Xho*I-cut pcDNA-Neo, yielding the plasmids pCB245 (Ate1^{1B7A}), pCB246 (Ate1^{1B7B}), pCB247 (Ate1^{1A7A}), and pCB248 (Ate1^{1A7B}) (Table S1). DNA fragments (cDNAs) encoding mouse Jmjd6 and the 40S ribosomal protein S14 were amplified by PCR from a preparation of total RNA (isolated from pooled mouse tissues). DNA encoding a triple-flag epitope was added to the 3' end of these cDNAs using a multi-step PCR protocol with primers flanked with *Hind*III and *Apa*I, followed by subcloning into *Hind*III/*Apa*I-cut pcDNA-Neo, a step that yielded pCB447 (expressing the 40S ribosomal protein S14) and pCB458 (expressing Jmjd6). Coimmunoprecipitation assays with ^{3ha}Liat1 vis-á-vis Jmjd6^{3f} and S14^{3f} were carried similarly to Liat1-Ate1 co-IPs described above, except that mouse EFs were transfected with pCB179 (expressing ^{3ha}Liat1) and either pCB447 (S14^{3f}) or pCB458 (Jmjd6^{3f}) and proteins were immunoprecipitated from cell extracts using magnetic anti-FLAG beads (Sigma), followed by SDS-12% PAGE, and immunoblotting with anti-ha and anti-flag.

GST pulldowns. A cDNA encoding untagged mouse Liat1 cDNA was amplified by PCR using primers flanked by *Eco*RI sites and subcloned into *Eco*RI-cut pGEX-2t,

yielding pCB404 (Table S1). The *GST-Liat1* ORF was amplified from pCB404 by PCR using primers flanked by *Nhe*I and *Eco*RI sites, and the resulting DNA fragment was subcloned into *NheI/Eco*RI digested pcDNA3.1 (Invitrogen), yielding pCB405 (Table S1). For expression of GST alone in mammalian cells, cDNA encoding GST was amplified from pGEX-2t using primers flanked by *Nhe*I and *Xho*I sites, and the resulting DNA fragment was subcloned into *Nhe*I/*Xho*I-cut pcDNA3.1, yielding pCB406. Wild-type mouse embryonic fibroblasts (EFs) were transiently transfected with pCB405 and pCB406 using BioT (Bioland Scientific) according to the manufacturers protocol. Cells were lysed 48 hr after transfection in 10% glycerol, 0.5% Triton-X100, 0.15 M NaCl, 40 mM HEPES (pH 7.6) containing *"complete EDTA-free protease inhibitors"* (Roche). Clarified cell lysates were then incubated with Glutathione Superflow Agarose (Qiagen) for 3 hrs at 4°C, followed by affinity chromatography in a glutathione-agarose mini-column. Proteins retained on the column were washed 3 times in lysis buffer, and proteins were eluted in lithium-dodecyl sulfate-PAGE sample buffer, followed by SDS-4-12% PAGE. Fractionated proteins were silver-stained, and a separate set of the same samples was fractionated identically but stained with Coomassie. Specific Coomassie-stained bands were excised from the gel and proteins in gel slices were analyzed (using tryptic digestion and LC-MS/MS) at the Columbia University Protein Core Facility (New York, NY).

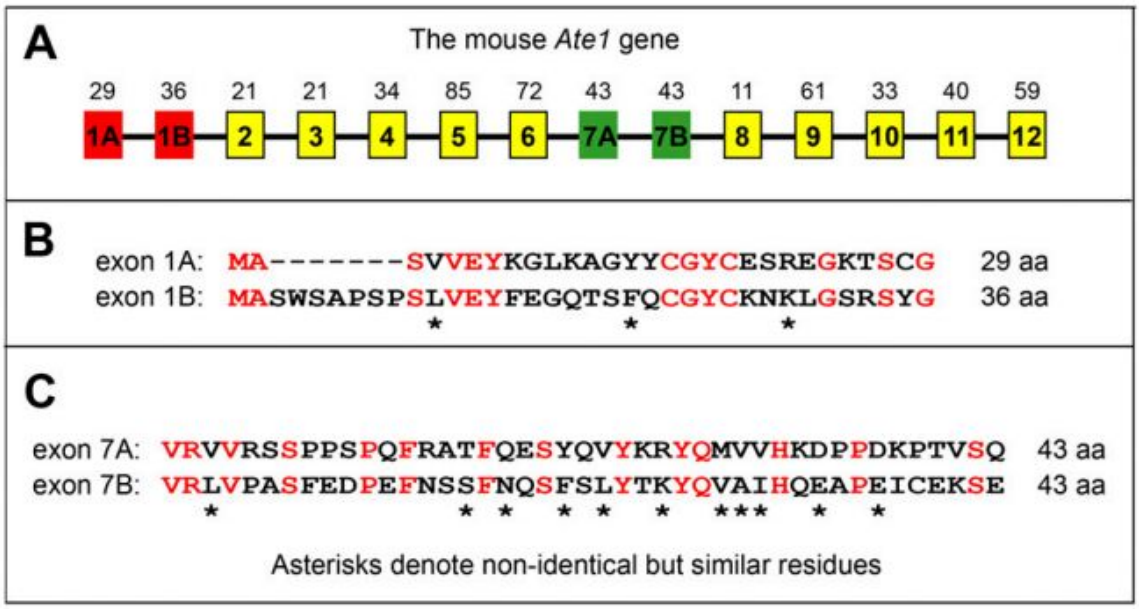


Fig. A3.S1. Mouse Ate1 arginyltransferase (R-transferase), its gene organization and its alternative exons. (*A*) *Ate1* exons, including alternative exons (1A/1B and 7A/7B), with deduced lengths of the corresponding polypeptide segments indicated on top. (*B*) The deduced amino acid sequences of the alternative mouse Ate1 exons 1A (29 residues) and 1B (36 residues), with identical residues in red and similar residues marked by asterisks. (*C*) Same as in *B*, but the 43-residue alternative exons 7A and 7B.

Fig. A3.S2. Sequelogies (sequence similarities; see Introduction and ref. (70)) among the Liat1 domains of different animals. This particularly highly conserved ~30-residue region of mouse Liat1, termed the Liat1 domain (it is highlighted in yellow in Figs. 3 and 4B), is aligned with Liat1 domains of other mammals, two non-mammalian vertebrates, and two invertebrates. At every position, a residue that is either identical or non-identical to the majority of residues at that position is either in red or in black, respectively. The numbers of last residues in each row are indicated on the right. Specific Liat1 proteins: Mouse (*Mus musculus*) (NP_941039). Rat (*Rattus norvegicus*) (NP_001103625.1). Dog (*Canis lupus*) (XP_537762.4). Rabbit (*Oryctolagus cuniculus*) (XP_002719100). Manatee (*Trichechus manatus*) (XP_004376319.1). Cow (*Bos taurus*) (DAA19029). Yak (*Bos grunniens*) (ELR61882). Dolphin (*Delphinus capensis*) (XP_004312659.1). Whale (*Balaenoptera musculus*) (XP_004267289.1). Rhinoceros (*Diceros bicornis*) (XP_004433508.1). Macaque (*Macaca mulatta*) (XP_001083490.1). Baboon (*Papio ursinus*) (XP_003912083). Bushbaby (*Otolemur garnettii*) (XP_003799182). Gibbon (*Hylobates lar lar*) (XP_003280369). Orangutan (*Pongo pygmaeus*) (XP_002826835). Gorilla (*Gorilla beringei beringei*) (XP_004058251.1). Chimpanzee (*Pan trogloditus*) (XP_001174024). Bonobo (*Pan paniscus*) (XP_003816870). Human (*Homo sapiens*) (NP_001013694.4). Dove (*Columba livia*) (EMC80592). Frog (*Xenopus tropicalis*) (XP_004911734). Sea anemone (*Nematostella vectensis*) (XP_001639507.1). Acorn worm (*Saccoglossus kowalewskii*) (XP_002738466).

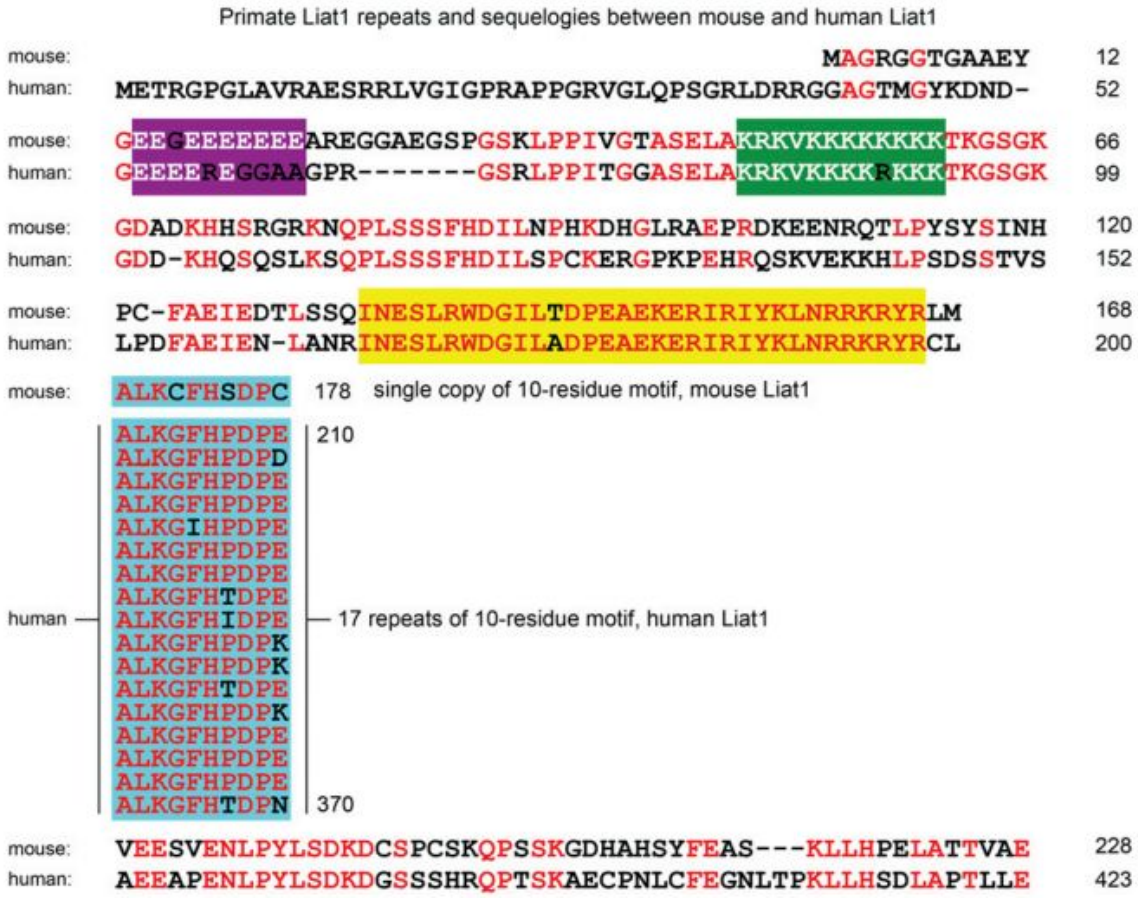


Fig. A3.S3. Sequence alignment of the full-length mouse and human Liat1 proteins. Note that significant sequelogies between mouse and human Liat1 are also present outside the nearly identical Liat1 domains (yellow rectangle). The tandem repeats in human Liat1 (they are absent in the mouse and other non-primate Liat1 proteins) are highlighted by their arrangement in a column. At every position, a residue that is either identical or non-identical to a majority-residue in this alignment is either in red or in black, respectively. The colors of the negatively charged region (purple), of the positively charged region (green), and of the Liat1 domain (yellow) (see also Figs. 4 and S2) are the same as in the corresponding regions of mouse Liat1 in Fig. 3A. The numbers of last residues in each row are indicated on the right.

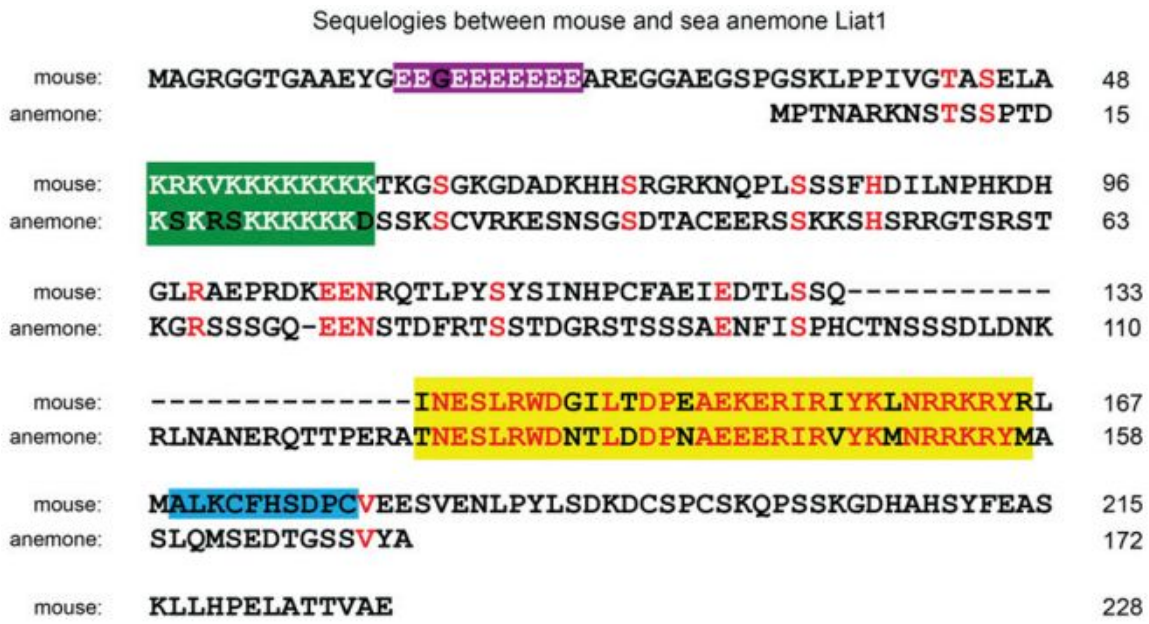


Fig. A3.S4. Sequence alignment of the full-length mouse (vertebrate) and sea anemone (invertebrate) Liat1 proteins. Note the near-absence of sequelogies outside the Liat1 domain (in yellow) and the positively charged region (in green). A 10-residue sequence of mouse Liat1 that becomes tandemly repeated in Liat1 of primates (see Figs. 4 and S3) is in blue. The numbers of last residues in each row are indicated on the right.

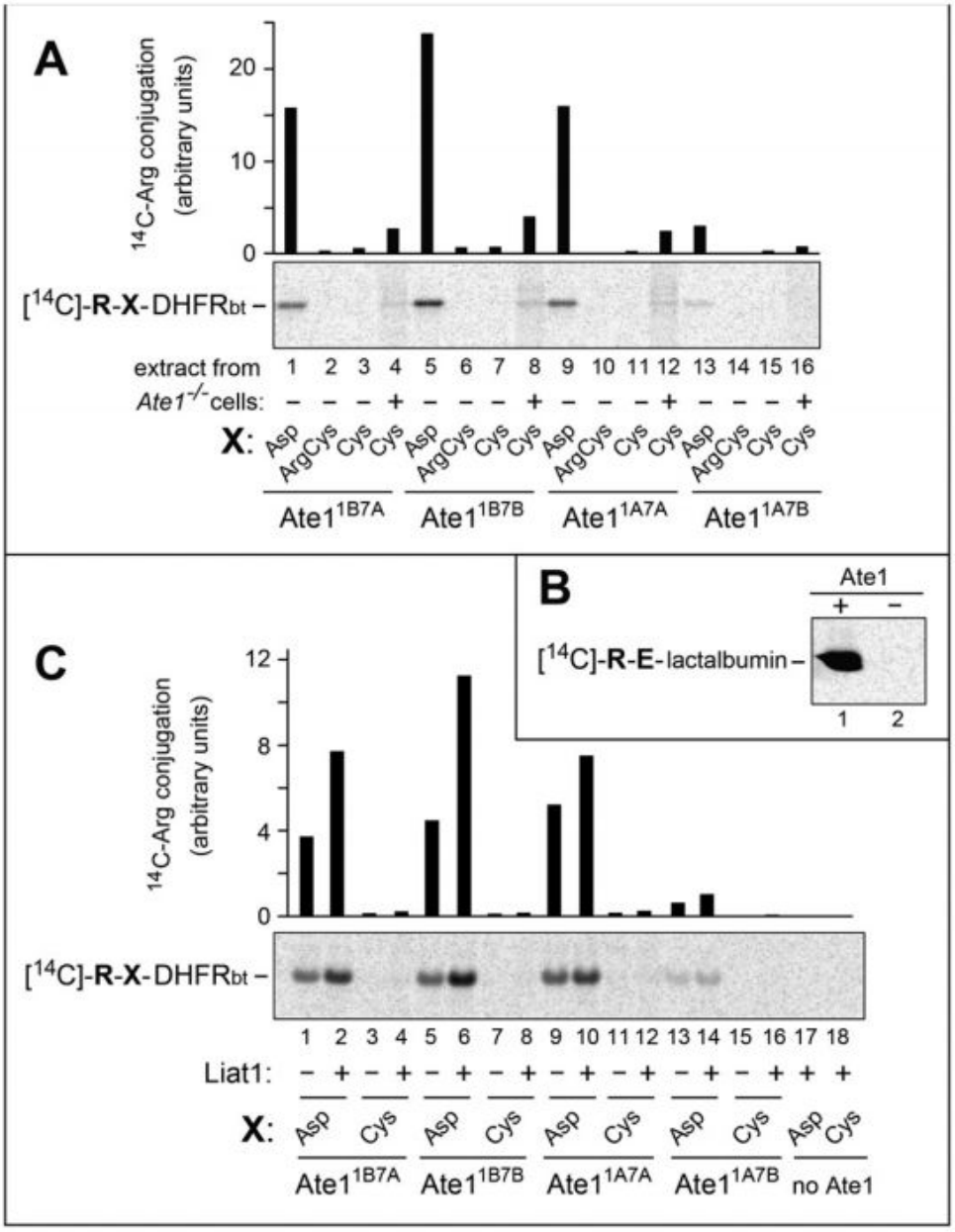
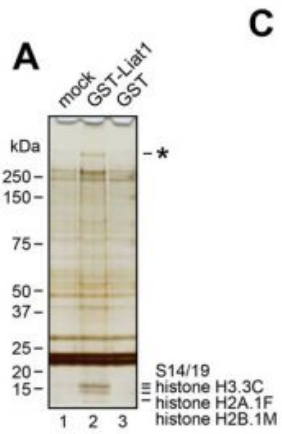
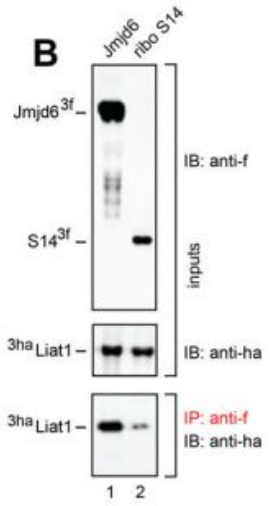


Fig. A3.S5. Purified Liat1 enhances the Ate1-mediated N-terminal arginylation *in vitro*. (*A*) Lane 1, purified mouse Ate1^{1B7A} R-transferase was incubated with Asp-DHFRbt, a test protein, and with other components of a ^{14}C-Arg-based Nt-arginylation system (see the main text and *SI Experimental Procedures*) followed by SDS-PAGE and autoradiography. Lane 2, same as in lane 1 but with Arg-Asp-DHFRbt. Same as in lane 1 but with Cys-DHFRbt. Lane 4, same as in lane 3 but in the presence of Ate1-lacking extract from *Ate1*$^{-/-}$ EF cells. Lanes 5-8, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1B7B} R-transferase. Lanes 9-12, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1A7A} R-transferase. Lanes 13-16, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1A7B} R-transferase. Relative levels of Nt-arginylation of X-DHFRbt are plotted above the autoradiogram. (*B*) Lane 1, ^{14}C-Arg-

based Nt-arginylation of lactalbumin (a model Nt-arginylation substrate bearing N-terminal Glu; see Fig. 1A). Lane 2, same as in lane 1 but in the absence of added Ate1^{1B7A} R-transferase. (*C*) Lane 1, same as in *A*, lane 1 (an independent assay). Lane 2, same as in lane 1 but in the presence of a ~4-fold molar excess (relative to Ate1) of the purified mouse Liat protein. Lanes 3 and 4, same as in lanes 1 and 2, respectively, but with Cys-DHFRbt. Lanes 5-8, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1B7B} R-transferase. Lanes 9-12, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1A7A} R-transferase. Lanes 13-16, same as in lanes 1-4, respectively, but with purified mouse Ate1^{1A7B} R-transferase. Lane 17, same as in lane 2 but no added Ate1 (in the presence of Liat1). Lane 18, same as in lane 17, but with Cys-DHFRbt.





Identified (in red) and potential mouse protein ligands of mouse Liat1

C

Liat1 ligand	# of hits	Accession #	Data
Ate1 R-transferase	see text	NP_038827.2 NP_001025066.1	Y2H; co-IP
mito ribosomal protein S33	9	Q9D2R8	Y2H
MycBP2	8	Q7TPH6	Y2H
Larp7	4	Q05CL8	Y2H
Unknown protein 1	2	NP_7664.24.2	Y2H
RING finger 103	2	Q9R1W3	Y2H
Cop9 signalosome subunit 5	2	O35864	Y2H
DEAD box 24	2	Q9ESV0	Y2H
Srp19	2	Q9D7A6	Y2H
Asparaginase-like 1	2	Q8C0M9	Y2H
Ub mofifier conjug. enzyme 1	2	Q9CR09	Y2H
mito ribosomal protein S18C	1	Q8R2L5	Y2H
Sec63	1	Q8VHE0	Y2H
DNAJc14	1	Q921R4	Y2H
RNA polymerase 2 subunit 3	1	P97760	Y2H
RNA polymerase 2 subunit 7	1	P62488	Y2H
Zing finger protein 60	1	Q8K0D9	Y2H
Zing finger protein 68	1	Q8VEM6	Y2H
Ncor1	1	Q60974	Y2H
DNA-binding protein SATB1	1	Q60611	Y2H
Cyclin C	1	Q62447	Y2H
Herc2	1	Q4U2R1	Y2H
Otub1	1	Q7TQI3	Y2H
Usp7	1	Q6A4J8	Y2H
Unknown protein 2	1	XR_108069.1	Y2H
Ccd2a	1	Q8CFW7	Y2H
Trmt11	1	Q9CWH5	Y2H
Lipoic acid synthetase	1	Q99M04	Y2H
Calm1	1	P62204	Y2H
Havcr2	1	Q8VIM0	Y2H
Schwannomin-inter. prot. 1	1	Q3TI53	Y2H
Exosome component 1	1	Q9DAA6	Y2H
Spink8	1	Q09TK9	Y2H
HBS1-like	1	Q3TGM7	Y2H
Prp	1	P04925	Y2H
Pgam1	1	Q9DBJ1	Y2H
Id4 dom-neg helix-loop-helix	1	P41139	Y2H
PLC-like 1	1	Q3USB7	Y2H
Cathepsin S	1	O70370	Y2H
Itm2a	1	Q61500	Y2H
Itm2b	1	Q89051	Y2H
ribosomal protein S14	–	P62264	MS
ribosomal protein S19	–	Q9CZX8	MS
histone H3.3C	–	P02301	MS
histone H2A.1F	–	Q8CGP5	MS
histone H2B.1M	–	P10854	MS
Jmjd6	–	Q6NYC1	Weimann et al. (Y2H; see SI)

Fig. A3.S6. Putative Liat1-binding proteins other than Ate1.
(*A*) GST-Liat1 pulldown with an extract from mouse EF cells expressing N-terminally tagged GST-Liat1, which was selectively retained, together with associated proteins, using glutathione-beads. Proteins were detected by silver staining after. Regions of a Coomassie-stained counterpart of the silver-stained gel shown here (lane 2) containing proteins co-isolated with GST-Liat1 were cut out (see *SI Experimental Procedures*) and analyzed using trypsin digestion and mass spectrometry (MS-MS). An indicated Liat1-binding large protein (>>250 kDa) could not be identified. (*B*) Coimmunoprecipitation of mouse ^{3ha}Liat1 (lane 1) with mouse Jmjd6^{3f} and the ribosomal protein S14^{3f} (see the main text). IP (in red), immunoprecipitation (with anti-flag). IB, immunoblotting (with anti-flag and anti-ha). (*C*) Putative Liat1-binding proteins (besides the independently characterized Ate1, in red) that have been identified through a separate Y2H screen using mouse Liat1 and mouse brain cDNA library. Five Liat1 ligands identified using GST-pulldowns (panel *A*) are cited last (before Jmjd6).

Table S1. Plasmids used in this study.

Plasmid	Description	Source or Reference
pcDNA3.0-Neo	Amp^R; Neo^R; Expression vector for cloning a gene of interest.	Invitrogen
pcDNA3.1-Hyg	Amp^R; Hyg^R; Expression vector for cloning a gene of interest.	Invitrogen
pACT2	Amp^R; *LEU2* selectable marker in yeast. Yeast two-hybrid expression vector with P_{ADH1} promoter. Produces ha-tagged Gal4 Activation Domain (AD) fusion.	Clontech
pAS2	Amp^R; (also called pAS1-CYH2); *CYH* selectable marker in yeast. Yeast two-hybrid expression vector with P_{ADH1} promoter. Produces ha-tagged Gal4 DNA binding domain (DBD) fusion.	Clontech
pGEX-2t	Amp^R; Bacterial expression vector for expression of fusion proteins containing an N-terminal glutathione S-transferase (GST) moiety.	GE Healthcare Life Sciences
pH10UE	Amp^R; Bacterial expression vector for expression of fusion proteins containing an N-terminal His_{10}-ubiquitin moiety.	(1)
pA6	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to mouse ^{3ha}Liat1 under the control of the yeast P_{ADH1} promoter.	This study
pCB132	Amp^R; pAS2-based plasmid encoding Gal4-DNA-binding domain (DBD) fused to mouse ^{ha}Ate1^{1B7A} under the control of the yeast P_{ADH1} promoter.	This study
pCB135	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to ^{ha}Ate1^{1B7A} under the control of the yeast P_{ADH1} promoter.	This study
pCB141	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to Cdc42 under the control of the yeast P_{ADH1} promoter.	This study

pCB142	Amp^R; pAS2-based plasmid encoding Gal4-DNA-binding domain (DBD) fused to WASP under the control of the yeast P_{ADH1} promoter.	This study
pCB164	Amp^R; pAS2-based plasmid encoding Gal4-DNA-binding domain (DBD) fused to $^{ha}ATE1^{1B7B}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB165	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}ATE1^{1B7B}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB166	Amp^R; pAS2-based plasmid encoding Gal4-DNA-binding (DBD) domain fused to $^{ha}ATE1^{1A7A}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB167	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}ATE1^{1A7A}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB179	Amp^R; pcDNA3.1-based plasmid encoding mouse ^{3ha}Liat1 under the control of the mammalian P_{CMV} promoter	This study
pCB239	Amp^R; pAS2-based plasmid encoding Gal4-DNA-binding (DBD) domain fused to ^{3ha}Liat1 under the control of the yeast P_{ADH1} promoter.	This study
pCB245	Amp^R; pcDNA-Neo-based plasmid encoding mouse $Ate1^{1B7A}$ (untagged) under the control of the mammalian P_{CMV} promoter.	This study
pCB246	Amp^R; pcDNA-Neo-based plasmid encoding mouse $Ate1^{1B7B}$ (untagged) under the control of the mammalian P_{CMV} promoter.	This study
pCB247	Amp^R; pcDNA-Neo-based plasmid encoding mouse $Ate1^{1A7A}$ (untagged) under the control of the mammalian P_{CMV} promoter.	This study

pCB248	Amp^R; pcDNANeo-based plasmid encoding mouse $Ate1^{1A7B}$ (untagged) under the control of the mammalian P_{CMV} promoter.	This study
pCB403	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and ^{3ha}Liat1.	This study
pCB404	Amp^R; Bacterial expression vector for expression of GST-Liat1.	This study
pCB405	Amp^R; pcDNA3.1-based plasmid encoding GST-Liat1 under the control of the mammalian P_{CMV} promoter.	This study
pCB406	Amp^R; pcDNA3.1-based plasmid encoding GST under the control of the mammalian P_{CMV} promoter.	This study
pCB407	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and $Ate1^{1B7A}$.	This study
pCB408	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and $Ate1^{1B7B}$.	This study
pCB409	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and $Ate1^{1A7A}$.	This study
pCB410	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and $Ate1^{1A7B}$.	This study
pCB432	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(1-228)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB433	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(61-228)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB434	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(113-228)}$ under the control of the yeast P_{ADH1} promoter.	This study

pCB435	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(135-228)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB436	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{3ha}Liat1^{(135-165)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB437	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{HA}mLiat1^{(1-165)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB438	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(\cdot\ 135-165)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB439	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(95-228)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCB447	Amp^R; pcDNA-Neo-based plasmid encoding 40S ribosomal protein $S14^{3f}$ under the control of the mammalian P_{CMV} promoter.	This study
pCB458	Amp^R; pcDNA-Neo-based plasmid encoding $Jmjd6^{3f}$ under the control of the mammalian P_{CMV} promoter.	This study
pCER006	Amp^R; pH10UE-based vector used for the bacterial expression of a fusion between His_{10}-ubiquitin and mouse Liat1.	This study
pCER019	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(\cdot\ 1-117)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCER020	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to $^{ha}Liat1^{(\cdot\ 1-124)}$ under the control of the yeast P_{ADH1} promoter.	This study
pCER021	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to	This study

	^{ha}Liat1$^{(113\text{-}165)}$ under the control of the yeast P_{ADH1} promoter.	
pCER022	Amp^R; pACT2-based plasmid encoding Gal4-activation domain (AD) fused to ^{ha}Liat1$^{(\bullet\ 153\text{-}228)}$ under the control of the yeast P_{ADH1} promoter.	This study

TABLE S1 REFERENCE

1. Catanzariti A-M, Soboleva TA, Jans DA, Board PG, & Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. *Protein Sci.* 13:1331-1339.

SI REFERENCES

1. Varshavsky A (2011) The N-end rule pathway and regulation by proteolysis. Prot. Sci. 20:1298-1345.
2. Hwang C-S, Shemorry A, & Varshavsky A (2010) N-terminal acetylation of cellular proteins creates specific degradation signals. Science 327:973-977.
3. Eisele F & Wolf DH (2008) Degradation of misfolded proteins in the cytoplasm by the ubiquitin ligase Ubr1. FEBS Lett. 582:4143-4146.
4. Heck JW, Cheung SK, & Hampton RY (2010) Cytoplasmic protein quality control degradation mediated by parallel actions of the E3 ubiquitin ligases Ubr1 and San1. Proc. Natl. Acad. Sci. USA 107:1106-1111.
5. Tasaki TS, Sriram SM, Park KS, & Kwon YT (2012) The N-end rule pathway. Annu. Rev. Biochem. 81:261-289.
6. Graciet E & Wellmer F (2010) The plant N-end rule pathway: structure and functions. Trends Plant Sci. 15:447-453.

7. Dougan DA, Micevski D, & Truscott KN (2011) The N-end rule pathway: from recognition by N-recognins to destruction by AAA+ proteases. Biochim. Biophys. Acta 1823:83-91.

8. Mogk A, Schmidt R, & Bukau B (2007) The N-end rule pathway of regulated proteolysis: prokaryotic and eukaryotic strategies. Trends Cell Biol. 17:165-172.

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

10. Hwang C-S, Shemorry A, & Varshavsky A (2010) The N-end rule pathway is mediated by a complex of the RING-type Ubr1 and HECT-type Ufd4 ubiquitin ligases. Nat. Cell Biol. 12:1177-1185.

11. Hwang C-S, Shemorry A, & Varshavsky A (2009) Two proteolytic pathways regulate DNA repair by co-targeting the Mgt1 alkylguanine transferase. Proc. Natl. Acad. Sci. USA 106:2142-2147.

12. Hwang C-S & Varshavsky A (2008) Regulation of peptide import through phosphorylation of Ubr1, the ubiquitin ligase of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 105:19188-19193.

13. Xia Z, et al. (2008) Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. J. Biol. Chem. 283:24011-24028.

14. Hu R-G, Wang H, Xia Z, & Varshavsky A (2008) The N-end rule pathway is a sensor of heme. Proc. Natl. Acad. Sci. USA 105:76-81.

15. Hu R-G, et al. (2005) The N-end rule pathway as a nitric oxide sensor controlling the levels of multiple regulators. Nature 437:981-986.

16. Hwang C-S, et al. (2011) Ubiquitin ligases of the N-end rule pathway: assessment of mutations in UBR1 that cause the Johanson-Blizzard syndrome. PLoS One 6:e24925.

17. Wang H, Piatkov KI, Brower CS, & Varshavsky A (2009) Glutamine-specific N-terminal amidase, a component of the N-end rule pathway. Mol. Cell 34:686-695.

18. Brower CS & Varshavsky A (2009) Ablation of arginylation in the mouse N-end rule pathway: loss of fat, higher metabolic rate, damaged spermatogenesis, and neurological perturbations. PLoS ONE 4:e7757.

19. Zenker M, et al. (2005) Deficiency of UBR1, a ubiquitin ligase of the N-end rule pathway, causes pancreatic dysfunction, malformations and mental retardation (Johanson-Blizzard syndrome). Nat. Genet. 37:1345-1350.

20. Prasad R, Kawaguchi S, & Ng DTW (2010) A nucleus-based quality control mechanism for cytosolic proteins. Mol. Biol. Cell 21:2117-2127.

21. Kurosaka S, et al. (2010) Arginylation-dependent neural crest cell migration is essential for mouse development. PLoS Genet. 6:e1000878.

22. Zhang F, Saha S, Shabalina SA, & Kashina A (2010) Differential arginylation of actin isoforms is regulated by coding sequence-dependent degradation. Science 329(1534-1537):1534-1537.

23. Lee MJ, et al. (2005) RGS4 and RGS5 are in vivo substrates of the N-end rule pathway. Proc. Natl. Acad. Sci. USA 102:15030-15035.

24. Piatkov KI, Brower CS, & Varshavsky A (2012) The N-end rule pathway counteracts cell death by destroying proapoptotic protein fragments. Proc. Natl. Acad. Sci. USA 109:E1839-E1847.

25. Piatkov KI, Colnaghi L, Bekes M, Varshavsky A, & Huang TT (2012) The auto-generated fragment of the Usp1 deubiquitylase is a physiological substrate of the N-end rule pathway. Mol. Cell 48:926-933.

26. Kwon YT, et al. (2002) An essential role of N-terminal arginylation in cardiovascular development. Science 297:96-99.

27. Lee MJ, et al. (2012) Characterization of arginylation branch of the N-end rule pathway in G-protein-mediated proliferation and signaling of cardiomyocytes. J. Biol. Chem. 287:24043-24052.

28. Choi WS, et al. (2010) Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. Nat. Struct. Mol. Biol. 17:1175-1181.

29. Matta-Camacho E, Kozlov G, Li FF, & Gehring K (2010) Structural basis of substrate recognition and specificity in the N-end rule pathway. Nat. Struct. Mol. Biol. 17:1182-1188.

30. Licausi F, et al. (2011) Oxygen sensing in plants is mediated by an N-end rule pathway for protein destabilization. Nature 479:419-422.

31. Sasidharan R & Mustroph A (2011) Plant oxygen sensing is mediated by the N-end rule pathway: a milestone in plant anaerobiosis. Plant Cell 23:4173-4183.

32. Gibbs DJ, et al. (2011) Homeostatic response to hypoxia is regulated by the N-end rule pathway in plants. Nature 479:415-418.

33. Rao H, Uhlmann F, Nasmyth K, & Varshavsky A (2001) Degradation of a cohesin subunit by the N-end rule pathway is essential for chromosome stability. Nature 410:955-960.

34. Kim ST, et al. (2013) The N-end rule proteolytic system in autophagy. Autophagy 9:1100-1103.

35. Zhang G, Lin RK, Kwon YT, & Li YP (2013) Signaling mechanism of tumor cell-induced up-regulation of E3 ubiquitin ligase UBR2. FASEB J. 27:2893-2901.

36. Tasaki T, et al. (2013) UBR box N-recognin-4 (UBR4), an N-recognin of the N-end rule pathway, and its role in yolk sac vascular development and autophagy. Proc. Natl. Acad. Sci. USA 110:3800-3805.

37. Fujiwara H, Tanaka N, Yamashita I, & Kitamura K (2013) Essential role of Ubr11, but not Ubr1, as an N-end rule ubiquitin ligase in Schizosaccharomyces pombe. Yeast 30:1-11.

38. Kitamura K & Fujiwara H (2013) The type-2 N-end rule peptide recognition activity of Ubr11 ubiquitin ligase is required for the expression of peptide transporters. FEBS Lett. 587:214-219.

39. An JY, et al. (2010) UBR2 mediates transcriptional silencing during spermatogenesis via histone ubiquitination. Proc. Natl. Acad. Sci. USA 107:1912-1917.

40. An JY, et al. (2012) UBR2 of the N-end rule pathway is required for chromosome stability via histone ubiquitylation in spermatocytes and somatic cells. PLoS One 7:e37414.

41. Sultana R, Theodoraki MA, & Caplan AJ (2012) UBR1 promotes protein kinase quality control and sensitizes cells to Hsp90 inhibition. Exp. Cell Res. 18:53-60.

42. Lee P, Sowa ME, Gygi SP, & Harper JW (2011) Alternative ubiquitin activation/conjugation cascades interact with N-recognin ubiquitin ligases of the

N-end rule pathway to control degradation of RGS proteins. Mol. Cell 43:392-405.

43. Román-Hernández G, Hou JY, Grant RA, Sauer RT, & Baker TA (2011) The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. Mol. Cell 43:217-228.

44. Yang F, et al. (2010) The ubiquitin ligase Ubr2, a recognition E3 component of the N-end rule pathway, stabilizes Tex.19.1 during spermatogenesis. PLoS One 5:e14017.

45. Ninnis RL, Spall SK, Talbo GH, Truscott KN, & Dougan DA (2009) Modification of PATase by L/F-transferase generates a ClpS-dependent N-end rule substrate in Escherichia coli. EMBO J. 28:1732-1744.

46. Schmidt R, Zahn R, Bukau B, & Mogk A (2009) ClpS is the recognition component for Escherichia coli substrates of the N-end rule degradation pathway. Mol. Microbiol. 72:506-517.

47. Holman TJ, et al. (2009) The N-end rule pathway promotes seed germination and establishment through removal of ABA sensitivity in Arabidopsis. Proc. Natl. Acad. Sci. USA 106:4549-4554.

48. Cai H, Hauser M, Naider F, & Becker JM (2007) Differential regulation and substrate preferences in two peptide transporters of Saccharomyces cerevisiae. Eukaryot. Cell 6:1805-1813.

49. Schnupf P, Zhou J, Varshavsky A, & D.A. P (2007) Listeriolysin O secreted by Listeria monocytogenes into the host cell cytosol is degraded by the N-end rule pathway. Inflammation & Immunity 75:5135-5147.

50. Hu R-G, et al. (2006) Arginyl-transferase, its specificity, putative substrates, bidirectional promoter, and splicing-derived isoforms. J. Biol. Chem.:32559-32573.

51. Graciet E, et al. (2006) Aminoacyl-transferases and the N-end rule pathway of prokaryotic/eukaryotic specificity in a human pathogen. Proc. Natl. Acad. Sci. USA 103:3078-3083.

52. Kurosaka S, et al. (2012) Arginylation regulates myofibrils to maintain heart function and prevent dilated cardiomyopathy. J. Mol. Cell. Cardiol. 53:333-341.

53. Wang J, et al. (2011) Arginyltransferase is an ATP-independent self-regulating enzyme that forms distinct functional complexes in vivo. Chem. Biol. 18:121-130.

54. Decca MB, et al. (2007) Post-translational arginylation of calreticulin: a new isospecies of calreticulin component of stress granules. J. Biol. Chem. 282:8337-8345.

55. Tobias JW, Shrader TE, Rocap G, & Varshavsky A (1991) The N-end rule in bacteria. Science 254:1374-1377.

56. Boso G, Tasaki T, Kwon YT, & Somia NV (2013) The N-end rule and retroviral infection: no effect on integrase. Virology J. 10:233.

57. Saha S & Kashina A (2011) Posttranslational arginylation as a global biological regulator. Dev. Biol. 358:1-8.

58. Belzil C, et al. (2013) A Ca2+-dependent mechanism of neuronal survival mediated by the microtubule-associated protein p600. J. Biol. Chem. 288:24452-24464.

59. Yamano K & Youle RJ (2013) PINK1 is degraded through the N-end rule pathway. Autophagy 9:1758-1769.

60. Kim H-K, et al. (2014) The N-terminal methionine of cellular proteins as a degradation signal. Cell 156:158-169.

61. Gibbs DJ, et al. (2014) Nitric oxide sensing in plants is mediated by proteolytic control of group VII ERF transcription factors. Mol. Cell 53:369-379.

62. Gibbs DJ, Bacardit J, Bachmair A, & Holdsworth MJ (2014) The eukaryotic N-end rule pathway: conserved mechanisms and diverse functions. Trends Cell Biol. 24:603-611.

63. Licausi F, Pucciariello C, & Perata P (2013) New role for an old rule: N-end rule-mediated degradation of ethylene responsive factor proteins governs low oxygen response in plants. J. Integr. Plant Biol. 55:31-39.

64. Weits DA, et al. (2014) Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. Nat Commun. 5:3425.

65. Arnesen T, et al. (2009) Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast to humans. Proc. Natl. Acad. Sci. USA 106:8157-8162.

66. Starheim KK, Gevaert K, & Arnesen T (2012) Protein N-terminal acettyltransferases: when the start matters. Trends Biochem. Sci. 37:152-161.

67. Van Damme P, et al. (2012) N-terminal acetylome analyses and functional insights of the N-terminal acetyltransferase NatB. Proc. Natl. Acad. Sci. USA 109:12449-12454.

68. Shemorry A, Hwang C-S, & Varshavsky A (2013) Control of protein quality and stoichiometries by N-terminal acetylation and the N-end rule pathway. Mol. Cell 50:540-551.

69. Brower CS, Veiga L, Jones RH, & Varshavsky A (2010) Mouse Dfa Is a Repressor of TATA-Box Promoters and Interacts with the Abt1 Activator of Basal Transcription. J. Biol. Chem. 285:17218-17234.

70. Varshavsky A (2004) Spalog and sequelog: neutral terms for spatial and sequence similarity. Curr. Biol. 14:R181-R183.

71. Catanzariti A-M, Soboleva TA, Jans DA, Board PG, & Baker RT (2004) An efficient system for high-level expression and easy purification of authentic recombinant proteins. Protein Sci. 13:1331-133