THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 281, NO. 43, pp. 32559–32573, October 27, 2006

Arginyltransferase, Its Specificity, Putative Substrates, Bidirectional Promoter, and Splicing-derived Isoforms*[S]

Received for publication, May 8, 2006, and in revised form, August 15, 2006 Published, JBC Papers in Press, August 30, 2006, DOI 10.1074/jbc.M604355200

Rong-Gui Hu[‡], Christopher S. Brower[‡], Haiqing Wang[‡], Ilia V. Davydov[§], Jun Sheng[‡], Jianmin Zhou[‡], Yong Tae Kwon[¶], and Alexander Varshavsky[‡1]

From the [‡]Division of Biology, California Institute of Technology, Pasadena, California 91125, [§]Meso Scale Discovery, Gaithersburg, Maryland 20877, and [¶]Center for Pharmacogenetics and Department of Pharmaceutical Sciences, School of Pharmacy, University of Pittsburgh, Pittsburgh, Pennsylvania 15261

Substrates of the N-end rule pathway include proteins with destabilizing N-terminal residues. Three of them, Asp, Glu, and (oxidized) Cys, function through their conjugation to Arg, one of destabilizing N-terminal residues that are recognized directly by the pathway's ubiquitin ligases. The conjugation of Arg is mediated by arginyltransferase, encoded by *ATE1*. Through its regulated degradation of specific proteins, the arginylation branch of the N-end rule pathway mediates, in particular, the cardiovascular development, the fidelity of chromosome segregation, and the control of signaling by nitric oxide. We show that mouse *ATE1* specifies at least six mRNA isoforms, which are produced through alternative splicing, encode enzymatically active arginyltransferases, and are expressed at varying levels in mouse tissues. We also show that the *ATE1* promoter is bidirectional, mediating the expression of both *ATE1* and an oppositely oriented, previously uncharacterized gene. In addition, we identified GRP78 (glucose-regulated protein 78) and protein-disulfide isomerase as putative physiological substrates of arginyltransferase. Purified isoforms of arginyltransferase that contain the alternative first exons differentially arginylate these proteins in extract from *ATE1*$^{-/-}$ embryos, suggesting that specific isoforms may have distinct functions. Although the N-end rule pathway is apparently confined to the cytosol and the nucleus, and although GRP78 and protein-disulfide isomerase are located largely in the endoplasmic reticulum, recent evidence suggests that these proteins are also present in the cytosol and other compartments *in vivo*, where they may become N-end rule substrates.

A protein substrate of the ubiquitin (Ub)[2]-proteasome system, which controls the levels of many intracellular proteins, is conjugated to Ub through the action of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzyme (E2), and ubiquitin-protein ligase (E3) (1–7). The term "Ub ligase" denotes either an E2-E3 holoenzyme or its E3 component. The selectivity of ubiquitylation is mediated largely by E3, which recognizes a degradation signal (degron) of a substrate (1, 3, 8–12). The E3 Ub ligases are an exceptionally large family, with more than 500 distinct E3s in a mammal (5, 12–14). A ubiquitylated protein bears a covalently linked poly-Ub chain and is targeted for processive degradation by the 26 S proteasome (15–19). Ub has other functions as well, including nonproteolytic ones (20, 21).

An essential determinant of one class of degrons, called N-degrons, is a destabilizing N-terminal residue of a substrate. The set of destabilizing residues in a given cell type yields a rule, called the N-end rule, which relates the *in vivo* half-life of a protein to the identity of its N-terminal residue (1, 22–26). In eukaryotes, the N-degron consists of three determinants: a destabilizing N-terminal residue of a protein substrate, its internal Lys residue(s) (the site of formation of a poly-Ub chain), and a conformationally flexible region(s) in the vicinity of these determinants that is required for the substrate's ubiquitylation and/or degradation (8, 23, 27–29).

The N-end rule has a hierarchic structure (Fig. 1*A*). In eukaryotes, N-terminal Asn and Gln are tertiary destabilizing residues (denoted as Nd^t) in that they function through their enzymatic deamidation (30, 31), to yield the secondary destabilizing N-terminal residues Asp and Glu (denoted as Nd^s). The destabilizing activity of the N-terminal Asp and Glu requires their conjugation, by *ATE1*-encoded isoforms of Arg-tRNA-protein transferase (arginyltransferase or R-transferase), to Arg, one of the primary destabilizing residues (denoted as Nd^p) (26, 32–37). The latter N-terminal residues are recognized by E3 Ub ligases of the N-end rule pathway, called N-recognins (23–25, 38, 39) (Fig. 1*A*). The N-end rule pathway of the yeast *Saccharomyces cerevisiae* is mediated by a single N-recognin, UBR1 (40, 41), while at least four N-recognins, including UBR1, mediate this pathway in mammals (24, 25, 38, 39, 42). In mammals and other eukaryotes that produce nitric oxide (NO), the set of arginylated residues contains not only Asp and Glu but also N-terminal Cys (C) (43), which is arginylated after its oxi-

* This work was supported by National Institutes of Health Grants GM31530, DK39520 (to A. V.), GM069482, and GM074000 (to Y. T. K.) and by grants from the Ellison Medical Foundation (to A. V.) and the American Heart Association (to Y. T. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2.

[1] To whom correspondence should be addressed. Tel.: 626-395-3785; Fax: 626-440-9821; E-mail: avarsh@caltech.edu.

[2] The abbreviations used are: Ub, ubiquitin; R-transferase, Arg-tRNA-protein transferase or arginyltransferase; Nd^p, primary destabilizing N-terminal residue; Nd^s, secondary destabilizing N-terminal residue; E3, ubiquitin-protein ligase; PDI, protein-disulfide isomerase; ER, endoplasmic reticulum; RT, reverse transcription; EF, embryonic fibroblast; MS/MS, tandem mass spectrometry; EST, expressed sequence tag; TR-PDI, translocon-resident protein-disulfide isomerase complex; βgal, β-galactosidase.

Supplemental Material can be found at: http://www.jbc.org/cgi/content/full/M604355200/DC1

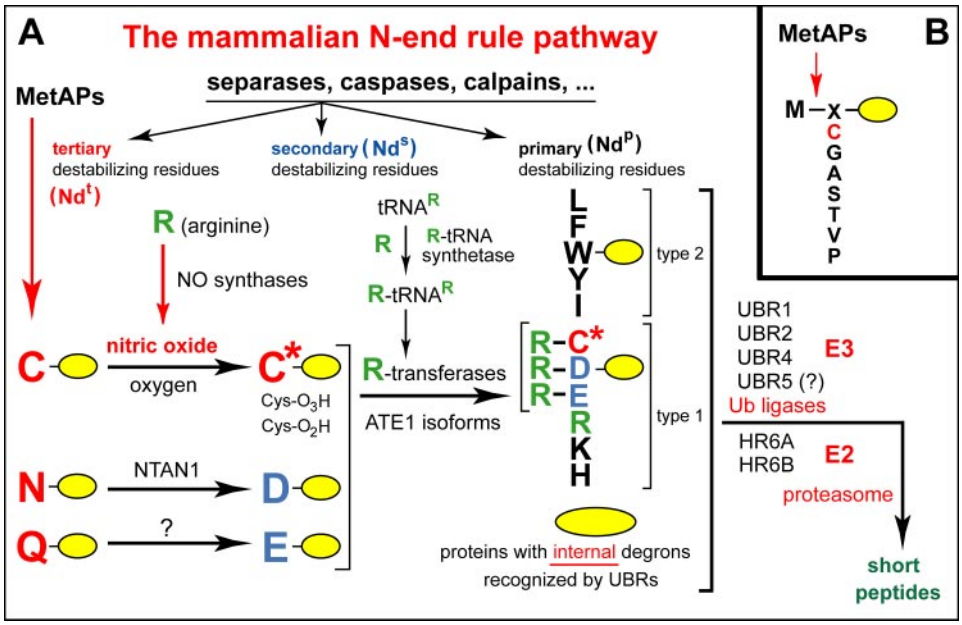

FIGURE 1. **The mammalian N-end rule pathway.** *A*, the pathway. N-terminal residues are indicated by single-letter abbreviations for amino acids. The *yellow ovals* denote the rest of a protein substrate. *MetAPs*, methionine aminopeptidases. *C** denotes oxidized N-terminal Cys, either Cys-sulfinic acid ($CysO_2(H)$) or Cys-sulfonic acid ($CysO_3(H)$), produced in reactions mediated by nitric oxide (NO) and oxygen (O_2) or its derivatives, with subsequent arginylation of oxidized Cys by the *ATE1*-encoded isoforms of Arg-tRNA-protein transferase (R-transferase) (26). (NO is produced from Arg by NO synthases.) The type 1 and type 2 primary destabilizing N-terminal residues are recognized by multiple E3 ubiquitin ligases (N-recognins) of the N-end rule pathway, including UBR1 and UBR2. Through their other substrate-binding sites, these E3s also recognize internal (non-N-terminal) degrons in other substrates of the N-end rule pathway, denoted by a *larger oval*. The marks of omission after "calpains" refer to the expectation that physiologically relevant N-degrons can also be produced by intracellular proteases additional to those cited. *B*, MetAPs remove Met from the N terminus of a polypeptide if the residue at position 2 belongs to the set of residues shown (23, 87).

dation (36). The *in vivo* oxidation of N-terminal Cys requires NO, as well as oxygen (O_2) or its derivatives (Fig. 1*A*) (26, 37).

Although prokaryotes lack Ub conjugation and Ub itself, they contain (Ub-independent) N-end rule pathways (44–47). The sets of destabilizing residues in prokaryotic N-end rules contain secondary destabilizing (Nd^s) residues as well. Their identities (Arg and Lys in *Escherichia coli*; Arg, Lys, Asp, and Glu in some other prokaryotes) are either overlapping with, or distinct from, the eukaryotic Nd^s residues (47). The activity of Nd^s residues in prokaryotes requires their conjugation to Leu (an Nd^P residue) by Leu-tRNA-protein transferases (L-transferases), in contrast to the conjugation of Arg (an Nd^P residue in eukaryotes) to N-terminal Asp, Glu, or (oxidized) Cys (Fig. 1*A*) (44, 47). Prokaryotic L-transferases are of two distinct types, differing in both amino acid sequence and substrate specificity. Remarkably, L-transferases of one class, encoded by *bpt* genes, are sequelogs of *ATE1*-encoded eukaryotic R-transferases, despite the fact that Bpt (prokaryotic) aminoacyltransferases conjugate Leu, rather than Arg, to the N termini of cognate substrates (47). (In this terminology, "sequelog" and "spalog" denote, respectively, a sequence that is similar, to a specified extent, to another sequence and a three-dimensional structure that is similar, to a specified extent, to another three-dimensional structure (48). Sequelog and spalog are mnemonically helpful, single word terms whose rigor-conferring advantage is their *evolutionary neutrality*. The sequelog terminology conveys the fact of sequence similarity (sequelogy) without evolutionary or functional connotations, in contrast to interpretation-laden terms such as homolog, ortholog, and paralog. The latter terms are compatible with the sequelog/spalog terminology and can be employed to convey understanding about functions and common descent, if this information is actually available (48).)

The functions of the N-end rule pathway include the regulation of import of short peptides (through the degradation, modulated by peptides, of the import's repressor) (41, 49), the fidelity of chromosome segregation (through the degradation of a conditionally produced cohesin fragment) (50), the regulation of apoptosis (through the degradation of a caspase-processed inhibitor of apoptosis) (51, 52), the regulation of meiosis (24), the leaf senescence in plants (53), as well as neurogenesis and cardiovascular development in mammals (26, 36, 37, 39). Mutations in human UBR1, one of several functionally overlapping N-recognins (Fig. 1*A*) (25, 39), are the cause of Johanson-Blizzard syndrome, which includes mental retardation, physical malformations, and severe pancreatitis (54). The abnormalities of $UBR1^{-/-}$ mice (38) include pancreatic insufficiency (54), a less severe version of the defect in human Johanson-Blizzard syndrome ($UBR1^{-/-}$) patients.

The cardiovascular and (probably) other functions of the N-end rule pathway involve the arginylation-mediated degradation of RGS4, RGS5, and RGS16. These "GTPase-activating" proteins function by inhibiting the signaling by specific G proteins, and are themselves down-regulated through the NO/O_2-dependent degradation by the N-end rule pathway. The N-terminal Cys residues of RGS4, RGS5, and RGS16 are oxidized *in vivo* at rates controlled by NO and oxygen, followed by the arginylation of oxidized Cys and processive proteolysis by the rest of the N-end rule pathway (Fig. 1*A*) (26, 35, 37). The arginylation branch of this pathway is also required for degradation of the *in vivo*-produced fragment of the mammalian RAD21/SCC1 subunit of cohesin.[3] This fragment, a part of circuits that control mitosis and meiosis, is produced through a cleavage by separase (55) and bears N-terminal Glu (an Nd^s residue) in mammals (56) but N-terminal Arg (an Nd^P residue) in *S. cerevisiae* (50). Given the tripartite structure of N-degrons, it is possible that some physiological substrates of R-transferase would be found to lack a complete N-degron, *i.e.* that their arginylation would not be followed by their degradation.

All eukaryotes examined, from fungi to plants and animals,

[3] J. Zhou, J. Sheng, R.-G. Hu, Y. T. Kwon, and A. Varshavsky, unpublished data.

contain both R-transferases and the N-end rule pathway. The former are sequelogous (similar in sequence) (48) throughout most of their ~60-kDa spans, even between highly divergent eukaryotes such as fungi and mammals (34). A single gene, *ATE1*, encodes R-transferase in both *S. cerevisiae* (32) and the mouse or human genomes (34, 36), whereas plants such as *Arabidopsis thaliana* contain two genes, *ATE1* and *ATE2*, which encode sequelogous R-transferases (53). Our previous studies described the cloning of yeast and mouse *ATE1*, and also the finding that mammalian (but not *S. cerevisiae*) R-transferase occurs as two isoforms, produced through alternative splicing of *ATE1* pre-mRNA (32, 34). The two mouse R-transferases were shown to be of identical sizes (516 residues), differing by a stretch of 43 residues, encoded by two alternative, adjacent and sequelogous 129-bp exons (34).

In this work, we continued the analysis of mouse *ATE1*, identifying six splicing-derived isoforms of *ATE1* mRNA and also discovering that the *ATE1* transcriptional promoter is bidirectional, driving the expression of both *ATE1* and an oppositely oriented, previously uncharacterized gene. In addition, we identified GRP78 (BiP) and protein-disulfide isomerase (PDI), two proteins located primarily in the endoplasmic reticulum (ER), as putative physiological substrates of R-transferases. Several lines of evidence (57–60) suggest that both GRP78 and PDI can be present in non-ER compartments as well, including the cytosol and the nucleus, where these proteins may be targeted by the N-end rule pathway.

A recent paper by Rai and Kashina (61) described two new splicing-derived mouse ATE1 isoforms, in addition to the two previously known ones (34). The authors claimed that the four R-transferases could arginylate unmodified N-terminal Cys and also that the two new R-transferases were inactive with N-terminal Asp or Glu (61). We show that both of these conclusions are incorrect: the activity of known R-transferases toward unmodified N-terminal Cys is negligible, and the two isoforms of R-transferase stated to be inactive with the N-terminal Asp and Glu (61) are actually active with these residues.

EXPERIMENTAL PROCEDURES

Yeast Strains, Plasmids, and β-Galactosidase Assay—Synthetic yeast media (62) contained 0.67% yeast nitrogen base without amino acids (Difco) and either 2% glucose (SD medium) or 2% galactose (SG medium). Synthetic media lacking appropriate nutrients were used to select for (and maintain) specific plasmids. Cells were also grown in rich medium (YPD) (62). The *S. cerevisiae* strains were YPH500 (*MATα ura3-52 lys-801 ade2-101 his3-Δ200 trp1-Δ63 his3-Δ200 leu2-Δ1*), SGY3 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1 ate1-Δ2::LEU2*), and JD55 (*MATα ura3-52 lys2-801 trp1-Δ63 his3-Δ200 leu2-3,112 ubr1-Δ1::HIS3*) (63). AVY34B, a *ubr1Δ ate1Δ* double mutant, was constructed by replacing the entire *ATE1* open reading frame in JD55 (*ubr1Δ*) by the antibiotic resistance gene *G418*, using homologous recombination with *G418* flanked on either side by 40 bp of *ATE1*-specific sequences (62). Mutants were selected on SD lacking His and containing G418. G418-resistant isolates were checked by PCR (for the absence of *ATE1*), by functional assays, and by N-terminal sequencing of X-β-galactosidase (X-βgal) reporters (26, 34) (see below). pUB23-X, a set of high copy plasmids expressing Ub-X-βgal reporter proteins (with X the varying residue) from the galactose-inducible P_{GAL1} promoter, was described previously (22). The cDNAs of mouse *ATE1* isoforms or mutants thereof were PCR-amplified and cloned into the low copy (*CEN*) vector p414GalS (64), which was modified by the addition of BamHI and XhoI restriction sites. The resulting plasmids, p414GalS-ATE1-X, expressed specific ATE1 isoforms (Fig. 1*D*) and their derivatives from the P_{GALS} promoter (64) in *S. cerevisiae*. Yeast transformation was performed using the lithium acetate method (62). The activity of βgal was measured as described previously (28, 30, 34, 40), using nitrophenyl-β-D-galactopyranoside as a βgal substrate and *S. cerevisiae* strains transformed with both pUB23-Asp and specific p414GalS-ATE1-X plasmids.

N-terminal Sequencing of X-βGal Reporters—*S. cerevisiae* AVY34B (*ubr1Δ ate1Δ*) was cotransformed with pUB23-Asp (expressing Ub-X-βgal) and either the p414GalS vector or p414GalS-ATE1–1A7A (expressing $ATE1^{1A7A}$). Cultures of both transformants were grown in SG to A_{600} of ~0.8, followed by the addition of galactose to 2%, to induce the expression of both Asp-βgal (Ub-Asp-βgal) and mouse $ATE1^{1A7A}$ (see Fig. 2*B* for ATE1 notations). Growth was allowed to continue until A_{600} of ~1.2, followed by collection and disruption of cells (26, 34). β-Galactosidase was purified by affinity chromatography on *p*-aminophenyl-1-thio-β-D-galactopyranoside-agarose beads (Mobitec, Göttingen, Germany) and further fractionated by SDS-10% PAGE. A Coomassie-stained band of βgal was electroblotted onto Sequiblot membrane (Bio-Rad), followed by N-terminal sequencing (at Caltech Microchemistry Facility) of ~30 pmol of βgal, using at least five cycles of Edman degradation and the Applied Biosystems 476A sequencer.

RT-PCR—Total RNA from mouse embryonic fibroblast (EF) cell lines, embryos, or tissues of adult mice was isolated using mini RNA isolation kit II (Zymo Research, Orange, CA), according to the manufacturer's instructions. The Oligotex kit (Qiagen, Valencia, CA) was then employed to isolate mRNA from total RNA. Using 5 μg of total RNA or 50 ng of total mRNA as templates, the first strand cDNAs were retrotranscribed using the Superscript kit and oligo(dT) primer (Invitrogen). Retrotranscription samples were used to carry out PCR, using standard procedures (62) and the following primers ("f" and "cr" denote "forward" and "complementary reverse" primers, respectively): 1Af, 5′-CAGAACATGGCCTCGGTGGTGGA-3′; 1Bf, 5′-CAGAACATGGCTTCTTGGAGCGCG-3′; 7Acr, 5′-GGAATCTTGTAAACTGGCTCACGGTT-3′; 7Bcr, 5′-GGAATCTTGTAAACTCAGACTTTT-3′ (these primers were used to clone *ATE1* cDNA isoforms); BamHIATE1Af, 5′-CGCGGATCCCAGGTTGTTGTACAGAACATGGCCTCGGTGGTGGA-3′; BamHIATE1Bf, 5′-CGCGGATCCCAGGTTGTTGTACAGAACATGGCTTCTTGGAGCGCG-3′; XhoIATEC1, 5′-CTAGCTCGAGTCAGTGTCTGAACAGCAGCATC-3′; mATEexon6cf, 5′-GCATCGCACAAGTTAGAGTTTACAAGATTCCTTTGCAGCTCACC-3′; mATEexon6cr, 5′-CTCTAACTTGTGCGATGCATTTTCTGGT-3′ (these primers were used to construct deletion derivatives $ATE1^{1AΔ(7AB)}$ and $ATE1^{1BΔ(7AB)}$).

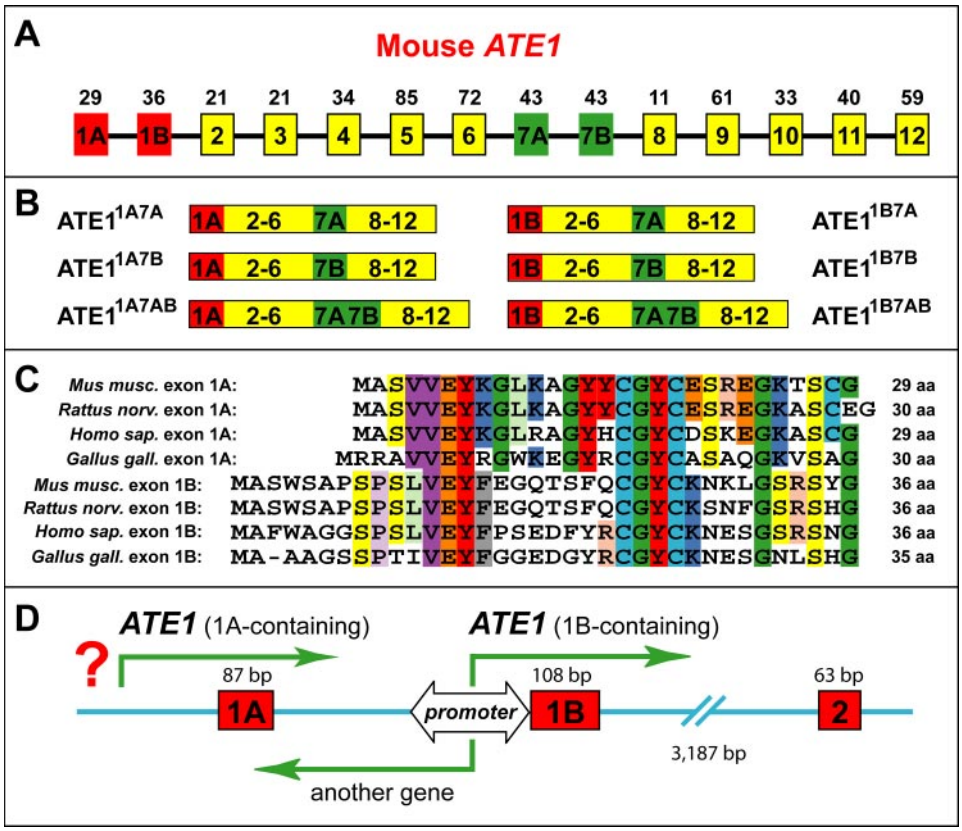

FIGURE 2. **Arginyltransferase isoforms and bidirectionality of the *ATE1* promoter.** *A*, the exons, including alternative exons, of the mouse *ATE1* gene, with deduced lengths of the corresponding peptide segments indicated on top. *B*, mouse R-transferase isoforms encoded by the *ATE1* mRNAs (see the main text). *C*, sequence comparisons of translated vertebrate *ATE1* exons 1A among themselves and to the set of longer but also sequelogous (48) alternative exons 1B. Most of recurrent amino acid identities are highlighted by color. *Mus musc.*, mouse; *Rattus norv.*, rat; *Homo sap.*, human; *Gallus gall.*, chicken. *D*, bidirectional promoter upstream of the *ATE1* exon 1B. The indicated sequence elements at the 5′-end of the mouse *ATE1* gene include the (alternative) exon 1A (87 bp; *question mark* denotes the currently unknown location of promoter element(s) for transcripts that include exon 1A); the (alternative) exon 1B (108 bp); an ~200-bp, high CpG DNA segment immediately upstream of exon 1B that functions as a bidirectional promoter (see the main text); and exon 2 (63 bp) of the multiexon *ATE1*. The stated sizes (in bp) of exons 1A and 1B refer to the protein-coding lengths; their actual lengths, which include their 5′-untranslated regions remain to be determined. *Green arrows* indicate transcriptional units, including a previously uncharacterized gene, oriented head-to-head relative to *ATE1* and transcribed, at least in part, from the bidirectional promoter[6] (see the main text).

Arginylation Assay—The *in vitro* arginylation assay with mouse embryo extracts was performed as described previously (36), with slight modifications. Briefly, the C-terminally His_6-tagged mouse ATE1 isoforms $ATE1^{1B7A}$, $ATE1^{1A7A}$, and $ATE1^{1B7B}$ (Fig. 2*B*) were expressed in Sf9 insect cells (Invitrogen) using the baculovirus system and were purified using nickel-nitrilotriacetic acid beads (Qiagen), as described in detail for the $ATE1^{1B7A}$ (ATE1-1) isoform (26). Either wild-type 12.5-day-old (E12.5) ($ATE1^{+/+}$) embryos (containing endogenous R-transferases) or $ATE1^{-/-}$ E12.5 embryos (26, 36) were lysed using motorized pestle (Kimble/Kontes, Vineland, NJ) in a hypotonic buffer (5 mM KCl, 5 mM $MgCl_2$, 1 mM EDTA, 5 mM β-mercaptoethanol, 20 mM HEPES, pH 7.5). The samples were centrifuged at 105,000 × *g* in the TL100 ultracentrifuge (Beckman Instruments, Fullerton, CA), and the supernatants, termed S105, which contained ~2 mg/ml of total protein, were collected. Specific components of [^{3}H]Arg-based arginylation assays, including embryo extracts, are mentioned under "Results and Discussion" and are described in detail elsewhere (36). The incubation time was 1 h at 37 °C. For one-dimensional electrophoretic analyses, the reaction was stopped by adding 20% CCl_3COOH to 5% (final concentration). Precipitates were collected by centrifugation, washed with 90% ethanol, dried, and resuspended in SDS-PAGE sample buffer, followed by SDS-10% PAGE, as described (36). For two-dimensional electrophoretic analyses, the reactions were stopped (in addition to removing unincorporated [^{3}H]Arg) by a quick buffer exchange to buffer A (5 mM Tris-HCl, pH 7.8), using Centricons (Millipore, Billerica, MA), followed by lyophilization.

Two-dimensional Electrophoresis and Mass Spectrometric Identification of Proteins—Two-dimensional electrophoretic analyses of our samples were carried out by Kendrick Laboratories, Inc. (Madison, WI). Isoelectric focusing was performed using 13-cm strips (pH range 4–10) (Amersham Biosciences). Second-dimension fractionation was by SDS-10% PAGE, using 20-cm-long gels. The latter were stained with Coomassie Blue, treated with Enhancer (PerkinElmer Life Sciences), and vacuum-dried, followed by fluorography with Kodak BioMax film for ~2 weeks at −80 °C. Image matching of fluorographs to Coomassie-stained gels was carried out manually. The relevant spots were excised from the gel, followed by processing for in-gel digestion with trypsin and mass spectrometry (MS/MS), which was performed at the Protein Analysis Facility of the Columbia University (New York).

Mammalian Cell Culture—NIH-3T3 cells were grown at 37 °C and 5% CO_2 in Dulbecco's modified Eagle's medium (Mediatech, Herndon, VA) containing 10% heat-inactivated fetal bovine serum and supplemented with penicillin/streptomycin/glutamine (Invitrogen). The medium was changed every 2–3 days, and cells were subcultured before reaching 90% confluency.

Northern Hybridization—Multiple tissue Northern blots containing 2 μg of mouse poly(A)$^+$ RNA per lane (Clontech) were probed with a 701-bp DNA fragment that was produced by PCR amplification from EST-AW414102 (GenBank^{TM}), using the primers 5′-ACTTTACAGTTGCTAGATAAGC-3′ and 5′-GTCCAATGACGAAGCGACAC-3′. The amplified fragment, corresponding to genomic DNA between −1603 and −903 relative to the start of *ATE1* exon 1B, was ^{32}P-labeled using the RediprimeII random prime labeling system (Amersham Biosciences) according to the manufacturer's protocol.

Hybridization was carried out for 12 h at 68 °C in ExpressHyb solution (Clontech). The blot was then washed twice for 15 min at room temperature in 2× SSC, 0.1% SDS, once for 15 min at room temperature in 0.1× SSC, 0.1% SDS, and once for 30 min at 50 °C in 0.1× SSC, 0.1% SDS, followed by autoradiography.

Luciferase Gene Fusions and Analysis of the ATE1 Promoter—pGL3-empty was constructed by removing the P_{SV40} promoter from the pGL3 promoter vector (Promega, Madison, WI), by digesting it with KpnI and HindIII, blunt-ending the resulting larger fragment with T4 DNA polymerase, and circularizing it using T4 DNA ligase. The pGL3-derived plasmids pCB30, pCB40, pCB41, pCB42, pCB43, pCB44, pCB107, and pCB108 were constructed as follows. Various 5′-untranslated regions of the mouse *ATE1* gene were amplified by PCR using the primer sets indicated below. Upon amplification, a 5′-KpnI site and a 3′-HindIII site were incorporated into the PCR-produced fragments. These (modified) genomic DNA fragments were used to replace the entire P_{SV40} promoter in pGL3 promoter vector, by digestion with KpnI and HindIII and ligation with T4 DNA ligase. The pGL3-derived plasmids pCB59F and pCB59R were constructed by PCR amplification of mouse genomic DNA (regions 1 and 2; see "Results and Discussion"), from 223 to 32 bp upstream of the *ATE1* exon 1B, using primers CB34 and CB33 and producing a HindIII site at both the 5′- and 3′-ends of the resulting fragment. It was then digested with HindIII and ligated to KpnI/HindIII-cut pGL3 promoter vector. This step yielded a mixture of nicked plasmids containing the above fragment (in either orientation) directly upstream of the firefly luciferase gene. The sample was then treated with T4 DNA polymerase and ligated using T4 DNA ligase. The above HindIII-flanked PCR fragment was also subcloned into pCB30, followed by digestion with HindIII, producing pCB60F and pCB60R. To minimize variability caused by differences in transfection efficiency, we cotransfected pGL3-derived plasmids, expressing the firefly luciferase gene, along with the pRL-SV40 plasmid (Promega), expressing the *Renilla* luciferase gene. The firefly and *Renilla* luciferases have dissimilar substrate requirements, making it possible, with the dual-luciferase reporter assay system (Promega), to measure their luminescence separately for each enzyme.

Primers used for construction of luciferase fusions are as follows: CB02, 5′-GATCAAGCTTCCGTTCCTCCTCTGTGAG-3′; CB28, 5′-GATCGGTACCGCTAGCACTTTACAGTTGCTAGATAAGC-3′; CB29, 5′-GATCAAGCTTGCCACGCCTCTCAGCGC-3′; CB32, 5′-GATCGGTACCGGGATTCCAGCACAGTG-3′; CB33, 5′-GATCAAGCTTGTCCTACGCAGCCGCCG-3′; CB34: 5′-GATCAAGCTTCTCACAGAGGAGGAACGGA-3′; CB112, 5′-GATCGGTACCGCTAGCGCGCTGAGAGGCGTGGC-3′; CB113, 5′-GATCGGTACCGCTAGCGTCCTACGCAGCCGCCG-3′; CB114, 5′-GATCAAGCTTGCGCTGAGAGGCGTGGC-3′. Primer sets for PCR of various *ATE1* 5′-regions (plasmid/forward/reverse primer) are as follows: pCB30/CB28/CB29; pCB40/CB32/CB29; pCB41/CB32/CB02; pCB42/CB32/CB33; pCB43/CB28/CB33; pCB44/CB28/CB33; pCB59/CB34/CB33; pCB60/CB34/CB33; pCB107/CB112/CB33; and pCB108/CB113/CB114.

Transient Transfection and Dual Luciferase Assay—About 20 h before transfection, NIH-3T3 cells were seeded at 1.5×10^5 cells per 3.5-cm well. Cells at 60–70% confluency were transfected with 1 μg of a pGL3-derived plasmid and 0.4 μg of pRL-SV40 vector (Promega) per well, using 10 μg of Lipofectamine and 15 μg of Plus reagent (Invitrogen), according to the manufacturer's protocol. 48 h post-transfection, cells were lysed by incubation in 0.5 ml of Passive Lysis Buffer (Promega) per 3.5-cm well on an orbital shaker at room temperature for 15 min. The extracts were clarified by centrifugation at 16,000 × *g* for 10 min at 4 °C. For luciferase assays, 10–40 μl of extract was mixed with 0.2 ml of LARII reagent (Promega), and the firefly luciferase levels were immediately measured over 10-s intervals, using a Wallac luminometer 1250. The levels of *Renilla* luciferase (the second, calibrating reporter) were then measured over 10-s intervals by adding 0.2 ml of "Stop & Glo" reagent (Promega). Light units were recorded as a ratio of firefly-to-*Renilla* luciferase levels, and values obtained were compared with those conferred by the pGL3 promoter vector (containing the P_{SV40} promoter) under the same conditions.

RESULTS AND DISCUSSION

ATE1 Exon Terminology—To accommodate, terminology-wise, the presence of an alternative *ATE1* exon 5′ to the previously described exon 1 (34), and to make the notations for R-transferase (ATE1) isoforms explicit in regard to the two pairs of alternative (and sequelogous) exons, the six *ATE1* mRNAs and corresponding proteins (see below) are denoted as follows: $ATE1^{1A7A}$, $ATE1^{1A7B}$, $ATE1^{1B7A}$, $ATE1^{1B7B}$, $ATE1^{1A7AB}$, and $ATE1^{1B7AB}$ (Fig. 2, *A* and *B*). The names of the first four ATE1 isoforms above correspond, respectively, to the following earlier names (34, 36, 61): ATE1–3, ATE1–4, ATE1–1, and ATE1–2.

Splicing-derived Isoforms of Mouse ATE1—Work by this laboratory has shown that prokaryotes, in addition to containing L-transferases (encoded by *aat* genes) (44, 45) that are nonsequelogous to the *ATE1*-encoded eukaryotic R-transferases, also contain a distinct family of *bpt*-encoded L-transferases that are sequelogs (48) of eukaryotic ATE1 R-transferases (see Introduction) (47). One of the most highly conserved sequences in ATE1 R-transferases is CGYC. It is present near N termini of these enzymes (Fig. 2*C*) and is important for their activity.[4] Although all examined R-transferases (they are confined, so far without exception, to eukaryotes) contain CGYC, a more general motif, *CXYX* (with restrictions on the identities of variable residues), is conserved in the eukaryotic/prokaryotic ATE1/Bpt superfamily of aminoacyltransferases. A search in data bases for mammalian sequelogs of CGYC-containing sequences identified an expressed sequence tag (EST) that was later found to correspond to the alternative first exon of ATE1, termed 1A (Fig. 2, *A–C*).

To determine whether *in vivo* transcripts containing other parts of ATE1-coding sequence contained exon 1A as well, RT-PCR was carried out using primers derived from both exon 1A and the sequences encoding the C termini of the alternative exons 7A or 7B of the previously characterized (34) isoforms

[4] R.-G. Hu, F. Du, and A. Varshavsky, unpublished data.

ATE1^{1B7A} and ATE1^{1B7B}. The results, using mouse brain mRNA (Fig. 3 and supplemental Fig. S1), showed that exon 1A, similarly to its previously characterized sequelog exon 1B, was present in transcripts of the expected length containing exon 7A or exon 7B, two previously characterized (34) alternative (and sequelogous) exons of equal sizes. We concluded that the *in vivo* synthesis and processing of *ATE1* pre-mRNA produces at least four distinct mRNAs, *ATE1*^{1A7A}, *ATE1*^{1A7B}; *ATE1*^{1B7A}, and *ATE1*^{1B7B}. Among the 10 sequenced *ATE1* cDNA isolates produced by RT-PCR (using exon 1A-specific forward primer and exon 12-specific complementary reverse primer), 9 isolates corresponded to *ATE1*^{1A7A} mRNA and only one corresponded to *ATE1*^{1A7B} mRNA, suggesting lower levels of the latter species in brain mRNA (see below).

RT-PCR was also employed to characterize the (approximate) relative levels of mouse *ATE1* mRNA isoforms. The previously identified (34) *ATE1*^{1B7A} mRNA was found to be a major species in every examined tissue except for spleen and skeletal muscle, whereas *ATE1*^{1B7B}, the other previously known isoform (34), was expressed at comparable or slightly lower levels in the brain, liver, and testis, was a minor species in the spleen, muscle, and heart, and was virtually absent from the kidney, in contrast to *ATE1*^{1B7A} (Fig. 3). The third isoform, *ATE1*^{1A7A}, was the major species in the muscle and was relatively high in the testis, kidney, heart, and spleen but was present at lower levels in the brain, and at still lower levels in the liver (Fig. 3). *ATE1*^{1A7B}, the fourth isoform, was expressed at readily detectable levels only in the kidney, testis, and muscle (Fig. 3), in agreement with low frequency of full-length *ATE1*^{1A7B} isolates produced by RT-PCR from brain mRNA.

Two additional species, of slightly larger than expected sizes, were amplified by RT-PCR from the muscle- and testis-derived samples specific for *ATE1*^{1A7B}, and from the testis-derived samples specific for *ATE1*^{1B7B} (Figs. 3 and supplemental Fig. S1). DNA sequencing revealed that these species corresponded, respectively, to *ATE1*^{1A7AB} and *ATE1*^{1B7AB} mRNAs, encoding slightly longer ATE1 isoforms that differed by the first alternative exons, 1A and 1B, and contained *both* of the alternative internal exons, 7A and 7B (Fig. 2, *A* and *B*). The removal of the (shorter) intron between the exons 7A and 7B, instead of a (longer) intron that in addition contained either 7A or 7B (Fig. 2*A*), lengthened but did not frameshift the *ATE1* open reading frame. The levels of *ATE1*^{1A7AB} in the muscle and testis were approximately equal to those of the (shorter) *ATE1*^{1A7B} isoform, whereas *ATE1*^{1B7AB} was significantly less abundant than *ATE1*^{1B7B} in the testis, the only tissue where *ATE1*^{1B7AB} was detectable (Fig. 3). The finding of *ATE1*^{1A7AB} and *ATE1*^{1B7AB} increases the total number of known *ATE1* mRNA isoforms to six (Fig. 2*B*). Whether all or only some of the mouse ATE1 isoforms (identified as mRNAs and characterized as recombinant proteins expressed in *E. coli* or *S. cerevisiae*) exist as naturally translated proteins in the mouse remains to be determined, because the current polyclonal antibody to mouse ATE1 was raised against one isoform, the full-length ATE1^{1B7A} (26), and would be expected to recognize all six isoforms (Fig. 2*B*). The larger isoforms, *ATE1*^{1A7AB} and *ATE1*^{1B7AB}, would be expected to differ from the other four (identically or similarly sized) ATE1 isoforms by at most ~5 kDa, and therefore size differences alone would not suffice for identification of the isoforms in mouse cells. Isoform-specific antibodies would be required to address these issues definitively. Work to produce such antibodies is under way.

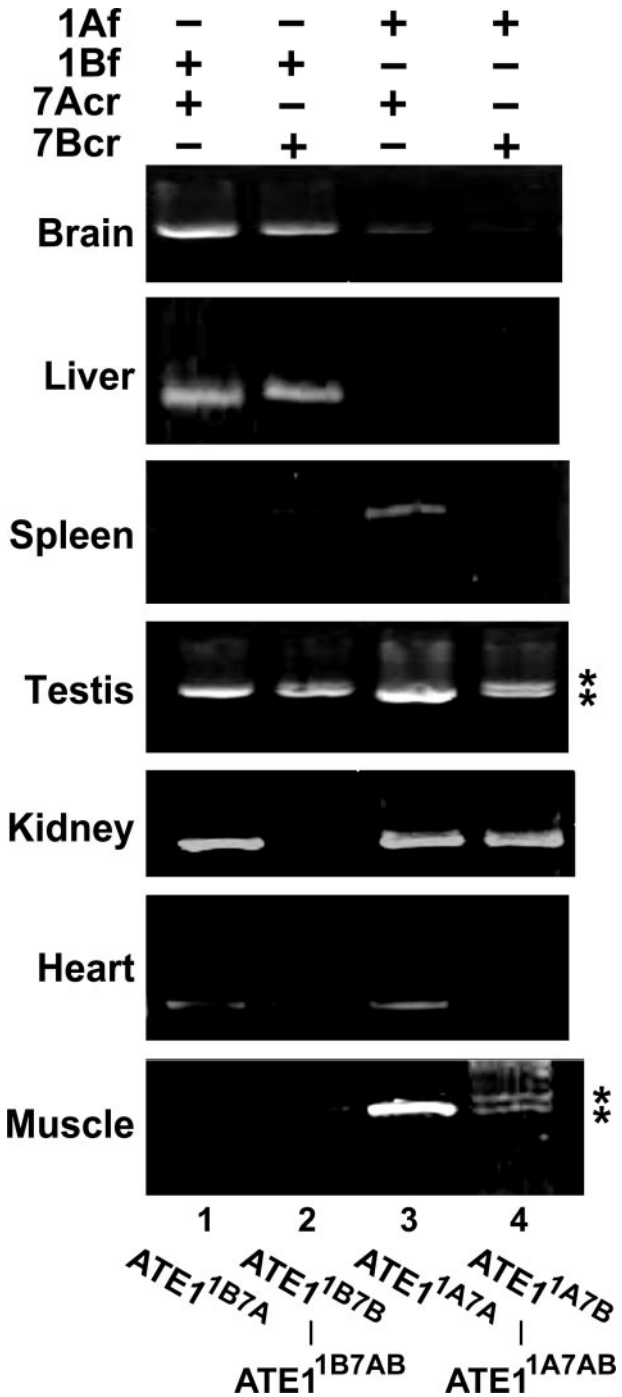

FIGURE 3. **Expression of *ATE1* mRNA isoforms in mouse tissues.** RT-PCR-based expression patterns of *ATE1*^{1A7A}, *ATE1*^{1A7B}, *ATE1*^{1B7A}, *ATE1*^{1B7B}, *ATE1*^{1A7AB}, and *ATE1*^{1B7AB} mRNAs in the indicated tissues. See "Experimental Procedures" for the names of primers and their sequences. See also Fig. *2, A* and *B,* and the main text. *Lane 1,* amplified cDNA (936 bp) encoding N-terminal region of ATE1^{1B7A}. *Lane 2,* amplified cDNA (936 bp) encoding N-terminal region of ATE1^{1B7B}. In a doublet of barely resolved bands from the testis-derived sample, the larger cDNA (1,065 bp) encoded ATE1^{1B7AB}. *Lane 3,* amplified cDNA (915 bp) encoding N-terminal region of ATE1^{1A7A}. *Lane 4,* amplified cDNA (915 bp) encoding N-terminal region of ATE1^{1A7B}. In a doublet of bands from either the testis-derived or muscle-derived samples, the larger cDNAs (1,044 bp) encoded ATE1^{1A7AB}. The doublets are indicated by *asterisks* on the *right.*

Our previous study (34) had shown that, in contrast to highly similar *ATE1* exon/intron patterns in the mouse and human genomes, including the presence of alternative exons 7A and 7B, both the *Drosophila melanogaster* (arthropod) and *Arabidopsis thaliana* (plant) *ATE1* genes lacked the alternative internal exon 7A/7B arrangement. *Drosophila* contains just one but still separate type 7 exon, whereas *Arabidopsis* contains a type 7 exon as a part of a larger exon, the result of a fusion of several exons that are separate in vertebrates. Examination of the sequenced *Drosophila* and *Arabidopsis* genomes confirmed the earlier conclusions and indicated that these nonvertebrate eukaryotes also lacked the mammalian *ATE1* pattern of alternative first exons 1A/1B. *Drosophila* and *Arabidopsis* contain only one such exon, and pufferfish (*Takifugu rubripes*), a vertebrate, apparently has a single first exon too (data not shown), in contrast to mammals and birds (Fig. 2, *A–C*).

To characterize ATE1 isoforms as enzymes and components of the N-end rule pathway, we asked whether each of them could confer metabolic instability, in *ate1*Δ *S. cerevisiae* (lacking its own R-transferase), on Asp-βgal, produced by the Ub fusion technique (65) from Ub-Asp-βgal and bearing N-terminal Asp, a secondary destabilizing residue (Fig. 1*A* and Introduction). Previous work (28, 34, 40, 63, 66) has shown that in yeast the steady-state level of an X-βgal reporter is a sensitive measure of its metabolic stability. SGY3 *S. cerevisiae* (*ate1*Δ) was cotransformed with a low copy plasmid expressing an isoform of mouse ATE1 from the P_{GALS} promoter and a high copy plasmid expressing Asp-βgal (Ub-Asp-βgal) from the P_{GAL1} promoter. Controls included either vector alone or an otherwise identical plasmid expressing *S. cerevisiae* ATE1. The levels of X-βgal proteins were determined by measuring the enzymatic activity of βgal in cell extracts. Using this assay (28, 34, 40, 63, 66), we found that, as expected, the previously characterized isoforms $ATE1^{1B7A}$ and $ATE1^{1B7B}$ (34) conferred metabolic instability on Asp-βgal, with $ATE1^{1B7B}$ being less active than $ATE1^{1B7A}$ (Fig. 4*B*), in agreement with both the earlier evidence (34) and a separate and direct arginylation test (see below). Furthermore, $ATE1^{1A7AB}$ and $ATE1^{1B7AB}$, the fifth and sixth isoforms (Fig. 2*B*) that contained *both* exons 7A and 7B, were much less active than the other four isoforms (Fig. 4*B*). Control immunoblotting tests (data not shown), using antibody to mouse ATE1 (26), confirmed that at least the $ATE1^{1A7A}$, $ATE1^{1B7A}$, and $ATE1^{1B7B}$ isoforms (Fig. 2*B*) were expressed at similar levels (less than 2-fold differences) in transfected *ate1*Δ *ubr1*Δ *S. cerevisiae* that yielded the results in Fig. 4*B*. Finally, $ATE1^{1BΔ(7AB)}$ and $ATE1^{1BΔ(7AB)}$, two artificial (engineered) deletion derivatives of the active R-transferases $ATE1^{1B7A}$ and $ATE1^{1B7B}$ that lacked *both* of the internal exons 7A and 7B, yielded the levels of Asp-βgal that were indistinguishable from those in *ate1*Δ yeast that lacked arginylation, strongly suggesting an essential requirement for the (alternative) internal exons 7A or 7B (Fig. 4*B*).

In a different approach to some of these questions, we employed an enzymatically direct *in vitro* test (as distinguished from the degradation-coupled *in vivo* assay (Fig. 4*B*)). The test used [^{3}H]Arg, purified ATE1 isoforms, and α-lactalbumin as a reporter bearing N-terminal Glu, one of three secondary destabilizing residues (Fig. 1*A*). In this previously described assay

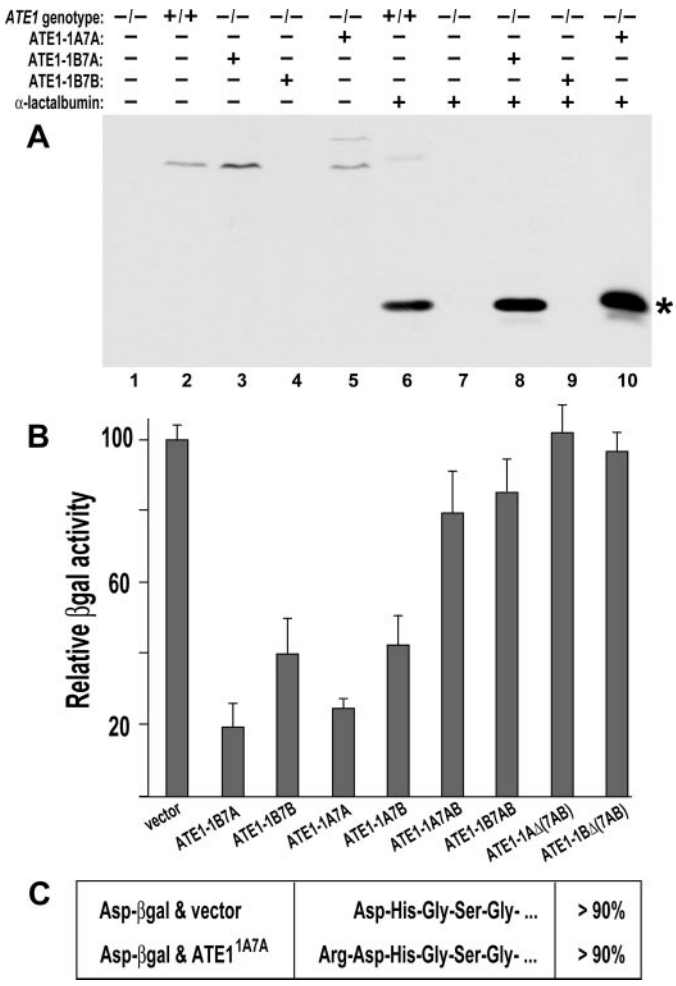

FIGURE 4. **Arginylation activity of mouse R-transferase isoforms.** *A*, purified $ATE1^{1A7A}$ can arginylate N-terminal Glu of a reporter protein. *Lane 1*, extract from $ATE1^{-/-}$ EF cells (0.25 mg of total protein) was incubated with [^{3}H]Arg under the conditions of arginylation assay (see "Experimental Procedures") but without added R-transferase, followed by SDS-PAGE and fluorography. *Lane 2,* same but with $ATE1^{+/+}$ extract. *Lanes 3–5*, same but with the added (purified) $ATE1^{1B7A}$, $ATE1^{1B7B}$, and $ATE1^{1A7A}$, respectively, at 1 μg each. *Lane 6,* same as *lane 2* but with α-lactalbumin (bearing N-terminal Glu). *Lane 7,* same but with $ATE1^{-/-}$ extract. *Lanes 8–10,* same as *lanes 3–5,* but with α-lactalbumin (the electrophoretic position of its arginylated derivative is indicated by *asterisk*). Note the absence of activity of $ATE1^{1B7B}$ at the level of sensitivity of this assay (see the main text). *B*, the six isoforms of R-transferase can metabolically destabilize, with different efficiencies, an otherwise long lived Asp-βgal reporter protein in *ate1*Δ *S. cerevisiae* (see "Experimental Procedures"). Values are the means ± S.D., from three independent measurements. The specific isoforms, as well as two deletion derivatives of ATE1, are indicated below the diagram. *C*, determination, through Edman degradation, of the N-terminal sequence of isolated Asp-βgal reporter that had been expressed in *ate1*Δ *ubr1*Δ *S. cerevisiae* either without or together with $ATE1^{1A7A}$.

(26, 36, 47), *identical amounts* of purified mouse $ATE1^{1A7A}$, $ATE1^{1B7A}$, and $ATE1^{1B7B}$ were incubated with α-lactalbumin, [^{3}H]Arg, and an extract from $ATE1^{-/-}$ mouse EF cells (which lacked R-transferases), followed by detection of the arginylated reporter by SDS-PAGE and fluorography. $ATE1^{1A7A}$ and $ATE1^{1B7A}$ were highly active as R-transferases in this assay (Fig. 4*A*). In contrast, the activity of $ATE1^{1B7B}$ was negligible at the same level of the assay's sensitivity (Fig. 4*A*) but was detectable (albeit weakly) in a different [^{3}H]Arg-based arginylation assay, with $ATE1^{-/-}$ embryo extracts (Fig. 6*A*), in agreement with the

weaker but detectable activity of ATE1^{1B7B} in the yeast-based reporter-degradation assay (Fig. 4*B*).

The alternative splicing of mouse *ATE1* pre-mRNA(s) involves two pairs of the alternative and sequelogous (48) exons, 1A/1B and 7A/7B (34) (Fig. 2*C*). Interestingly, the alternative exons 7A and 7B (Fig. 2*A*) (34) contain 5′- and 3′-splice junction consensus sequences that are characteristic of introns rather than exons (data not shown), suggesting the origins of these alternative exons, over evolutionary time, from introns, through a set of processes called "exonization" (67). Are the two pairs of alternative exons in the *ATE1* genes of mammals and birds (Fig. 2*C*) a relatively recent feature, acquired during the evolution of vertebrates, or does the alternative-exon organization of *ATE1* predate the divergence of animals and plants, having been lost in most lineages but retained in some (not all) extant vertebrate species? One difficulty in addressing this issue definitively is that positive and negative selection pressures are not the only evolutionary forces at work in such settings, where random drift, through almost neutral mutations, contributes significantly as well, especially in organismal lineages with relatively low population sizes such as vertebrates (68). The evolutionary understanding of *ATE1* would be advanced by the discovery of specific functions of R-transferase isoforms (Fig. 2*B*).

While our findings about the isoforms of mouse ATE1 were being prepared for publication, Rai and Kashina (61) published a paper that described the identification of two of these isoforms, which they termed ATE1–3 and ATE1–4. These isoforms are identical to ATE1^{1A7A} and ATE1^{1A7B}, respectively, of the present work (Fig. 2*B*). Two additional new isoforms described in the present work, ATE1^{1A7AB} and ATE1^{1B7AB} (Figs. 2*B*, 3, and 4), bring the current number of mouse R-transferase isoforms to six. In addition to describing two ATE1 isoforms, Rai and Kashina (61) presented evidence for two specific properties of ATE1 R-transferases. First, they concluded that the two new R-transferase isoforms (ATE1^{1A7A} and ATE1^{1A7B} in the current terminology; see Fig. 2*B*) could not arginylate substrates bearing N-terminal Asp or Glu (61), in contrast to the previously characterized isoforms ATE1^{1B7A} and ATE1^{1B7B} (34). Second, they concluded, on the basis of circumstantial evidence (cycloheximide-based pulse-chases with mouse *ATE1*$^{-/-}$ EF cells and the RGS4 protein reporter), that all four R-transferase isoforms could arginylate *unmodified* N-terminal Cys (61).

Our data do not support the above conclusions. Specifically, both ATE1^{1A7A} and ATE1^{1A7B} (which were described by Rai and Kashina (61) as inactive with N-terminal Asp and Glu) are actually active with these residues, as demonstrated by three different tests (Fig. 4, *A–C*). At least one of these tests, the yeast-based X-βgal activity assay (Fig. 4*B*), is identical to the assay used by Rai and Kashina (61). As mentioned above, the control immunoblotting tests (data not shown), using antibody to mouse ATE1 (26), confirmed that ATE1^{1A7A}, ATE1^{1B7A}, and ATE1^{1B7B} (Fig. 2*B*) were expressed at similar levels (less than 2-fold differences) in the transfected *ate1*Δ *ubr1*Δ *S. cerevisiae*. We further verified these results by employing a second assay, also used by the above authors (61), namely the N-terminal sequencing of X-βgal reporters. Asp-βgal was expressed in *ate1*Δ *ubr1*Δ *S. cerevisiae* (deletion of *UBR1* precluded degradation of arginylated Asp-βgal) in the presence of (coexpressed) ATE1^{1A7A}. The resulting βgal reporter was isolated, purified, and sequenced by Edman degradation, as described previously (26, 34). More than 90% of Asp-βgal isolated from *ate1*Δ *ubr1*Δ *S. cerevisiae* that coexpressed ATE1^{1A7A} was found to be the arginylated Arg-Asp-βgal (Fig. 4*C*), thus directly confirming the conclusion from yeast-based X-βgal activity assays (Fig. 4*B*).

We do not know the reason for this direct discrepancy between our experimental data (Fig. 4) and the data by Rai and Kashina (61). Because no controls verifying that R-transferases were actually expressed (as proteins) in the tester yeast strain were mentioned in the cited study, one possibility (which, if correct, would account for the above discrepancy) is that the apparent lack of activity reported by the authors (61) stemmed from negligible expression levels of the tested mouse ATE1s in *S. cerevisiae*.

Rai and Kashina (61) also observed that mouse RGS4, an N-end rule substrate that bears N-terminal Cys and is targeted for degradation by the arginylation branch of the N-end rule pathway (26, 35, 37) (see Introduction), was short lived in arginylation lacking *ATE1*$^{-/-}$ mouse EF cells that had been transiently transfected with cDNAs encoding any one of the four tested mouse ATE1s. From this (inherently indirect) evidence, they concluded that mouse R-transferases, in contrast to *S. cerevisiae* R-transferase, could mediate the arginylation of unmodified N-terminal Cys (61). However, our earlier study has shown that RGS4 isolated from mouse cells that had been treated with proteasome inhibitor (to prevent the destruction of arginylated RGS4) was not only arginylated but also contained a modified Cys residue at position 2, specifically Cys sulfonate (36). This result was followed by our recent study that demonstrated, using direct enzymatic tests with purified R-transferases, that the activity of either mouse or *S. cerevisiae* ATE1s toward unmodified N-terminal Cys was negligible and that NO-mediated oxidation of N-terminal Cys was *required* for its (subsequent) arginylation, both *in vitro* and *in vivo* (26).

The ATE1 Promoter—Analyses, using PromoterInspector (Genomatix) of the ~20-kb segment upstream of mouse *ATE1*, pinpointed an ~700-bp region immediately upstream of exon 1B that appeared to comprise a part of *ATE1* promoter (P_{ATE1}). To identify putative *cis*-acting sequences of P_{ATE1}, we tested fragments of genomic DNA that encompassed the above region, using a double-luciferase reporter system and transient cotransfections of NIH-3T3 cells. The measured levels of the firefly luciferase reporter, whose expression was driven by various DNA regions upstream of exon 1B (Fig. 2*D*), were compared with the levels of firefly luciferase driven by the control (P_{SV40}) promoter (Fig. 5, *B, D* and *E*). Whereas constructs containing the segment from −1,603 to −543 bp (pCB30; region 3) or the segment from −2,083 to −543 bp (pCB40; region 3–4) upstream of exon 1B were unable to drive luciferase expression, the constructs that contained the segment −2,083 to −223 bp (pCB41; region 2–4) or the segment from −1,603 to −223 bp (pCB43; region 2–3) yielded significant but modest levels of luciferase expression, ~50 and 30% of the control (P_{SV40}-conferred) levels, respectively (Fig. 5, *B, D,* and *E*). In contrast, the

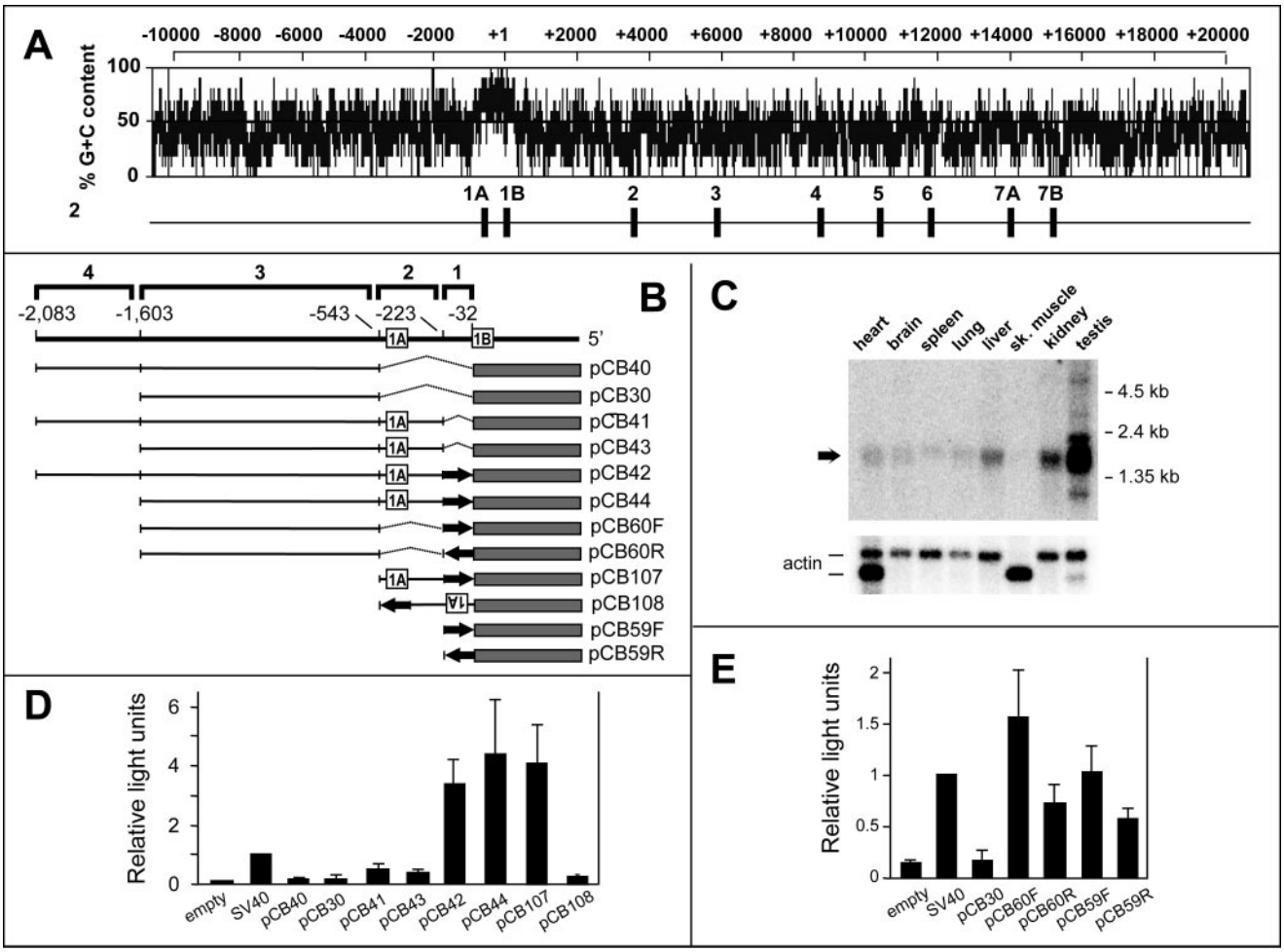

FIGURE 5. **Detection and analysis of a bidirectional promoter element in the mouse *ATE1* gene.** *A*, diagram of (G + C) content of genomic DNA (~30 kb) encompassing the 5′-region of *ATE1* reveals a CpG island between the alternative exons 1A and 1B. Positions of the first nine *ATE1* exons (1A through 7B) are shown below the (G + C) pattern. The first nucleotide of *ATE1* exon 1B is denoted as +1. *B*, the 192-bp segment immediately upstream of *ATE1* exon 1B can function as bidirectional promoter. The diagram shows segments of genomic DNA upstream of *ATE1* exon 1B that were tested for their ability to direct the expression of luciferase gene in transfected mouse cells. *C*, Northern analysis of expression of a previously uncharacterized mouse gene adjacent to *ATE1* and transcribed in the opposite direction[6] (see Fig. 2*D* and "Experimental Procedures"). The *arrow* indicates the main RNA species, with strongest expression in the testis. Actin mRNA probe was used to verify the uniformity of total RNA inputs. *D* and *E*, expression of luciferase, measured using a dual luciferase assay with extracts from NIH-3T3 cells transfected with pGL3-based plasmids that contained upstream DNA segments indicated in *B*. Luciferase levels are plotted in relative light units, normalized against the sample with P_{SV40} plasmid, arbitrarily assigned 1 light unit. Values are the means ± S.D., from three independent measurements.

segment from −2,083 to −32 bp (pCB42; region 1–4) produced ~3.5-fold higher levels of luciferase than the control (P_{SV40}) promoter, with a further increase to ~4.5-fold for regions 1–3, from −1,603 to −32 bp (pCB44) (Fig. 5, *B, D,* and *E*). The latter result suggested that the region from −2,083 to −1,603 bp (region 4) contained motifs that repressed transcription. In summary, the addition of a 192-bp fragment, located between −223 and −32 bp (region 1), to the −2,083/−223-bp segment (region 2–4) increased luciferase expression by ~6-fold (compare pCB42 and pCB41) and by ~14-fold in the absence of the repressive region 4 (compare pCB44 and pCB43) (Fig. 5, *B, D,* and *E*).

The 192-bp region-1 segment could drive luciferase expression by itself, at the level comparable with that of the control (P_{SV40}) promoter. This region, directly upstream of exon 1B, sufficed as a promoter under these conditions (compare pCB59F to pSV40; Fig. 5, *B* and *E*). Similarly to the sequences of *ATE1* exons, the sequence of this promoter region is highly conserved among the human, mouse, rat, and other examined vertebrate genomes, whereas DNA sequences outside this region are much less conserved, as determined using percent identity plot. Interestingly, this promoter motif is located between the alternative exons 1A and 1B, indicating that additional promoter elements must exist upstream of (alternative) exon 1A that drive the expression of *ATE1* mRNAs containing exon 1A (Fig. 2*D* and Fig. 5, *A* and *B*).

Promoter-relevant regions predicted by PromoterInspector encompassed, in addition to the region 1, about 500 bp of the adjacent upstream sequence as well. The addition of the region 2 (−543/−223 bp) yielded an ~4-fold increase in luciferase levels over that of the region 1 alone (compare pCB59F to pCB107) (Fig. 5, *B, D* and *E*). Furthermore, although the much larger region 1–3 (−1603/−32 bp) (pCB44) conferred luciferase levels ~4.5-fold higher than the control (P_{SV40}) promoter, the otherwise identical segment that lacked the region from −543 to −223 bp (pCB60F; region 1 and 3) yielded luciferase

levels that were only ~1.6 times higher than those by P_{SV40} or by the 192-bp fragment alone (Fig. 5, *B, D,* and *E*). We concluded that positive promoter elements of P_{ATE1} encompassed both the critical 192-bp region 1 (−223/−32 bp) and the adjacent upstream region 2 (−543/−223 bp) (Fig. 2*D* and Fig. 5, *A* and *B*).

The observed spatiotemporal complexity of *ATE1* expression in both embryonic (36) and adult tissues,[5] as well as the presence of (alternative) exon 1A upstream of the currently defined promoter region, imply that the *cis*-acting elements identified so far (Fig. 2*D* and Fig. 5*B*) are but a part of the multisite, modular P_{ATE1} promoter, whose complete span remains to be determined. Our attempts, based on the dual-luciferase assay, to identify transcription-enhancing promoter elements upstream of exon 1A were unsuccessful thus far (pCB30 and pCB40; Fig. 5, *B, D,* and *E*). In addition, analyses of the ~10-kb region upstream of exon 1A by PromoterInspector did not yield candidate sequences either significantly or immediately upstream of exon 1A. Further work to identify such sequences is under way.

Bidirectionality and CpG Island of the ATE1 Promoter—The mean G + C content of ~30 kb of the mouse genomic DNA, from ~10 kb upstream to ~20 kb downstream of *ATE1* exon 1B was ~40%. In contrast, an ~800-bp region containing both exons 1A and 1B, from ~680 bp upstream of exon 1B to ~120 bp downstream of exon 1B, had the mean G + C content of 75% (Fig. 5*A* and supplemental Fig. S2). Closer inspection identified 85 CpG dinucleotide repeats in this region (Figs. 5*A* and supplemental Fig. S2). About half of these CpGs resided in a segment directly upstream of the *ATE1* exon 1B that includes the highly conserved 192-bp region 1, shown to function as the core promoter element between the alternative exons 1A and 1B (Fig. 2*D* and Fig. 5*A*). High density of CpG repeats is a characteristic feature of previously identified bidirectional promoters in metazoan genomes (69–77). Although the number of actually characterized bidirectional promoters remains low, current estimates suggest that more than 10% of genes in a mammalian genome may be present as closely apposed head-to-head pairs transcribed from bidirectional promoters (75–77).

The 192-bp region-1 DNA segment was indeed found to drive the expression of a luciferase reporter in both orientations (Fig. 5, *B, D,* and *E*). Moreover, linking the region 3 (−1,603/−543 bp) to the oppositely oriented region 1 enhanced the expression of luciferase, similarly to the effect of region 3 (−1,603/−543 bp) on the naturally oriented region 1 segment (compare pCB59R to pCB60R and pCB59F to pCB60F; Fig. 5, *B* and *E*). We also made constructs, pCB107 and pCB108, that contained the entire region 1–2 (−543/−32 bp) in either orientation (Fig. 5*B*). Whereas pCB107 conferred high levels of luciferase expression, the same fragment in the opposite orientation was essentially inactive, indicating that the property of bidirectionality is confined mostly to the smaller (192 bp) CpG-rich region 1 promoter element embedded in a larger CpG island (Figs. 2*D*, Fig. 5, *A, B,* and *D,* and supplemental Fig. S2). Thus, the region 1, an evolutionarily conserved, CpG-rich genomic DNA fragment at the −223/−32 position relative to the start codon of the mouse *ATE1* exon 1B can function as a bidirectional promoter in transient transfection assays with luciferase reporter.

To determine whether the CpG-rich region-1 bidirectional element functioned similarly in the context of genomic *ATE1* locus, we carried out Northern hybridization of poly(A)$^+$ RNA from adult mouse tissues, using a probe derived from genomic DNA upstream of the bidirectional promoter region. To detect transcripts that were distinct from those encoding ATE1 isoforms (*e.g.* those containing exon 1A), the probe was designed so that it did not overlap with either *ATE1* exon 1A or its (expected) 5′-untranslated region, the latter suggested by information in the GenBank™ EST entries AW105867 and CJ144976 (data not shown). A low level ~2-kb transcript was detected in the heart, brain, spleen, lung, liver, and kidney but not in skeletal muscle (Fig. 5*C*). Although we still do not know the location of promoter elements that yield *ATE1* transcripts containing exon 1A (see above), the ongoing analyses[6] of a new gene that is transcribed in the direction opposite to that of *ATE1* made the assignment of transcripts described below definitive (data not shown). In particular (to cite just the evidence relevant to the specificity of detection of oppositely oriented transcripts by Northern), the exon 1A-containing *ATE1* transcripts were clearly expressed in skeletal muscle, as detected by RT-PCR (Fig. 3, *lanes 3* and *4*). In contrast, the Northern probe used to detect transcripts of the oppositely oriented gene did not detect any such transcripts in the muscle, in contrast to other tissues (Fig. 5*C*).

Interestingly, the ~2-kb transcript was highly expressed in the testis, together with additional probe-hybridizing transcripts, of >4.5, ~2.3, and <1.35 kb (Fig. 5*C*). This and related results identified a set of novel transcripts in mouse cells that are encoded by a previously uncharacterized gene in the vicinity of the 192-bp bidirectional promoter element but in the opposite orientation to *ATE1*, and partially overlapping with it. Investigations of this gene and its products are underway.[6]

Substrate Preferences of $ATE1^{1A7A}$ Versus $ATE1^{1B7A}$—In part to attempt *in vitro* identification of new physiological substrates of the N-end rule pathway, and also because we wished to determine whether different R-transferase isoforms had preferences for specific substrates, we employed the [^{3}H]Arg-based arginylation assay, using extracts from 12.5-day-old (E12.5) mouse embryos (26, 36). To preclude the *in vivo* arginylation and to make certain that only a specific (added) isoform of R-transferase was responsible for arginylation in this *in vitro* assay, the extracts were prepared from previously described (36) $ATE1^{-/-}$ embryos, which lacked arginylation. Specific epitope-tagged R-transferase isoforms ($ATE1^{1A7A}$, $ATE1^{1B7A}$, and $ATE1^{1B7B}$; see Fig. 2*B*) were expressed in insect cells using recombinant baculoviruses and were purified to near homogeneity. The mouse embryo extracts used, termed S105, were supernatants after centrifugation of total extracts at 105,000 × *g*, a step that removes the bulk of ribosomes. Besides an S105 extract (either +/+ or $ATE1^{-/-}$), the reaction mixture

[5] J. Sheng, R. Solomatina, K.-J. Shin, R.-G. Hu, and A. Varshavsky, unpublished data.

[6] C. Brower, L. Veiga, R. Jones, and A. Varshavsky, unpublished data.

contained the following: [^{3}H]Arg; a purified mouse R-transferase (ATE1^{1A7A}, ATE1^{1B7A} or ATE1^{1B7B}); puromycin to inhibit possible (residual) translation in S105; ATP regeneration system; a mixture of *E. coli* tRNAs and aminoacyl-tRNA synthetases to produce Arg-tRNA; the Arg-Ala dipeptide to inhibit possible ubiquitylation of arginylated proteins in the extract; MG132, a proteasome inhibitor, to reduce possible degradation of arginylated proteins; and a set of standard protease inhibitors. The reactions were carried out at 37 °C for 1 h, followed by removal of unincorporated [^{3}H]Arg and the processing of samples for electrophoresis and fluorography.

This approach stemmed from our earlier detection, using specific antibodies, of greatly increased *in vivo* levels of RGS4, RGS5, and RGS16 in *ATE1*$^{-/-}$ embryos (26). That *in vivo*-based finding, in conjunction with additional evidence, indicated that these regulatory proteins (see Introduction) were physiological substrates of the arginylation branch of the N-end rule pathway (26). This result also implied that comparing the levels of *in vitro*-arginylated proteins between otherwise identical extracts from wild-type (+/+) and *ATE1*$^{-/-}$ embryos would help distinguish, under such conditions, between (putative) physiological substrates of R-transferase and artifactually arginylated proteins, because the former, but not the latter, would be expected to be present at higher levels in *ATE1*$^{-/-}$ extracts, having been spared the *in vivo* arginylation and degradation. A problem that makes this *ATE1*$^{-/-}$-based approach particularly helpful stems from the fact that compartments such as the ER and Golgi contain a number of proteins that bear destabilizing (including arginylatable) N-terminal residues. These proteins had been cleaved by signal peptidase during their translocation into the ER, and the cleavage specificity of this peptidase often yields a destabilizing residue at the N terminus of a translocated, compartmentalized protein (23). Even if detergents are absent during preparation of extracts, a compartmentalized protein may partially "leak" (from membrane vesicles to be removed by centrifugation) into the cytosolic fraction and become arginylated in an *in vitro* assay. An *ATE1*$^{-/-}$ extract and a +/+ extract would be expected not to differ significantly by the levels of "leaked" proteins, in contrast to the levels of proteins that are arginylated and degraded by the N-end rule pathway in +/+ cells. (Note that some ER or Golgi proteins can be present, *in vivo*, in the cytosol and/or the nucleus as well. Thus, not all ER or Golgi proteins, if they are found to be arginylated *in vitro*, are necessarily artifacts; see below.)

In the absence of added R-transferases, no incorporation of [^{3}H]Arg could be detected with *ATE1*$^{-/-}$ embryo extracts, confirming the absence of both R-transferases and (residual) translation in this setting (Fig. 6*A*, *lane 1*). The same test with +/+ embryo extracts showed low levels of [^{3}H]Arg incorporation, mediated by endogenous R-transferases (Fig. 6*A*, *lane 2*). When a purified R-transferase, such as ATE1^{1B7A}, was added to the above assay, the resulting patterns of [^{3}H]Arg incorporation were different between +/+ and *ATE1*$^{-/-}$ extracts, as could be shown by two-dimensional electrophoresis (Fig. 6, *B–D*). This result was obtained with ATE1^{1A7A} and ATE1^{1B7A} R-transferases (Fig. 6 and data not shown). Remarkably, this approach revealed that

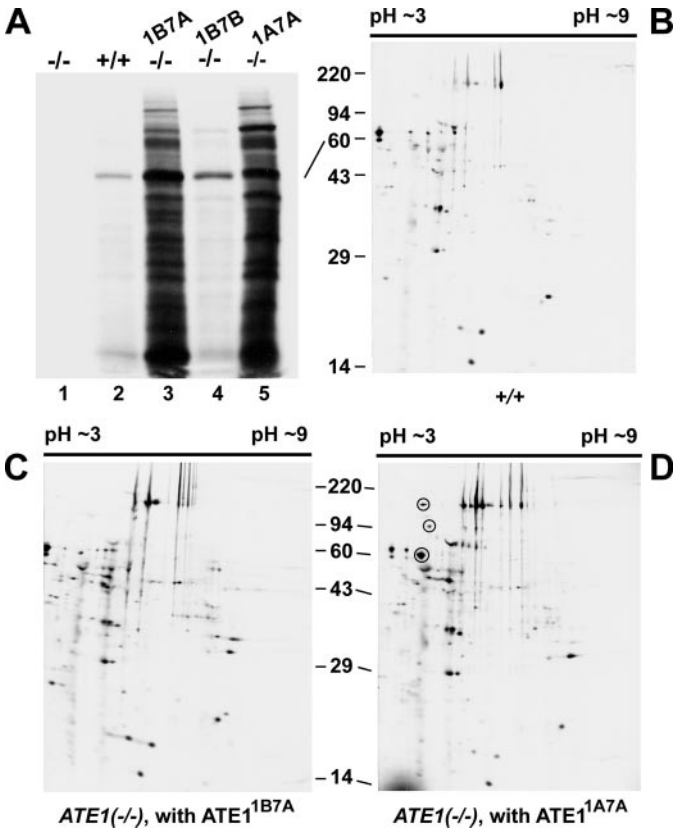

FIGURE 6. **GRP78 (BiP) and protein-disulfide isomerase as putative N-end rule substrates.** *A*, arginylation assay, with [^{3}H]Arg, mouse embryo extracts, and one-dimensional SDS-PAGE, followed by fluorography. *Lane 1,* with extract from E12.5 *ATE1*$^{-/-}$ embryos, lacking R-transferases. *Lane 2,* same but with *ATE1*$^{+/+}$ embryos. *Lane 3,* same as *lane 1* but with purified ATE1^{1B7A} added to *ATE1*$^{-/-}$ extract. *Lane 4,* same but with purified ATE1^{1B7B}. *Lane 5,* same but with purified ATE1^{1A7A}. (Note a much lower activity of ATE1^{1B7B}, in agreement with the results of assays in Fig. 3, *A* and *B*.) *B*, arginylation assay, with [^{3}H]Arg, *ATE1*$^{+/+}$ embryo extract (with endogenous R-transferases) and two-dimensional electrophoretic analysis, followed by fluorography. *C*, same but with E12.5 *ATE1*$^{-/-}$ embryos and ATE1^{1B7A}. *D*, same but with ATE1^{1A7A}. Molecular masses of marker proteins (not shown) are indicated. Three *circled protein spots* in *D* indicate, from the *top circle down*, an unknown protein, GRP78, and PDI, respectively. The latter proteins were identified by mass spectrometry (see the main text). Other also readily observable differences between two-dimensional patterns with *ATE1*$^{+/+}$ embryos (*B*) *versus ATE1*$^{-/-}$ embryos (*C*) and between ATE1^{1B7A} (*C*) *versus* ATE1^{1A7A} (*D*) remain to be addressed and are unmarked.

purified ATE1^{1A7A} or ATE1^{1B7A}, when added at identical levels to identical samples of the E12.5 *ATE1*$^{-/-}$ embryo extract, produced overlapping but significantly different patterns of arginylation, suggesting that these two R-transferases, which differ exclusively by their (sequelogous) N-terminal exons 1A and 1B (Fig. 2, *B* and *C*), have distinct preferences for specific protein substrates. Both of these R-transferases had comparable activities in the yeast-based *in vivo* degradation assay with an arginylation reporter such as Asp-βgal (Fig. 4*B*). However, the yeast-based assay, in which R-transferases were overexpressed, would have missed significant but not all-or-none differences in the levels of activity of specific R-transferase isoforms. Some of the differences between the ATE1^{1B7A}-specific and ATE1^{1A7A}-specific arginylation patterns are indicated by *circles* in Fig. 6*D*.

Thus, specific isoforms of R-transferase may have distinct functions. For example, although all of the isoforms are capable of arginylating N-terminal Asp, Glu, and (oxidized) Cys, the

relative rates of arginylation, *e.g.* of Asp *versus* Glu, may differ among the isoforms. Yet another (nonalternative) possibility is that specific sequence and conformational features downstream of (arginylatable) N-terminal residues of a substrate may lead to physiologically relevant differences in the rate of substrate's arginylation by different isoforms of R-transferase. Our previous work (34) and more recent mutagenesis studies[7] indicated that N-terminal regions of R-transferases are essential for their activity and are likely to be a part of a substrate-binding domain. Thus, the distinct but sequelogous exons 1A and 1B, encoding N-terminal regions of $ATE1^{1A7A}$ and $ATE1^{1B7A}$ (Fig. 2, *B* and *C*), respectively, may confer physiologically relevant substrate preferences on these isoforms of R-transferase. In addition, different spatial locations of R-transferase isoforms in a cell (little is known about this at present (34)) may further modulate their intrinsic preferences for specific substrates. There is extensive precedent for the notion that different splicing-derived isoforms of an enzyme can have distinct locations in a cell and thereby different functions (78–83).

Putative N-end Rule Substrates GRP78 and PDI—Some of the spots of proteins that were arginylated by $ATE1^{1A7A}$ but not by $ATE1^{1B7A}$ in $ATE1^{-/-}$ extracts (Fig. 6, *C* and *D*) were excised and sequenced, using digestion with trypsin and MS/MS mass spectrometry (see "Experimental Procedures"). Two of these arginylated proteins were found to be GRP78 (glucose-regulated protein 78) and PDI, in agreement with their apparent molecular masses and isoelectric points in two-dimensional electrophoretic patterns (Fig. 6*D* and data not shown). Other proteins that were arginylated by $ATE1^{1A7A}$ or $ATE1^{1B7A}$ in $ATE1^{-/-}$ but not in $ATE1^{+/+}$ embryo extracts (Fig. 6 and data not shown) remain to be identified. Mammalian GRP78, also referred to as BiP, is an abundant 70-kDa protein, a member of the family of HSP70 stress proteins, and a major chaperone in the lumen of ER, where the bulk of GRP78 resides (84). Mammalian PDI (PDIA1) is a 55-kDa protein-disulfide isomerase. It catalyzes disulfide bond formation, reduction, or isomerization, is a member of the family of structurally related PDI enzymes, functions also as a molecular chaperone, and resides largely in the lumen of ER (85). As nascent proteins, mammalian GRP78 and PDI bear N-terminal signal sequences that are cleaved by signal peptidase upon their translocation into the ER, yielding Glu and Asp, respectively, at the N termini of translocated GRP78 and PDI. Both of these N-terminal residues are Nd^s residues in the N-end rule (Fig. 1*A*) and would therefore be expected to be arginylated by R-transferases, in agreement with our *in vitro* results (Fig. 6*D*).

The [^{3}H]Arg-labeled spots of GRP78 and PDI were present in the extracts from $ATE1^{-/-}$ embryos (containing an added specific R-transferase) but not in extracts from $ATE1^{+/+}$ embryos (containing a set of endogenous R-transferases) (Fig. 6, *B–D*). This strongly suggested that the *in vitro*-arginylated GRP78 and PDI did not result from artifactual "leakage" of ER proteins into the cytosolic fraction during preparation of these extracts, an inference consistent with the earlier lines of evidence that both PDI and GRP78 can also occur, *in vivo*, outside the ER (57–60). Nevertheless, our evidence that the two proteins are physiological substrates of R-transferase (implying their cytosolic and/or nuclear localization), although strongly suggestive, is still indirect, because it remains formally possible that subsets of GRP78 and PDI might have leaked from the ER and were arginylated by endogenous R-transferases in the $ATE1^{+/+}$ extract (but not in the $ATE1^{-/-}$ extract) before the addition of [^{3}H]Arg, despite the near −0 °C temperature during the processing of extracts and despite a likely decrease of endogenous Arg-tRNA under *in vitro* conditions, before the addition of aminoacyl-tRNA synthetases and tRNA.

Yet another possibility is that the absence of arginylation in $ATE1^{-/-}$ embryos may increase the expression levels of ER-localized GRP78 and PDI, in comparison to $ATE1^{+/+}$ embryos. Note, however, that even if GRP78 and PDI were, in fact, induced in the absence of R-transferase, that effect alone would not preclude the possibility that these proteins are *bona fide in vivo* substrates of the N-end rule pathway. The "overexpression" interpretation was made unlikely by the data (not shown) that indicated similar total levels of, respectively, GRP78 and PDI in E12.5 $ATE1^{-/-}$ embryos *versus* $ATE1^{+/+}$ embryos, as determined by immunoblotting with antibodies to GRP78 and PDI. Although informative otherwise, such data cannot distinguish between the levels of GRP78 and PDI in the ER (their major location) and the presumably much lower (and independently changing) levels of GRP78 and PDI at their non-ER locations in a cell (see below). This is an illustration of the difficulty in addressing, definitively, the biological meaning of the above *in vitro* results (Fig. 6, *B–D*). We should also mention that the above identification of two spots of arginylated proteins in Fig. 6 as GRP78 and PDI, although direct and robust as an MS/MS-based evidence, is, strictly speaking, an interpretation of data, because of the formal possibility that other (arginylatable) protein species comigrated with GRP78 or PDI in two-dimensional electrophoretic fractionations. Although unlikely for several reasons, this possibility is formally unexcluded. Having identified GRP78 and PDI as putative arginylation substrates, our further work in this area aims to bypass *in vitro* approaches, by developing an *in vivo*-based method to solve the problem of which GRP78 and PDI are the first examples: how to identify, *in vivo*, a subset of molecules of a given protein that may undergo arginylation (or other modifications) in the cytosol, for example, if the bulk of this protein resides in another compartment such as the ER.

Remarkably, there is an independent and functionally relevant evidence for the *in vivo* location of a fraction of both GRP78 and PDI on the cytosolic side of the ER. Stockton *et al.* (57) identified a protein complex, termed the translocon-resident protein-disulfide isomerase complex (TR-PDI), that included both PDI and GRP78 and was present in the cytosol, where it interacted with the cytosolic face of SEC61, the major component of ER translocation channel. Specific aspects of the TR-PDI complex remain to be clarified, including its function in protein translocation, the paths that GRP78 and PDI take to form this complex, and the metabolic fates of TR-PDI's components, including GRP78 and PDI (57). For example, in order to form the TR-PDI complex, must these proteins be retrotranslocated from the ER into the cytosol, with their signal

[7] R.-G. Hu, H. Wang, and A. Varshavsky, unpublished data.

sequences cleaved off earlier? Or does TR-PDI consist of proteins that did not enter the ER, as yet, and therefore retained their signal sequences, bearing the initial, stabilizing N-terminal residues? One possibility is that the former route to TR-PDI is the actually taken one, at least with GRP78 and PDI. If so, these proteins bearing, respectively, Glu and Asp at the N termini (see above) may be arginylated and degraded by the N-end rule pathway upon, for instance, dissociation of the TR-PDI, if such is the dynamics of TR-PDI *in vivo*. This model is consistent with *in vitro* properties of the TR-PDI complex (57). One way to address these issues would be to determine whether the N termini of GRP78 and PDI in the TR-PDI complex are either already arginylated or can be arginylated upon dissociation of TR-PDI. One can also ask, using previously constructed mouse *ATE1*$^{-/-}$ cell lines (26, 36), whether TR-PDI-derived GRP78 and PDI in these cells (a subset of total GRP78 and PDI) become short lived upon re-expression of R-transferase.

Another study, by Grune *et al.* (60), showed that transient treatment of a rat liver cell line with H_2O_2 resulted in a degradation-mediated decrease of PDI, one of two proteins identified above as putative *in vivo* substrates of R-transferase. The oxidation-induced *in vivo* degradation of PDI was mediated by the proteasome (60), and therefore must have taken place in the cytosol and/or the nucleus. The N-terminal sequence of PDI from these cells, determined by Edman sequencing, lacked the (initially present) signal sequence, thus identifying the bulk of sequenced PDI as either a luminal ER protein (the main location of PDI) or a protein retrotranslocated from the ER. In the N-terminally sequenced PDI, its (arginylatable) Asp-26 residue was the second residue, and the identity of the first residue was unclear (60). One possibility (not the only one) is that the first residue was not the encoded Ser-25 but the post-translationally conjugated Arg residue. If so, this (oxidatively damaged?) species of PDI could be a specific subset of PDI in a cell, retrotranslocated from the ER and arginylated by *ATE1*-encoded R-transferases, thus becoming a target for the rest of the N-end rule pathway and thereby possibly accounting for the findings of Grune *et al.* (60). This currently speculative model can be tested using approaches mentioned above in the context of GRP78.

Concluding Remarks—N-terminal arginylation, a protein modification universal among eukaryotes and apparently confined to them (47), is mediated by *ATE1*-encoded R-transferases and is a part of the N-end rule pathway, a Ub-dependent proteolytic system (Fig. 1*A*). The functions of the arginylation branch of the N-end rule pathway include the regulation of cardiovascular development and homeostasis (36, 39), the regulation of signaling by NO and oxygen (26, 37) (this regulation is likely to subserve, in particular, the pathway's cardiovascular functions), the regulation of apoptosis (51, 52), the regulation of leaf senescence in plants (53), and the regulation of chromosome mechanics through the arginylation-dependent degradation of the separase-produced fragment of a subunit of mammalian cohesin[3] (see Introduction for additional information and references). Although it is possible that mouse *ATE1* gives rise to more than six splicing-derived mRNA isoforms described above, including two new isoforms (Fig. 2, *B–D*), both our present work and preceding efforts to identify *ATE1* isoforms (34, 36, 61) suggest that the current set may be the complete one. We do not know, as yet, whether R-transferase isoforms have distinct functions. The current circumstantial evidence suggests this possibility, given strong variations in expression levels of specific *ATE1* mRNA isoforms in different mouse tissues (34) (Fig. 3), and also the observed differential activity of the ATE1^{1A7A} *versus* ATE1^{1B7A} isoforms toward putative physiological substrates (Fig. 6).

Another set of new findings reported and discussed above includes the discovery of bidirectionality of the *ATE1* promoter (P_{ATE1}), specifically the ability of an ~200-bp DNA segment (region 1) immediately upstream of the start codon of exon 1B (which encodes one of two alternative N termini of R-transferase; Fig. 2, *B–D*) to function as a promoter in both orientations. This short CpG-rich segment of P_{ATE1} is bidirectional not only in model (transfection-based) settings but also in mouse tissues *in vivo*, where it mediates, at least in part, the transcription of both *ATE1* and an oppositely oriented, previously uncharacterized gene, which is expressed in several mouse tissues (Fig. 2*D* and Fig. 5*D*). Our findings about this (second) gene will be described elsewhere.[6] Although bidirectional promoters (whose features include the presence of CpG islands) appear to mediate the expression of more than 10% of genes in a mammal, relatively few examples of transcriptional bidirectionality have been characterized thus far in mammals, and even fewer bidirectional promoters were analyzed in detail (69–77). Our finding that P_{ATE1} is bidirectional adds an unexpected dimension to the understanding of this promoter, whose spatiotemporal and cell type-specific regulation is known to be complex[5] (36). Although the ~200-bp bidirectional promoter element is upstream of the start codon of the *ATE1* exon 1B, it is downstream of the alternative *ATE1* exon 1A (Fig. 1*F*). This arrangement is consistent with interesting possibilities, such as regulation of *ATE1* and the (oppositely oriented) adjacent gene through transcriptional interference and/or, for example, the formation, through bidirectional transcription in this region, of double-stranded RNA molecules, a hallmark of RNA interference (86). One of many remaining questions about P_{ATE1} is the location of its element(s) that mediates the formation of mRNAs encoding ATE1^{1A7A} and the other two exon 1A-containing ATE1 isoforms (Fig. 2, *B–D*).

Yet another set of new results is the identification of two putative physiological substrates of mouse R-transferases that are distinct from the previously known (and definitively identified) RGS4, RGS5, and RGS16 (26, 37) (see Introduction). The above RGS proteins are conditionally short lived proteins that are targeted for the arginylation and degradation via the NO-dependent activation of their Cys-based N-degrons. Given the tripartite structure of N-degrons, it is possible that some physiological substrates of R-transferase would be found to lack a complete N-degron, *i.e.* that their arginylation would not be followed by their degradation.

The two substrates of R-transferase described in the present work are GRP78 (BiP), an ER-localized chaperone, and PDI, another ER-localized enzyme. Both of these proteins are still putative N-end rule substrates, given their *in vitro* (but genetically based) identification (Fig. 6 and discussion above), and the localization of the bulk of these proteins in the ER, a compartment that lacks the N-end rule pathway (23). However, several

lines of evidence, described above, suggest that both GRP78 and PDI can be present in other cellular compartments as well, including the cytosol and the nucleus (57–60), where these proteins may be N-end rule substrates. Whether this is actually so *in vivo* and the functions of degradation of GRP78 and PDI by the N-end rule pathway remain to be determined.

The findings of the present work about mouse *ATE1*, its bidirectional promoter, alternative splicing, the resulting R-transferase isoforms, and their putative physiological substrates raised a number of new questions. Studies to address them are under way.

Acknowledgments—We thank N. Kendrick and J. Johansen (Kendrick Laboratories, Inc.) for two-dimensional electrophoresis; M. A. Gawinowicz (Columbia University) for mass spectrometric analyses; F. Rusnak, G. Hathaway, and J. Zhou (Caltech) for protein sequencing by Edman degradation; J. Arvizu, L. del Carmen Sandoval, B. W. Kennedy, and S. Pease (Caltech) for advice and assistance with mouse mutants; and E. Schwarz (Caltech) for advice on bioinformatic searches. We also thank E. Graciet for comments on the manuscript and acknowledge purification of the isoforms of mouse R-transferase by I. K. Nangiana and the late P. Snow (Protein Expression Center, Caltech).

REFERENCES

1. Varshavsky, A. (2006) *Protein Sci.* **15,** 647–654
2. Hershko, A., Ciechanover, A., and Varshavsky, A. (2000) *Nat. Med.* **10,** 1073–1081
3. Pickart, C. (2004) *Cell* **116,** 181–190
4. Elsasser, S., and Finley, D. (2005) *Nat. Cell Biol.* **7,** 16–23
5. Petroski, M. D., and Deshaies, R. J. (2005) *Nat. Rev. Mol. Cell Biol.* **6,** 9–20
6. Fang, S., and Weissman, A. M. (2004) *Cell. Mol. Life Sci.* **61,** 1546–1561
7. Hochstrasser, M. (2006) *Cell* **124,** 27–34
8. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032
9. Glotzer, M., Murray, A. W., and Kirschner, M. (1991) *Nature* **349,** 132–138
10. Johnson, E. S., Bartel, B., W., and Varshavsky, A. (1992) *EMBO J.* **11,** 497–505
11. Sheng, J., Kumagai, A., Dunphy, W. G., and Varshavsky, A. (2002) *EMBO J.* **21,** 6061–6071
12. Ardley, H. C., and Robinson, P. A. (2005) *Essays Biochem.* **41,** 15–30
13. Cardozo, T., and Pagano, M. (2004) *Nat. Rev. Mol. Cell Biol.* **5,** 739–751
14. Peters, J.-M. (2002) *Mol. Cell* **9,** 931–943
15. Baumeister, W., Walz, J., Zühl, F., and Seemüller, E. (1998) *Cell* **92,** 367–380
16. Rape, M., and Jentsch, S. (2002) *Nat. Cell Biol.* **4,** E113–E116
17. Wolf, D. H., and Hilt, W. (2004) *Biochim. Biophys. Acta* **1695,** 19–31
18. Rechsteiner, M., and Hill, C. P. (2005) *Trends Cell Biol.* **15,** 27–33
19. Collins, G. A., and Tansey, W. P. (2006) *Curr. Opin. Genet. Dev.* **16,** 197–202
20. Finley, D., Bartel, B., and Varshavsky, A. (1989) *Nature* **338,** 394–401
21. Schnell, J. D., and Hicke, L. (2003) *J. Biol. Chem.* **278,** 35857–35860
22. Bachmair, A., Finley, D., and Varshavsky, A. (1986) *Science* **234,** 179–186
23. Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 12142–12149
24. Kwon, Y. T., Xia, Z. X., An, J. Y., Tasaki, T., Davydov, I. V., Seo, J. W., Xie, Y., and Varshavsky, A. (2003) *Mol. Cell. Biol.* **23,** 8255–8271
25. Tasaki, T., Mulder, L. C. F., Iwamatsu, A., Lee, M. J., Davydov, I. V., Varshavsky, A., Muesing, M., and Kwon, Y. T. (2005) *Mol. Cell. Biol.* **25,** 7120–7136
26. Hu, R.-G., Sheng, J., Xin, Q., Xu, Z., Takahashi, T. T., and Varshavsky, A. (2005) *Nature* **437,** 981–986
27. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291
28. Suzuki, T., and Varshavsky, A. (1999) *EMBO J.* **18,** 6017–6026
29. Prakash, S., Tian, L., Ratliff, K. S., Lehotzky, R. E., and Matouschek, A. (2004) *Nat. Struct. Mol. Biol.* **11,** 830–837
30. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074
31. Kwon, Y. T., Balogh, S. A., Davydov, I. V., Kashina, A. S., Yoon, J. K., Xie, Y., Gaur, A., Hyde, L., Denenberg, V. H., and Varshavsky, A. (2000) *Mol. Cell. Biol.* **20,** 4135–4148
32. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471
33. Li, J., and Pickart, C. M. (1995) *Biochemistry* **34,** 15829–15837
34. Kwon, Y. T., Kashina, A. S., and Varshavsky, A. (1999) *Mol. Cell. Biol.* **19,** 182–193
35. Davydov, I. V., and Varshavsky, A. (2000) *J. Biol. Chem.* **275,** 22931–22941
36. Kwon, Y. T., Kashina, A. S., Davydov, I. V., Hu, R.-G., An, J. Y., Seo, J. W., Du, F., and Varshavsky, A. (2002) *Science* **297,** 96–99
37. Lee, M. J., Tasaki, T., Moroi, K., An, J. Y., Kimura, S., Davydov, I. V., and Kwon, Y. T. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15030–15035
38. Kwon, Y. T., Xia, Z., Davydov, I. V., Lecker, S. H., and Varshavsky, A. (2001) *Mol. Cell. Biol.* **21,** 8007–8021
39. An, J. Y., Seo, J. W., Tasaki, T., Lee, M. J., Varshavsky, A., and Kwon, Y. T. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 6212–6217
40. Xie, Y., and Varshavsky, A. (1999) *EMBO J.* **18,** 6832–6844
41. Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T., and Varshavsky, A. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14110–14115
42. Siepmann, T. J., Bohnsack, R. N., Tokgoz, Z., Baboshina, O. V., and Haas, A. L. (2003) *J. Biol. Chem.* **278,** 9448–9457
43. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712
44. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254,** 1374–1377
45. Shrader, T. E., Tobias, J. W., and Varshavsky, A. (1993) *J. Bacteriol.* **175,** 4364–4374
46. Erbse, A., Schmidt, R., Bornemann, T., Schneider-Mergener, J., Mogk, A., Zahn, R., Dougan, D. A., and Bukau, B. (2005) *Nature* **439,** 753–756
47. Graciet, E., Hu, R. G., Piatkov, K., Rhee, J. H., Schwarz, E. M., and Varshavsky, A. (2006) *Proc. Natl. Acad. Sci. U. S. A.* **103,** 3078–3083
48. Varshavsky, A. (2004) *Curr. Biol.* **14,** R181–R183
49. Turner, G. C., Du, F., and Varshavsky, A. (2000) *Nature* **405,** 579–583
50. Rao, H., Uhlmann, F., Nasmyth, K., and Varshavsky, A. (2001) *Nature* **410,** 955–960
51. Ditzel, M., Wilson, R., Tenev, T., Zachariou, A., Paul, A., Deas, E., and Meier, P. (2003) *Nat. Cell Biol.* **5,** 467–473
52. Varshavsky, A. (2003) *Nat. Cell Biol.* **5,** 373–376
53. Yoshida, S., Ito, M., Gallis, J., Nishida, I., and Watanabe, A. (2002) *Plant J.* **32,** 129–137
54. 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) *Nat. Genet.* **37,** 1345–1350
55. Uhlmann, F., Lottspeich, F., and Nasmyth, K. (1999) *Nature* **400,** 37–42
56. Hauf, S., Waizenegger, I. C., and Peters, J.-M. (2001) *Science* **293,** 1320–1323
57. Stockton, J. D., Merkert, M. C., and Kellaris, K. V. (2003) *Biochemistry* **42,** 12821–12834
58. Turano, C., Coppari, S., Altieri, F., and Ferraro, A. (2002) *J. Cell. Physiol.* **193,** 154–163
59. Sun, F. C., Wei, S., Li, C. W., Chang, Y. S., Chao, C. C., and Lai, Y. K. (2006) *Biochem. J.* **396,** 31–39
60. Grune, T., Reinheckel, T., Li, R., North, J. A., and Davies, K. J. A. (2002) *Arch. Biochem. Biophys.* **397,** 407–413
61. Rai, R., and Kashina, A. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10123–10128
62. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (2002) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
63. Ghislain, M., Dohmen, R. J., Levy, F., and Varshavsky, A. (1996) *EMBO J.*

15, 4884–4899
64. Mumberg, D., Muller, R., and Funk, M. (1994) *Nucleic Acids Res.* **22,** 5767–5768
65. Varshavsky, A. (2005) *Methods Enzymol.* **399,** 777–799
66. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2374–2378
67. Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. (2003) *Science* **300,** 1288–1291
68. Lynch, M., Koskella, B., and Schaack, S. (2006) *Science* **311,** 1727–1730
69. Patton, J., Block, S., Coombs, C., and Martin, M. E. (2006) *Gene* (*Amst.*) **369,** 35–44
70. Emoto, M., Miki, M., Sarker, A. H., Nakamura, T., Seki, Y., Seki, S., and Ikeda, S. (2005) *Gene* (*Amst.*) **357,** 47–54
71. Travers, M. T., Cambot, M., Kennedy, H. T., Lenoir, G. M., Barber, M. C., and Joulin, V. (2005) *Genomics* **85,** 71–84
72. Saleh, A., Davies, G. E., Pascal, V., Wright, P. W., Hodge, D. L., Cho, E. H., Lockett, S. J., Abshari, M., and Anderson, S. K. (2004) *Immunity* **21,** 55–66
73. Meier-Noorden, M., Flindt, S., Kalinke, U., and Hinz, T. (2004) *Gene* (*Amst.*) **338,** 197–207
74. Whitehouse, C., Chambers, J., Catteau, A., and Solomon, E. (2004) *Gene* (*Amst.*) **326,** 87–96
75. Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004) *Genome Res.* **14,** 62–66
76. Takai, D., and Jones, P. A. (2004) *Mol. Biol. Evol.* **21,** 463–467
77. Adachi, N., and Lieber, M. R. (2002) *Cell* **109,** 807–809
78. Stetefeld, J., and Ruegg, M. (2005) *Trends Biochem. Sci.* **30,** 515–521
79. Maniatis, T., and Tasic, B. (2002) *Nature* **418,** 236–243
80. Lareau, L. F., Green, R. E., Bhatnagar, R. S., and Brenner, S. E. (2004) *Curr. Opin. Struct. Biol.* **14,** 273–282
81. Preger, E., Ziv, I., Shabtay, A., Sher, I., Tsang, M., Dawid, I. B., Altuvia, Y., and Ron, D. (2004) *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1229–1234
82. Trotter, K. W., Fraser, I. D. C., Scott, G. K., Stutts, M. J., Scott, J. D., and Milgram, S. L. (1999) *J. Cell Biol.* **147,** 1481–1492
83. Vendel, A. C., Terry, M. D., Striegel, A. R., Iverson, N. M., Leuranguer, V., Rithner, C. D., Lyons, B. A., Pickard, G. E., Tobet, S. A., and Horne, W. A. (2006) *J. Neurosci.* **26,** 2635–2644
84. Schroder, M., and Kaufman, R. J. (2005) *Annu. Rev. Biochem.* **74,** 739–789
85. Wilkinson, B., and Gilbert, H. F. (2004) *Biochim. Biophys. Acta* **1699,** 35–44
86. Fire, A. (2006) *Q. Rev. Biophys.* **6,** 1–7
87. Bradshaw, R. A., Brickey, W. W., and Walker, K. W. (1998) *Trends Biochem. Sci.* **23,** 263–267