Identification of mammalian arginyltransferases that modify a specific subset of protein substrates

Reena Rai and Anna Kashina*

University of Pennsylvania School of Veterinary Medicine, 3800 Spruce Street, Philadelphia, PA 19104

Communicated by Roald Hoffmann, Cornell University, Ithaca, NY, May 31, 2005 (received for review April 25, 2005)

Posttranslational N-terminal protein arginylation, mediated by Arg-tRNA-protein transferase 1 (ATE1), is essential for cardiovascular development and angiogenesis in mammals but is nonessential in yeast. Evidence suggests that many proteins are arginylated *in vivo* in both mammals and yeast; however, in yeast, N-terminal arginylation can occur only on proteins bearing an N-terminal Asp or Glu, whereas in mammals, N-terminal Cys residues are also arginylation targets, suggesting that Cys arginylation contributes to the essential role of ATE1 in mammals. To date, all of the characterized forms of ATE1 in yeast and mammals have been shown to arginylate only Asp and Glu, leaving open to speculation whether Cys arginylation is possible only through other components of mammalian arginylation machinery and whether Cys-specific forms of Arg-transferase exist in mammals. Here, we report the identification of two forms of Arg-transferase in mice that are specific for N-terminal Cys. We also show that the two previously identified mammalian forms of ATE1 can arginylate Cys-containing substrates in addition to Asp- and Glu-containing substrates. This finding provides insights into the significance of Cys-specific protein arginylation in mammals and suggests possibilities of the determinants of substrate specificity within the ATE1 molecule.

Arg-tRNA-protein transferase | N-terminal protein modifications

Posttranslational incorporation of Arg into proteins was discovered ≈40 years ago (1); however, unlike other posttranslational modifications, protein arginylation has remained a mystery. Arginylation is mediated by Arg-tRNA-protein transferase 1 (ATE1) (2, 3), which transfers Arg from tRNA onto the N terminus of proteins, forming a peptide bond. ATE1 is an evolutionarily conserved enzyme, present in multiple species. Yeast ATE1 specifically arginylates proteins bearing an Asp or Glu residue at the N terminus, and no other recognition sequence is apparently necessary for the arginylation reaction to occur (4). Yeast ATE1 is a soluble protein of ≈50 kDa, with conserved Cys residues in positions 20, 23, and 94/95 that are critical for its activity (5, 6).

Studies of protein arginylation in mammalian systems have revealed that, in addition to Asp and Glu, N-terminal Cys residues can also be arginylated (7), suggesting a more complex role for mammalian protein arginylation. Furthermore, it has been found that, unlike invertebrates and unicellular organisms, mammalian species contain more than one form of ATE1, produced by alternative splicing from a single gene. Mouse ATE1-1 and ATE1-2, identical except for a single exon substitution in the middle of the molecule, have different activity, tissue specificity, and intracellular localization; however, they both have been found to arginylate only Asp- and Glu-containing substrates in a yeast complementation assay (8). The questions of whether they can also arginylate Cys and whether other Cys-specific ATE1 forms exist in higher species have remained unsolved.

Evidence indicates that many proteins are arginylated *in vivo* (9–12) and that ATE1 functions as a component of the ubiquitin (Ub)-dependent N-end rule pathway that relates the metabolic stability of a protein to the identity of its N-terminal residue (13). Although some proteins are, indeed, destabilized by the addition of N-terminal arginine (11, 13, 14), the exact relationship between arginylation and Ub protein degradation *in vivo* remains to be determined.

ATE1 knockout in yeast has no significant phenotype; however, the deletion of the *Ate1* gene in mice results in embryonic lethality and severe cardiovascular defects (15). Thus, it is evident that protein arginylation in mammals plays a vital role and that arginylation of Cys-containing substrates may have added importance, because Cys-specific arginylation occurs only in mammals (where ATE1 presence is critical) but not in yeast (where ATE1 is nonessential). In view of this recent finding, it appears critically important to understand whether the previously identified mouse ATE1-1 and ATE1-2 can arginylate Cys-containing substrates and whether other Cys-specific ATE1 forms exist in mammals.

Here, we report the identification of two previously unknown ATE1 forms in mice, ATE1-3 and ATE1-4, and show that these forms specifically arginylate substrates containing N-terminal Cys. We also show that the previously identified ATE1-1 and ATE1-2 can also arginylate Cys, in addition to the Asp- and Glu-containing substrates, and that ATE1-1, previously considered to be more active, does so with less efficiency. This finding provides insights into the significance of Cys-specific protein arginylation in mammals and suggests possibilities of the determinants of substrate specificity within the ATE1 molecule.

Materials and Methods

Yeast Strains and Plasmids. *Saccharomyces cerevisiae ate1*Δ and *ubr1*Δ strains and the background wild-type strain were obtained from Open Biosystems (Huntsville, AL). Yeast strains were grown by following standard procedures, as described in ref. 16. The double mutant *ate1*Δ *ubr1*Δ was obtained by mating haploid strains of *ate1*Δ (type **a**) and *ubr1*Δ (type α), followed by tetrad dissection and analysis, as described in ref. 16. Mouse ATE1 forms were isolated by RT-PCR from mRNA derived from mouse embryonic fibroblasts and subcloned into pBluescript SK(+) for initial sequencing, low-copy-number p325gal1 vector for yeast expression and complementation assays (17), or pEGFP-N2 (Clontech) for mammalian expression and localization studies. Mouse DNA encoding the GTPase-activating protein regulator of G protein signaling (RGS)4 was isolated by RT-PCR from mRNA derived from mouse brain tissue and subcloned into pCDNA3 (Invitrogen) containing mammalian pCMV promoter and C-terminal V5 and His_6 epitope tags for mammalian degradation assay. High-copy-number pUB23-X plasmids expressing Ub-X-β-gal proteins from P_{GAL} promoter were described in ref. 13; in this study, we used plasmids encoding Ub-X-β-gal in which X was M, C, D, E, or R for yeast

Freely available online through the PNAS open access option.

Abbreviations: ATE1, Arg-tRNA-protein transferase 1; RGS, regulator of G protein signaling; Ub, ubiquitin.

Data deposition: The cDNA sequences reported in this article have been deposited in the GenBank database (accession nos. DQ093639 and DQ093640).

*To whom correspondence should be addressed. E-mail: akashina@vet.upenn.edu.

BIOCHEMISTRY

complementation assay and C, D, or E for Edman sequencing and mass spectrometry.

mRNA Isolation and Tissue-Specific RT-PCR. mRNA was isolated from freshly excised mouse tissues and cultured cells by using TRIzol reagent (Invitrogen). RT-PCR was conducted by using Super Script II reverse transcriptase (Invitrogen) according to the procedure described in the product insert, followed by *Ate1*-specific PCR using primers corresponding to either the beginning and end of ATE1 ORF (for identification of ATE1-3 and ATE1-4) or to the specific sequences within the alternatively spliced exons (for tissue-specific PCR; see explanation in the *Results*).

Cell Culture, Transfection, and Fluorescence Microscopy. Immortalized cultures of mouse wild-type and *Ate1*$^{-/-}$ embryonic fibroblasts (15) were grown in 1:1 DMEM/F10 medium supplemented with 10% FBS and antibiotics, as described in ref. 15. Transient transfection was performed by using Lipofectamine Plus reagent (Invitrogen). Transfection conditions were optimized to achieve at least 30% transfection efficiency. GFP fluorescence was observed by using a Nikon TE300 inverted microscope, recorded by using a Hamamatsu OrcaAG digital camera, and analyzed by using METAMORPH imaging software (Universal Imaging, Downingtown, PA).

Yeast Complementation Assay. Yeast wild-type and *ate1*Δ mutants were cotransformed with plasmids carrying ATE1-1, ATE1-3, or ATE1-4 cDNA and one of the five Ub-X-β-gal protein substrates. Protein expression was induced by 1% galactose, and β-gal activity was measured by using *o*-nitrophenyl β-D-galactoside liquid assay as described in ref. 8.

Substrate Preparation for Edman Sequencing and Mass Spectrometry. Yeast double mutant *ate1*Δ *ubr1*Δ was cotransformed with plasmids carrying ATE1-1, ATE1-3, or ATE1-4 cDNA and Ub-D-β-gal, Ub-E-β-gal, or Ub-C-β-gal protein substrates. Expression was induced by 1% galactose, yeast extracts were prepared by using the liquid nitrogen procedure (18), and β-gal substrates were immunoprecipitated from the yeast extracts by the addition of monoclonal anti-β-gal (Promega) and protein-A-agarose beads (Invitrogen). β-gal immunoprecipitates were fractionated on SDS/polyacrylamide gels (8% acrylamide, acrylamide/methylene bisacryamide = 30:0.8), and 116-kDa bands corresponding to the β-gal-derived polypeptides were excised from either the Coomassie stained gels (for mass spectrometry) or the amido-black-stained poly(vinylidene difluoride) membranes after transfer (for N-terminal Edman microsequencing). Edman sequencing was performed at the W. M. Keck facility at Yale University (New Haven, CT). Mass spectrometry was performed at Midwest Bio Services (Overland Park, KS). For identification of the N-terminal β-gal peptide by mass spectrometry, β-gal substrates were eluted from the gel and digested by EndoLysC (Roche Applied Science, Indianapolis).

Mammalian Assay for ATE1-Dependent RGS4 Degradation. Mouse wild-type and *Ate1*$^{-/-}$ immortalized embryonic fibroblasts were cotransfected with pCDNA3-RGS4 and plasmids bearing ATE1-1, -2, -3, or -4 GFP fusions, or pEGFP-N2. Transfected cells were equally divided into wells of 24-well plates and cultured overnight to produce semiconfluent cultures. The medium in each well was supplemented with 10 μg/ml cycloheximide, and the contents of the entire well was resuspended in 100 μl of SDS sample buffer at 0 time point and at regular time intervals after cycloheximide addition (0, 1, 2, 3, 5, and 7 h). Equal volumes of samples were analyzed by SDS/PAGE and the amount of RGS4 in each sample was quantified by immunoblotting with anti-V5 antibody (Invitrogen) by using ECF alkaline phosphatase substrate and a Storm imaging system (Amersham Pharmacia).

Results

Identification of ATE1-3 and ATE1-4. Because the previously identified ATE1-1 and ATE1-2 showed no ability to arginylate N-terminal Cys in yeast complementation assays (8), we hypothesized that the mouse genome encodes additional ATE1 forms. To identify these forms, we first performed BLAST searches of mouse genomic databases by using short sequences corresponding to highly conserved motifs within all Arg-transferases identified to date, and we found no additional sequences that could encode the novel Arg-transferases. We then analyzed the sequence of the *Ate1* gene from 37 kb upstream (up to the next gene encoded upstream of *Ate1* on mouse chromosome 7) to ≈3 kb downstream of the first exon, in an attempt to find additional areas containing the Kozak initiation sequence [(GCC) GCCRCCATGG (19)] that might indicate an alternative ATE1 translation initiation site. Whereas no downstream or internal regions of the *Ate1* gene contained apparent coding sequences corresponding to alternatively spliced exons, we found a Kozak consensus sequence followed by an in-frame stretch of 30 codons ≈0.6 kb upstream of the originally identified first exon of the *Ate1* gene.

To test whether this sequence indeed corresponded to an alternative first exon, we designed a set of primers corresponding to the beginning of the newly found putative exon (forward primer) and the end of ATE1-1/2 coding sequence (reverse primer) and analyzed a preparation of mouse mRNA by RT-PCR. This analysis yielded a mixture of two cDNAs of ≈1.5 kb, whose sequence closely resembled that of ATE1-1 and ATE1-2, with the exception of the first exon (Fig. 1), suggesting that the newly identified forms are produced by alternative splicing of exons 1 and 2 in the *Ate1* gene (Fig. 1). To follow the existing ATE1 nomenclature, we named these forms ATE1-3 and ATE1-4.

To check whether the *Ate1* gene may encode additional ATE1 forms, we analyzed the entire 37-kb sequence upstream of the *Ate1* gene in search for any stretches of sequence containing the Met coding sequence followed by an uninterrupted in-frame stretch of sequence encoding at least 10 amino acid residues. We found five such sequences. Although none of them contained apparent Kozak consensus, we designed forward primers corresponding to each of these sequences and performed RT-PCR using the reverse primer corresponding to the end of ATE1 cDNA. None of these reactions yielded a product, suggesting that none of the five sequences corresponded to the ATE1 coding sequence. Thus, we conclude that the mouse *Ate1* gene encodes a total of four ATE1 forms.

To search whether other species contain ATE1 forms corresponding to the mouse ATE1-3 and ATE1-4, we performed a BLAST search with the sequence corresponding to the newly identified first exon of the *Ate1* gene (30 aa; Fig. 1). Several species of higher vertebrates, including humans, rats, and primates, were found to contain predicted coding sequences with high homology to the first 30 residues of the ATE1-3/4 sequence (identical except for five homologous amino acid substitutions), suggesting that the four ATE1 forms are evolutionarily conserved in vertebrates. Remarkably, the highly conserved motif CGYC, shown to be critical for ATE1 activity (5), was present in a similar position in all ATE1 forms (boxed in Fig. 1), indicating that this motif is, indeed, universal.

Tissue Expression and Intracellular Localization of ATE1-3 and ATE1-4. The presence of multiple ATE1 forms in mice could result in different tissue-specific expression and intracellular localization, as shown for mouse ATE1-1 and ATE1-2 in ref. 8. To analyze the presence or absence of expression of ATE1 forms in different tissues, we performed RT-PCR on mRNA preparations from a number of adult and embryonic tissues (see Fig. 2*A* for full list) by using sets of primers specific for individual ATE1 forms. The

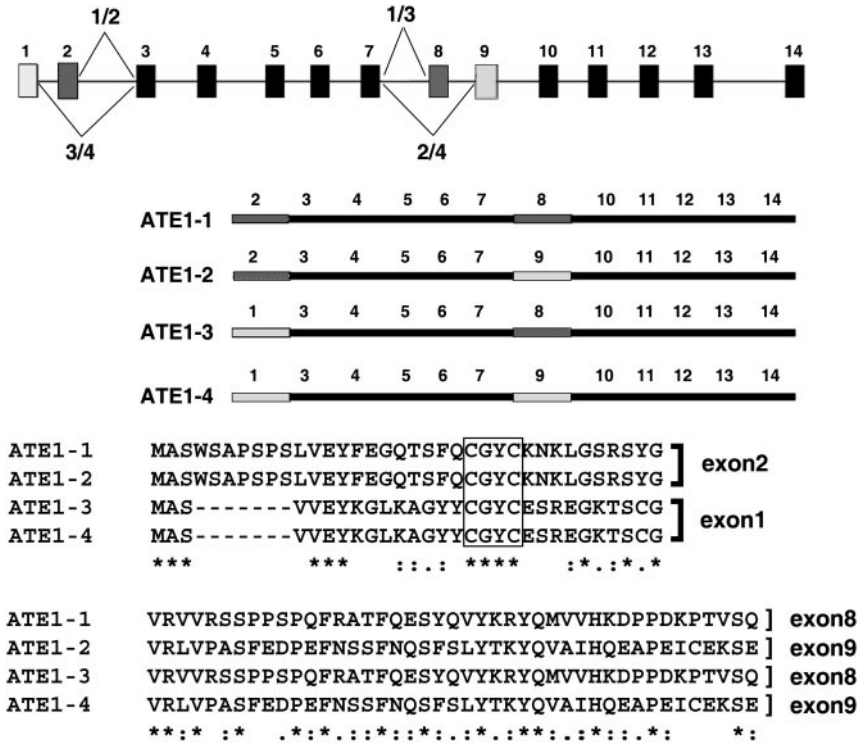

Fig. 1. Identification of mouse ATE1-3 and ATE1-4. (*Top*) A diagram representing the intron–exon structure of the mouse *Ate1* gene, which is encoded by a 200-kb sequence on chromosome 7, containing a total of 14 exons. Four ATE1 forms are produced by alternative splicing of exons 1 and 2 and exons 8 and 9. (*Middle*) Schematic representation of cDNAs of ATE1 alternatively spliced forms. Exons encoding the corresponding cDNA sequences are listed on the top. (*Bottom*) Sequence alignment of the alternatively spliced exons reveals high homology among mouse ATE1 forms. The boxed motif CGYC, previously shown to be critical for ATE1 activity, is present in all four forms. *, Identical residues in all four forms; : and . denote residues with higher and lower degrees of homology, respectively.

primers were designed corresponding to the nonhomologous sequences within exons 1 or 2 (forward primers, F) and within exons 8 or 9 (reverse primers, R). These primers, named F1, F2, R8, and R9, respectively, were used in the following combinations: F1 + R8 for ATE1-3, F1 + R9 for ATE1-4, F2 + R8 for ATE1-1, and F2 + R9 for ATE1-2, yielding an ≈1-kb product for each ATE1 form. The specificity of these primer combinations was confirmed with the use of purified plasmids containing cDNAs for individual ATE1 forms (data not shown).

We found that, whereas ATE1-1 and ATE1-2 were expressed in all tissues at all stages of development, ATE1-3 and ATE1-4 showed distinct tissue specificity (Fig. 2*A*). ATE1-3 was found only in adult animals and immortalized fibroblasts in culture and was expressed with high abundance in testis, kidney, liver, and brain. ATE1-4 showed some level of expression in embryos and pups and increased levels in several tissues of the adult animal, including brain, heart, liver, spleen, testis, and, to a lesser extent, spinal cord. Thus, it appears that ATE1-3 and ATE1-4 have selective functions in postembryogenesis.

To analyze the intracellular localization of ATE1-3 and ATE1-4, we transiently transfected immortalized mouse embryonic fibroblasts in culture with vectors encoding GFP fusions of ATE1-3 and ATE1-4 (with GFP fused to the C terminus of ATE1, away from all of the known ATE1 functional regions) and analyzed the intracellular pattern of GFP fluorescence. It was previously shown that, under similar conditions, ATE1-1 and ATE1-2 show differences in intracellular localization, with ATE1-1 localizing either exclusively in the cytoplasm or preferentially in the nucleus and ATE1-2 always excluded from the nucleus. In agreement with these observations, we found ATE1-3 (with higher identity to ATE1-1) to localize in several distinct patterns, including or excluding the nucleus (Fig. 2*B Top*), and ATE1-4 (with higher identity to ATE1-2) to localize exclusively in the cytoplasm (Fig. 2*B Bottom*). In some cells, ATE1-3 also showed preferential localization at the leading edge of motile cells (Fig. 2*B*). The physiological significance of such localization remains to be understood.

ATE1-3 and ATE1-4, Unlike All of the Previously Identified Arg-Transferases, Cannot Arginylate Protein Substrates Containing N-Terminal Asp or Glu. To study the activity and substrate specificity of ATE1-3 and ATE1-4, we first performed a yeast complementation assay to test their ability to induce degradation of β-gal-derived substrates with different amino acid residues at the N terminus. This assay, previously used to characterize N-end rule degradation and all of the previously identified forms of ATE1 (8), involves co-transformation of yeast *ate1*Δ cells with ATE1 and a β-gal-derived substrate containing either Met or Arg (stabilizing and destabilizing residue for positive and negative control of β-gal activity, respectively), or Asp, Glu, or Cys (possible arginylation substrates) in the N-terminal position. It has been shown that, under these conditions, the presence of yeast ATE1 or mouse ATE1-1 and ATE1-2 induces degradation of Asp- and Glu-containing β-gal substrates, resulting in the loss of β-gal activity (ref. 8 and Fig. 3, filled bars). However, neither ATE1-3 nor ATE1-4 induced degradation of Asp-, Glu-, or Cys-containing β-gal substrates, suggesting that the two newly identified mouse ATE1 forms are unable to arginylate any of these substrates in yeast (Fig. 3, white bars).

To confirm that the absence of β-gal degradation was indeed due to the absence of arginylation and not to the partial N-terminal blockage of the experimental substrates that may reduce their degradation and obscure the results of the β-gal assay, we performed a similar assay using an *ate1*Δ *ubr1*Δ yeast

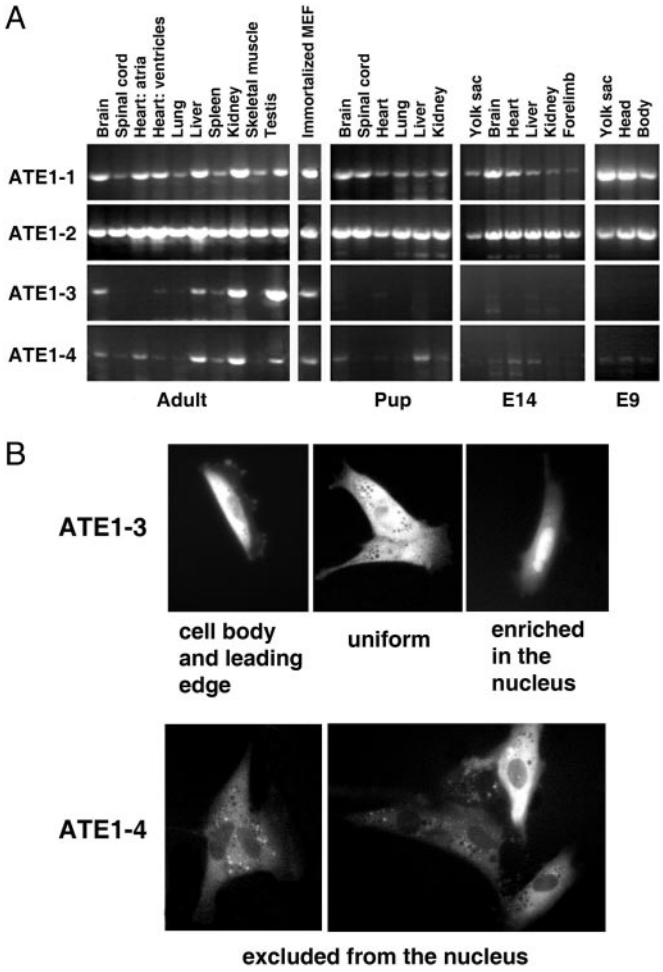

Fig. 2. Mouse ATE1 forms have different patterns of tissue-specific expression and intracellular localization. (*A*) RT-PCR of embryonic and adult mouse tissues using sets of primers specific for individual ATE1 forms. Whereas ATE1-2 is ubiquitously expressed and ATE1-1 is present throughout embryogenesis and adulthood, ATE1-3 and ATE1-4 show distinct patterns of tissue and developmental-stage expression and are present mostly in select adult tissues. (*B*) Intracellular localization of GFP fusions of ATE1-3 and ATE1-4 in wild-type mouse embryonic fibroblasts. Whereas ATE1-4 localizes uniformly in the cytoplasm and is always excluded from the nucleus, ATE1-3 has three distinct localization patterns: cell body and leading edge, uniformly in the nucleus and cytoplasm, or enriched in the nucleus.

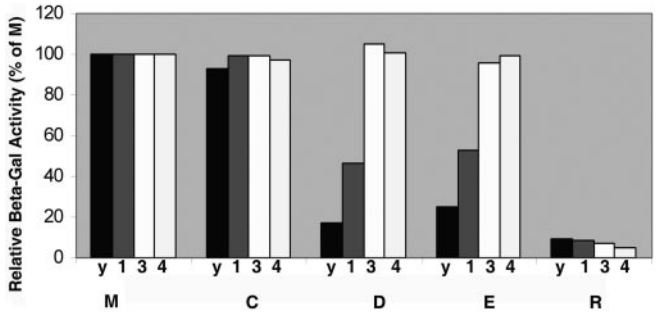

Fig. 3. ATE1-3 and ATE1-4, unlike ATE1-1, are unable to arginylate substrates with N-terminal Asp or Glu. Liquid assays of β-gal substrates in wild-type yeast (y) or in *ate1*Δ yeast mutants in the presence of mouse ATE1-1 (1), ATE1-3 (3) or ATE1-4 (4). Letters on the bottom of the chart (M, C, D, E, and R) correspond to the amino acid residue in the N-terminal position of each β-gal substrate. In yeast, substrates containing N-terminal Met (M) and Cys (C) are stable, whereas substrates containing N-terminal Arg (R) are highly unstable, resulting in dramatic reduction of β-gal activity. In the presence of ATE1, β-gal substrates containing N-terminal arginylatable residues receive an Arg, resulting in their destabilization and rapid degradation, which can be detected by the loss of β-gal activity. In this assay, Asp (D)- and Glu (E)-containing substrates get destabilized in the presence of yeast ATE1 or mouse ATE1-1, suggesting that these ATE1 forms can arginylate the corresponding β-gal-derived substrates, whereas ATE1-3 and ATE1-4 have no effect on β-gal stability, suggesting that they are not able to arginylate Asp, Glu, or Cys in yeast. β-gal activity is represented as the percentage of activity of M-containing substrate for each data set. Bars represent the average of seven independent experiments.

strain, where the knockout of the N-end rule degradation component ubr1 results in stability of arginylated substrates. It has been previously shown that, under such conditions, β-gal-derived substrates can be isolated in an arginylated form, and the addition of the N-terminal arginine can be verified by N-terminal Edman sequencing (8). In agreement with the previous observations, Asp- and Glu-containing substrates, isolated from *ate1*Δ *ubr1*Δ cells in the presence of ATE1-1, were completely arginylated, whereas Cys-containing substrates were not. However, we found no N-terminal Arg on any of the substrates obtained in the presence of either ATE1-3 or ATE1-4, confirming the results of the color assay (Table 1). Thus, ATE1-3 and ATE1-4, unlike all of the previously characterized Arg-transferases, are unable to arginylate the N-terminal Asp or Glu in a yeast complementation assay.

ATE1-3 and ATE1-4 Are Specific for N-Terminal Cys. Because ATE1-3 and ATE1-4 showed no activity in yeast complementation assays, two possibilities remained. First, it was conceivable that, despite their high sequence similarity to other Arg-transferases and the apparent presence of all of the motifs known to be essential for Arg-transferase activity, ATE1-3 and ATE1-4 were inactive. Second, it was possible that the two unique ATE1 forms were specific for N-terminal Cys, which has been shown to be non-arginylatable in yeast, presumably due to the absence of Cys-modifying machinery (8, 15) and, thus, could work only in a mammalian-based experimental system.

To distinguish between these two possibilities and to find out whether ATE1-3 and ATE1-4 are active Arg-transferases, we performed an assay similar to the yeast complementation assay but using cultured mouse fibroblasts derived from the embryos of *Ate1*-knockout mice (15). In this assay, we cotransfected *Ate1*$^{-/-}$

Table 1. N-terminal Edman sequencing and mass spectrometry confirm the inability of ATE1-3 and ATE1-4 to arginylate Asp, Glu, or Cys in yeast complementation assay

ATE1 form	Unmodified substrate sequence	N-terminal Edman sequence	Substrate recovery, %
ATE1-1	**D**HGSGAWLLPVSLVK---	**RD**GSG----------	100
	EHGSGAWLLPVSLVK---	**RE**GSG----------	100
	CHGSGAWLLPVSLVK---	**C**GSG----------	20
ATE1-3	**D**HGSGAWLLPVSLVK---	**D**GSG----------	100
	EHGSGAWLLPVSLVK---	**E**GSG----------	100
	CHGSGAWLLPVSLVK---	**C**GSG----------	20
ATE1-4	**D**HGSGAWLLPVSLVK---	**D**GSG----------	100
	EHGSGAWLLPVSLVK---	**E**GSG----------	100
	CHGSGAWLLPVSLVK---	**C**GSG----------	20

Asp and Glu β-gal substrates isolated from *ate1* Δ *ubr1* Δ mutants (where the absence of ubr1 makes the arginylated substrates stable) in the presence of ATE1-1 are arginylated, whereas those isolated in the presence of ATE1-3 and ATE1-4 are not. As previously shown, 80% of Cys-containing substrate *in vivo* is N-terminally blocked for Edman sequencing, but neither of the Arg-transferases is able to arginylate the unmodified portion of N-terminal Cys residues in the yeast-based assay, suggesting that, in agreement with previous observations, Cys is not accessible for arginylation machinery in yeast. All results were confirmed by mass spectrometry.

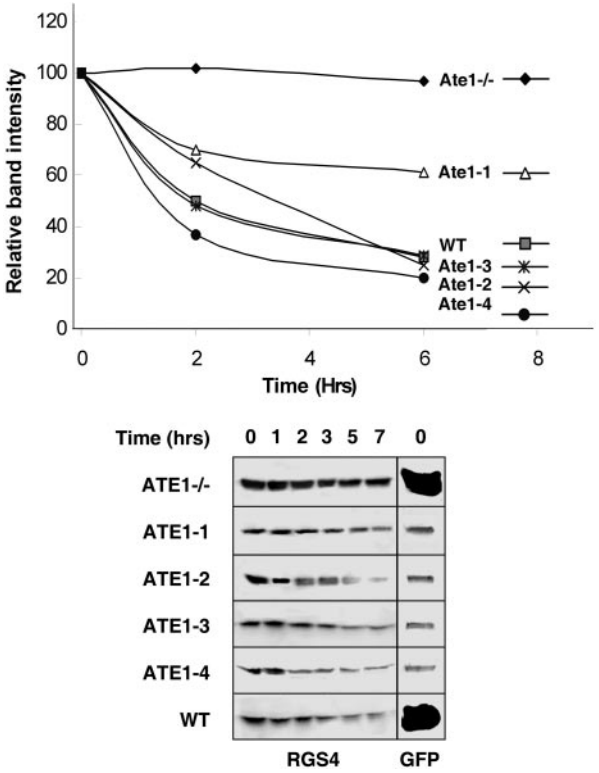

Fig. 4. Mouse ATE1 forms are able to arginylate substrates with N-terminal Cys. (*Upper*) the quantification results of RGS4 levels in cells incubated with cycloheximide (to inhibit *de novo* protein synthesis) at increased time intervals. Mouse *Ate1*$^{-/-}$ embryonic fibroblasts were cotransfected with mouse RGS4 and either pEGFP-N2 vector (negative control) or one of the ATE1 forms, followed by cell collection at various time intervals and quantification of RGS4 levels by immunoblotting. For positive control, wild-type mouse embryonic fibroblasts were cotransfected with pEGFP-N2 and RGS4. RGS4 levels were quantified at 0, 1, 2, 3, 5, and 7 h after cycloheximide addition in five independent experiments for ATE1-1 and ATE1-2 and three independent experiments for controls and ATE1-3 and ATE1-4. To obtain the best fitted curves, RGS4 readings were first averaged between experiments and then averaged between two neighboring time intervals as follows: 0 + 1 h = 0 time point on the curve, 2 + 3 h = 2 h time point on the curve, and 5 + 7 h = 6 h time point on the curve. RGS4 levels were calculated as the percentage of 0 time point. (*Lower*) Images of representative immunoblots quantified in the chart. (*Left*) RGS4 immunoblots at the corresponding time intervals (indicated on the top) for *Ate1*$^{-/-}$, ATE1 forms, and *Ate1*$^{+/+}$ cells. (*Right*) Anti-GFP immunoblots at 0 time point. For each ATE1 form ≈80-kDa fusion protein was present in approximately equal amounts throughout the experiment. In control cells, the ≈27-kDa GFP protein was present at extremely high levels, reflecting the elevated efficiency of GFP expression in the absence of N-terminal fusion.

cells with (*i*) ATE1-1, -2, -3, or -4 and (*ii*) mouse RGS4 [the only Cys-specific arginylation substrate identified to date (14)]. Because the RGS4 half-life is significantly reduced upon arginylation (14), ATE1 activity can be monitored by its ability to destabilize RGS4 in cultured cells. Indeed, RGS4 was completely stable in *Ate1*$^{-/-}$ cells over a 6-h time course but degraded by >50% over the course of 2 h in wild-type *Ate1*$^{+/+}$ fibroblasts (Fig. 4). Moreover, cotransfection of *Ate1*$^{-/-}$ cells with RGS4 and ATE1-2, -3, or -4 resulted in RGS4 degradation with kinetics similar to wild-type cells (Fig. 4). ATE1-1 also induced degradation of RGS4; however, its effect was much less prominent, suggesting that its ability to arginylate RGS4 is much less efficient than that of the other ATE1 forms. Control experiments using blots with anti-GFP antibody showed that ATE1 forms were expressed equally in all cells at 0 time point of cycloheximide addition (Fig. 4, *Lower Right*) and throughout the experiment (not shown). This experiment demonstrates that all ATE1 forms can arginylate substrates bearing N-terminal Cys. In addition, the experiment demonstrates that the two newly identified ATE1 forms are active Arg-transferases specific for Cys-containing substrates.

Discussion

Our data demonstrate that Arg-transferases can differ by substrate specificity and that the mammalian-specific branch of protein arginylation that modifies N-terminal Cys involves an additional subset of ATE1 enzymes. It has been shown that Cys-specific arginylation occurs in mammals but not in yeast; our finding of Cys-specific mammalian ATE1 forms suggests that Cys-arginylation may have additional importance and, thus, require an additional set of Arg-transferases.

Because *Ate1* knockout in mice results in embryonic lethality (15), there have been no studies of the role of protein arginylation in adult organisms. However, the Cys-specificity of ATE1-3 and ATE1-4 expressed predominantly in selected adult tissues suggests that Cys-specific protein arginylation has an added role in the functioning of these tissues that constitute some of the vital organs in the body and are also rich in ATE1-1 and ATE1-2. It would be interesting to study the role of ATE1 and protein arginylation in the adult organisms past embryonic and postnatal development, and its specific role in the functioning of the vital tissues and organs.

ATE1-3 and ATE1-4 differ from the previously identified Arg-transferases by a relatively small region of 30 amino acid residues of homologous sequence; however, this difference is apparently enough to induce inability to arginylate Asp- and Glu-containing substrates that are so efficiently arginylated by other ATE1 forms. This fact allows us to speculate about the determinants within the ATE1 molecule that specify substrate binding and/or Arg transfer reaction through changes in secondary or tertiary structure. It would be interesting to perform a detailed domain analysis and point mutagenesis to determine structural features and amino acid residues that define the universality of substrates arginylated by ATE1-1 and ATE1-2 and the narrow specificity of ATE1-3 and ATE1-4 for N-terminal Cys.

It has been suggested that Cys-arginylation requires its prior modification to cysteic acid, which makes it structurally similar to Asp and Glu (15) and that this modification occurs only in mammals, explaining the inability of Cys to be arginylated in yeast. We now find that, in addition to the possible modification of Cys itself, there are additional enzymes that ensure its efficient arginylation in mammals. Thus, Cys-arginylation likely plays an important regulatory role. It would be important to identify Cys-specific substrates in mammalian cells and to study the role of arginylation in their stability and functions.

We thank Dr. Cornelia Kurischko and Dr. Frank Luca for their help and guidance in setting up the yeast work, Dr. Ilia Davydov for advice about RGS4 degradation assays, Dr. Vladmir Rodionov and Dr. John Pehrson for helpful discussions, and Dr. Serge Fuchs for critical reading of the manuscript. We express special thanks to Professor Roald Hoffman for communicating our manuscript. This work was supported, in part, by the Abramson Cancer Center of the University of Pennsylvania Pilot Projects Program, American Cancer Society Grant IRG-78-002-27, and U.S. Department of Defense Congressionally Directed Medical Research Program Grant PC040372.

1. Kaji, A., Kaji, H. & Novelli, G. D. (1963) *Biochem. Biophys. Res. Commun.* **10,** 406–409.
2. Balzi, E., Choder, M., Chen, W. N., Varshavsky, A. & Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471.
3. Kaji, H. (1968) *Biochemistry* **7,** 3844–3850.
4. Soffer, R. L. (1973) *J. Biol. Chem.* **248,** 2918–2921.
5. Li, J. & Pickart, C. M. (1995) *Biochemistry* **34,** 15829–15837.
6. Li, J. & Pickart, C. M. (1995) *Biochemistry* **34,** 139–147.

7. Gonda, D. K., Bachmair, A., Wunning, I., Tobias, J. W., Lane, W. S. & Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712.
8. Kwon, Y. T., Kashina, A. S. & Varshavsky, A. (1999) *Mol. Cell. Biol.* **19,** 182–193.
9. Bongiovanni, G., Fidelio, G. D., Barra, H. S. & Hallak, M. E. (1995) *Neuroreport* **7,** 326–328.
10. Bongiovanni, G., Fissolo, S., Barra, H. S. & Hallak, M. E. (1999) *J. Neurosci. Res.* **56,** 85–92.
11. Chakraborty, G. & Ingoglia, N. A. (1993) *Brain Res. Bull.* **30,** 439–445.
12. Hallak, M. E., Bongiovanni, G. & Barra, H. S. (1991) *J. Neurochem.* **57,** 1735–1739.
13. Bachmair, A., Finley, D. & Varshavsky, A. (1986) *Science* **234,** 179–186.
14. Davydov, I. V. & Varshavsky, A. (2000) *J. Biol. Chem.* **275,** 22931–22941.
15. Kwon, Y. T., Kashina, A. S., Davydov, I. V., Hu, R. G., An, J. Y., Seo, J. W., Du, F. & Varshavsky, A. (2002) *Science* **297,** 96–99.
16. Burke, D., Dawson, D. & Stearns, T. (2000) *Methods in Yeast Genetics* (Cold Spring Harbor Lab. Press, Woodbury, NY).
17. Mumberg, D., Muller, R. & Funk, M. (1994) *Nucleic Acids Res.* **22,** 5767–5768.
18. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1996) *Current Protocols in Molecular Biology* (Wiley, New York).
19. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.