JOURNAL OF BACTERIOLOGY, July 1993, p. 4364–4374
0021-9193/93/144364-11$02.00/0

Vol. 175, No. 14

# The N-End Rule in *Escherichia coli*: Cloning and Analysis of the Leucyl, Phenylalanyl-tRNA-Protein Transferase Gene *aat*

THOMAS E. SHRADER,[1,2]* JOHN W. TOBIAS,[1,3] AND ALEXANDER VARSHAVSKY[1,4]

*Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139[1]; Department of Biochemistry, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461[2]*; Department of Microbiology and Molecular Genetics, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115[3]; and Division of Biology, California Institute of Technology, Pasadena, California 91125[4]*

Received 22 January 1993/Accepted 6 May 1993

**The N-end rule relates the in vivo half-life of a protein to the identity of its N-terminal residue. Distinct versions of the N-end rule operate in bacteria, fungi, and mammals. We report the cloning and analysis of *aat*, the *Escherichia coli* gene that encodes leucyl, phenylalanyl-tRNA-protein transferase (L/F-transferase), a component of the bacterial N-end rule pathway. L/F-transferase is required for the degradation of N-end rule substrates bearing an N-terminal arginine or lysine. The *aat* gene maps to the 19-min region of the *E. coli* chromosome and encodes a 234-residue protein whose sequence lacks significant similarities to sequences in data bases. In vitro, L/F-transferase catalyzes the posttranslational conjugation of leucine or phenylalanine to the N termini of proteins that bear an N-terminal arginine or lysine. However, the isolation and sequence analysis of a β-galactosidase variant engineered to expose an N-terminal arginine in vivo revealed the conjugation of leucine but not of phenylalanine to the N terminus of the β-galactosidase variant. Thus, the specificity of L/F-transferase in vivo may be greater than that in vitro. The *aat* gene is located ~1 kb from *clpA*, which encodes a subunit of ATP-dependent protease Clp. Although both *aat* and *clpA* are required for the degradation of certain N-end rule substrates, their nearly adjacent genes are convergently transcribed. The *aat* gene lies downstream of an open reading frame that encodes a homolog of the mammalian multidrug resistance P glycoproteins.**

Among the functions of intracellular proteolysis are the elimination of abnormal proteins, the maintenance of amino acid pools in cells during stresses such as starvation, and the generation of protein fragments that act as hormones, antigens, or other effectors. Yet another role of proteolytic pathways is to confer short half-lives on proteins whose concentrations must vary with time or alterations in the state of a cell. Thus, either a constitutive or a transient metabolic instability is a property of many regulatory proteins. Many other proteins, while long-lived as components of larger complexes, such as ribosomes and oligomeric proteins, are metabolically unstable as free subunits (for recent reviews, see references 15, 20, 22, 28, 34, 36, and 45).

Amino acid sequences, conformational determinants, or chemically modified protein structures that confer metabolic instability are called degradation signals, or degrons (44). An essential component of one degradation signal is the N-terminal residue of the protein (2). The presence of this signal, called the N-degron, is manifested as the N-end rule, which relates the in vivo half-life of a protein to the identity of its N-terminal residue. Distinct versions of the N-end rule have been shown to operate in all organisms examined, from mammals to bacteria (reviewed in reference 45).

In eukaryotes, linear ubiquitin (Ub) fusions to a test protein, such as β-galactosidase (βgal), are rapidly and precisely deubiquitinated by endogenous Ub-specific processing proteases (2). Deubiquitination occurs irrespective of the identity of a residue at the Ub-βgal junction and results in free Ub and a βgal variant bearing the previously junctional residue at its new N terminus. This in vivo method for generating different residues at the N termini of otherwise identical test proteins led to the discovery of the N-end rule in eukaryotes (2). Subsequent isolation of *UBP1*, the gene encoding one of the Ub-specific processing proteases of the yeast *Saccharomyces cerevisiae* (43), and expression of this enzyme in *Escherichia coli* made possible the use of Ub fusions to test for the presence of an N-end rule in bacteria, which lack Ub and Ub-specific enzymes.

Tobias and colleagues detected and defined the N-end rule in *E. coli* and also showed that the gene *aat* is required for the degradation of N-end rule substrates that bear N-terminal Arg or Lys, whereas all classes of N-end rule substrates (i.e., those bearing N-terminal Phe, Leu, Trp, Tyr, Arg, or Lys) are stabilized by disruptions of the gene *clpA* (42). The previously cloned *clpA* encodes a subunit of ATP-dependent protease Clp (also known as protease Ti; for reviews, see references 18 and 20). The *aat* locus was originally defined by use of a mutant that lacked the activity of leucyl, phenylalanyl-tRNA-protein transferase (L/F-transferase), an enzyme that conjugates Leu or Phe to Arg or Lys at the N termini of test substrates (40). The function of this enzyme, which is present in bacteria but not in eukaryotes, was revealed by the finding that N-end rule substrates bearing N-terminal Arg or Lys are short-lived in *E. coli* in the presence but not in the absence of L/F-transferase (42).

The N-end rule is thus organized hierarchically, with N-terminal Arg and Lys being secondary destabilizing residues, in that they function as substrates for the conjugation of either Leu or Phe, two of the four primary destabilizing residues (Leu, Phe, Trp, and Tyr) in the *E. coli* N-end rule. The N-end rule is also hierarchical in eukaryotes, but the details are quite different (19). For example, in the yeast *S.*

* Corresponding author.

*cerevisiae*, N-terminal Arg and Lys are primary destabilizing residues, while N-terminal Asp and Glu are secondary destabilizing residues, in that they function as substrates for the conjugation of Arg, one of the primary destabilizing residues in eukaryotes. The conjugating enzyme, arginyl-tRNA-protein transferase (R-transferase), is not found in bacteria (14).

The present work extends our initial findings about the *E. coli* N-end rule (42). First, we describe the cloning and analysis of the *aat* gene. Second, an N-end rule substrate, such as Arg-βgal (a β-gal variant engineered to expose an N-terminal arginine), is modified in vivo through the N-terminal conjugation of Leu but not of Phe. Third, *aat* is only ~1 kb away from the previously characterized, convergently transcribed *clpA* gene, which encodes a subunit of protease Clp, another component of the N-end rule pathway. Finally, the *aat* gene is part of a new operon that also encodes a homolog of the mammalian multidrug resistance (Mdr) P glycoproteins.

## MATERIALS AND METHODS

**Bacterial strains and media.** *E. coli* MC1061 [*hsdR mcrB araD139* Δ(*araABC-leu*)*7679* Δ*lacX74 galU galK rpsL thi*] was used as the parental strain throughout this work. Variants of this strain were constructed by P1 transduction (38). In particular, strain TS360 (MC1061 *recD1903*::mini*tet*), which allows targeted homologous recombination to be carried out by transformation with linear DNA fragments, was constructed from MC1061 by use of a P1*vir* lysate grown on strain DPB271 (*recD1903*::mini*tet*) (6). TS365 (MC1061 *pcnB80 zad*::mini*kan*), in which ColE1-like plasmids, such as pBR322, have a decreased copy number, was constructed by use of a P1*vir* lysate grown on strain MJC112 (*pcnB80 zad*::mini*kan*) (32; a gift from M. Carson, Harvard Medical School, Boston, Mass.). The liquid medium was Luria broth (1% Bacto Tryptone, 0.5% yeast extract [Difco], 1% NaCl). The solid medium contained 2% Bacto Agar (Difco). Antibiotics (Sigma) were added as needed to the following final concentrations: 100 μg of ampicillin, 10 μg of tetracycline, 34 μg of chloramphenicol, and 25 μg of kanamycin per ml. 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal; Boehringer), a chromogenic βgal substrate, was added as needed to 0.006%.

**Plasmids.** Plasmid pAS1 is a derivative of pBR322 (8) that was constructed by replacing the ~2-kb *Eco*RI-*Pvu*II fragment of pBR322 with a ~0.12-kb *Eco*RI-*Pvu*II fragment of pGEM2-Ava that contains a polylinker cloning region (37).

Each of the pUBP-X-βgal plasmids expressed both the Ub-specific processing protease Ubp1 of *S. cerevisiae* and 1 of the 20 Ub-X-βgal fusion proteins (where X is any amino acid) (see above). This set of plasmids was constructed by isolating the ~5.3-kb *Sca*I fragments of pUB23-X plasmids (2), which encode the Ub-X-βgal fusion proteins, ligating *Bgl*II linkers to the ends of the fragments, and inserting each of the fragments into the *Bam*HI site of plasmid pJT184, which encodes Ubp1 (43). The resulting pUBP-X-βgal plasmids expressed Ub-X-βgal and Ubp1 from yeast promoters that are weakly active in *E. coli* (42). These plasmids contained the p15A replicon and a *cat* gene from plasmid pACYC184 that conferred, respectively, compatibility with ColE1 replicon-based plasmids and resistance to chloramphenicol (11). The construction of other plasmids is described below.

**Transposon mutagenesis.** Mini*tet* transposon insertions into the *E. coli* chromosome that inhibited the degradation of normally short-lived Arg-βgal were generated by the "λ-hop" method (46). In brief, MC1061 cells carrying the pUBP-R-βgal plasmid were infected with λNK1098 (46), plated on medium containing tetracycline, and screened for tetracycline-resistant (Tet$^r$) colonies. MC1061 is *sup*$^0$ and therefore does not support the replication or integration of λNK1098, which carries a *P*am80 amber mutation in its *P* gene and the *b*522 deletion, which removes the phage attachment site. Thus, a Tet$^r$ colony would be derived from a cell in which the mini*tet* transposon of λNK1098 has "hopped" from the phage into the *E. coli* chromosome. Mini*tet* transposon insertions that inhibit Arg-βgal degradation result in elevated intracellular levels of Arg-βgal; such cells form blue colonies in the presence of X-Gal (see Results and reference 42). Approximately 15,000 Tet$^r$ colonies were screened for high Arg-βgal levels on X-Gal plates. Putative Arg-βgal degradation mutants were tested further by curing the cells of pUBP-R-βgal through growth in the absence of chloramphenicol and retransforming cured cells with independently isolated pUBP-X-βgal plasmids.

**Identification of mini*tet* insertion sites.** The insertion sites were determined by direct sequencing of mini*tet*-*E. coli* DNA junctions. Total DNA from mutants TS351 and TS357, which were defective in the degradation of Arg-βgal (see Results), was prepared (38) and digested with *Bam*HI. Since *Bam*HI does not cut within the mini*tet* sequence, complete digestion yields a mini*tet*-containing DNA fragment flanked on both sides with *E. coli* DNA. Mini*tet*-containing fragments were isolated by ligating the *Bam*HI-produced *E. coli* DNA fragments into *Bam*HI-digested pAS1, transforming the resulting library into TS365, and selecting for Tet$^r$ transformants. TS365 contains a mutation in the *pcnB* gene that lowers the copy number of ColE1-like plasmids such as pBR322 and pAS1 (32) and therefore allows the use of the Tn*10*-derived *tet* gene, which is lethal when expressed from a multicopy plasmid in wild-type cells (29). For preparation of larger quantities of mini*tet*-*E. coli* DNA junctions for sequencing, mini*tet*-containing *Bam*HI fragments were digested with *Eco*RI, which cuts within the *tet* gene, yielding fragments that were no longer lethal at a high copy number. These fragments were ligated into *Bam*HI-*Eco*RI-cut pAS1 and transformed into MC1061. Plasmid DNA was isolated from transformants by the alkaline lysis procedure (1). Nucleotide sequencing was carried out by the chain termination method with standard procedures (1) and primers complementary to the ends of the mini*tet* element (47). The sequences of the actual mini*tet* insertion sites, 5′-CACA CAGCA-3′ (TS351 mutant) and 5′-CGCAAAGCA-3′ (TS357 mutant), were similar to the consensus sequence for Tn*10* insertions, 5′-NGCTNAGCN-3′ (29). These sequences are identical to positions 2363 to 2371 of Fig. 1, which lie within the *aat* gene (see below), and to positions 2773 to 2781 of the previously published sequence of the *clpA* gene (21); the latter positions correspond to positions 4592 to 4584 in the numbering scheme of Fig. 1.

**Mapping, cloning, and sequencing of the *aat* gene.** Hfr and P1 mapping of the class I and II mini*tet* insertions (see Results), which inhibited the degradation of Arg-βgal, was performed as described previously (38) with mutants that carried pUBP-R-βgal. Inclusion of this plasmid allowed for easy monitoring of the ability of a particular Hfr donor strain or P1*vir* lysate to correct the mutant cell's defect in Arg-βgal degradation. The restoration of Arg-βgal degradation by Hfr mating or P1 transduction resulted in the appearance of white colonies on X-Gal plates. Polymerase chain reaction mapping was carried out with a GeneAmp kit (Perkin-Elmer

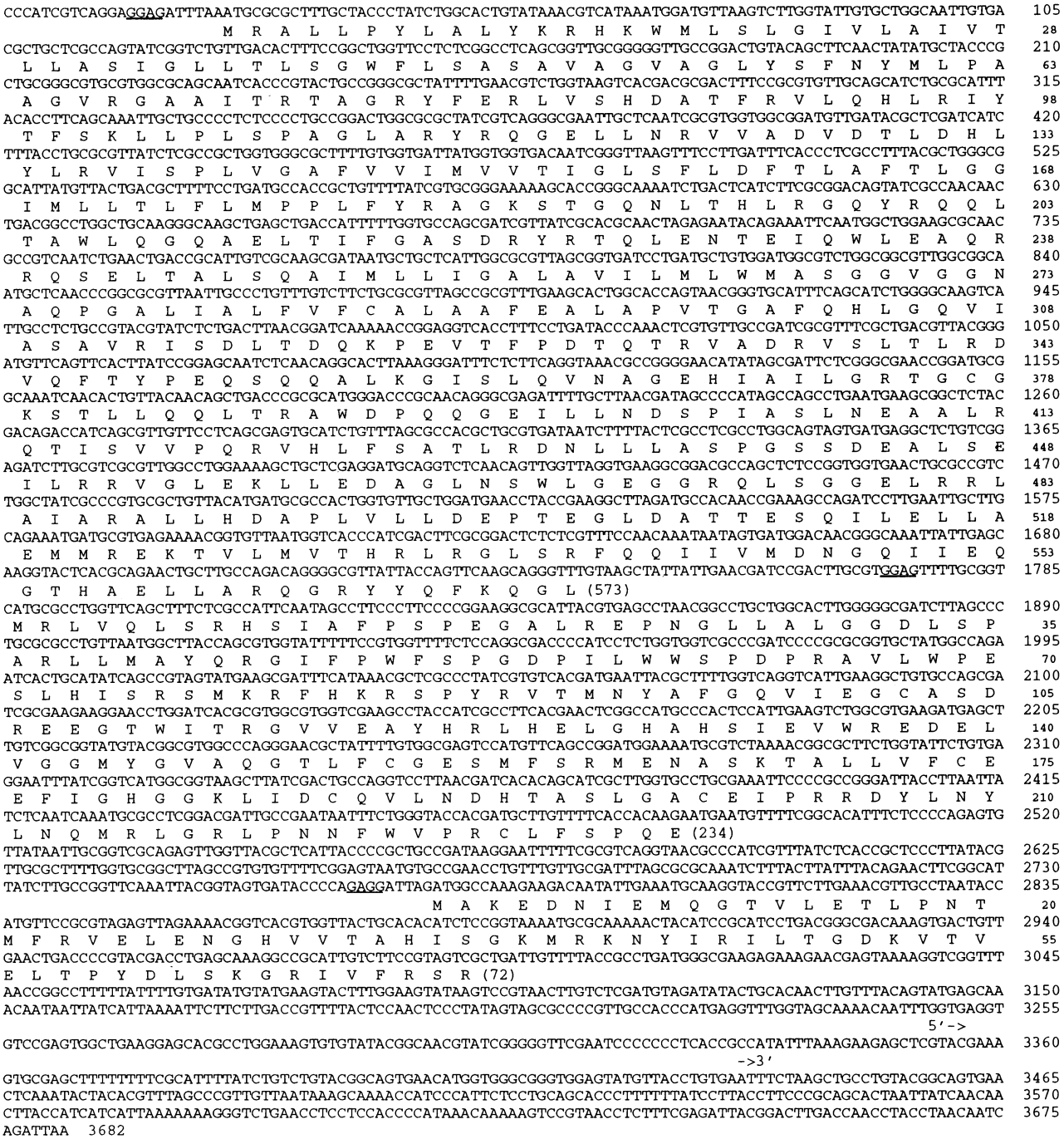

FIG. 1. Nucleotide sequence of a ~3.5-kb region of the *E. coli* chromosome at the 19-min (~930-kb) map position and deduced amino acid sequences of the corresponding proteins. The nucleotide sequence is presented in the direction opposite that in the standard genetic map to indicate the deduced translation products. The sequence shown contains the *mdrH* (24 to 1745), *aat* (1787 to 2491), *infA* (2776 to 2994), and *serW* (3213 to 3335) genes. Nucleotides 3679 to 3681 of this sequence correspond to the stop codon of the convergently transcribed *clpA* gene (21). The putative Shine-Dalgarno sequences for *mdrH*, *aat*, and *infA* are underlined. Nucleotide sequence data for a portion of this region have been published elsewhere (12, 21). Our data differ from those published in reference 12 at positions 3585 (no nucleotide in reference 12), 1583 (C in reference 12), and 1620 (no nucleotide in reference 12). These differences result in an alteration of the deduced amino acid sequence of the MdrH protein encoded by this region. In addition, our sequence data differ from those reported in reference 21 at positions 3585 (no nucleotide in reference 21) and 3602 (no nucleotide in reference 21).

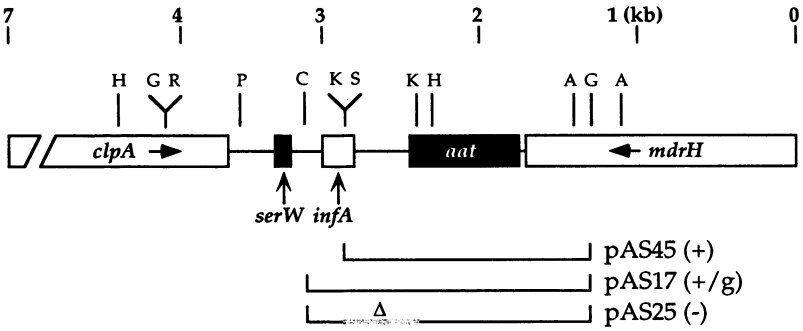


FIG. 2. Diagram of the *aat* region of the *E. coli* chromosome. Also shown are the genomic DNA fragments that complemented the Arg-βgal degradation defect in TS351 cells and a noncomplementing clone that contains an internal deletion. Abbreviations: H, *Hin*dIII; G, *Bgl*II; R, *Eco*RI; P, *Pst*I; C, *Sca*I; K, *Kpn*I; S, *Ssp*I; A, *Ava*I; (+), complementation of the Arg-βgal degradation defect; (−), lack of complementation; (+/g), complementation with a small-colony phenotype.

Cetus) and genomic DNA isolated from strains TS351 and TS353, which contained distinct mini*tet* insertions in the *aat* gene. One polymerase chain reaction primer (5′-GCCG GATCCGTAATCTCGAAAGAGG-3′) was complementary to the 3′ end of the *clpA* gene (21); the other primers were complementary to unique sequences near the ends of the mini*tet* sequence (5′-GCCGGATCCATTGCTGTTGAC-3′ and 5′-GCCGGACCCTTTGGTGACCAACGC-3′) (47). A ~6-kb *Eco*RI genomic DNA fragment containing the 3′ end of the *clpA* gene and downstream sequences was isolated from phage 213 (1H1) of the Kohara library (30) and cloned by use of the pAS1 cloning vector. This *Eco*RI fragment is toxic in wild-type cells, so the initial cloning and subcloning steps were carried out with strain TS365, in which the copy number of these ColE1-based plasmids is reduced to nontoxic levels (32). After construction in strain TS365, subclones were tested for toxic effects by transformation into MC1061 cells. The region of the *Eco*RI fragment shown in Fig. 1 was sequenced on both strands by the chain termination method (1).

**Subcloning and deletion analysis of the *aat* gene.** A series of deletions near the 5′ end of the *aat* open reading frame (ORF) were produced by published procedures (1) with minor modifications. In brief, the 1,711-bp *Bgl*II-*Sca*I fragment that contained the entire *aat* gene (Fig. 2) was ligated to *Bam*HI-*Sma*I-digested pGEM2 (Promega), yielding pAS17. This plasmid (10 μg) was linearized with *Ava*I and *Pst*I to yield a fragment with one 5′ overhang and one 3′ overhang (see Fig. 4A). This fragment was treated with exonuclease III (New England Biolabs) for 20, 30, 60, 105, 135, and 180 s to produce a nested set of deletions of the 5′ strand that encompassed the 5′ end of the *aat* ORF. The resulting 3′ overhangs and the 3′ overhang generated by *Pst*I were removed by treating the samples sequentially with S1 nuclease and Klenow polymerase I (Boehringer). The DNA was purified to remove nuclease buffers by use of GeneClean (Bio 101, Inc.), resuspended in TE buffer (1 mM EDTA, 10 mM Tris-HCl [pH 8.0]), and circularized by use of T4 DNA ligase (Bethesda Research Laboratories). Ligation mixtures from different exonuclease III time points were cotransformed into *E. coli* TS351 (MC1061 *aat-2*::mini*tet*) together with pUBP-R-βgal. Transformation mixtures were plated on a solid medium containing chloramphenicol, ampicillin, tetracycline, and X-Gal. The transformation mixture from the 30-s exonuclease III digestion had a particularly convenient mixture of white and blue colonies. Plasmid DNA was prepared by alkaline lysis (1) from 13 transformants derived from the 30-s time point, and the 5′ region of *aat* in these isolates was sequenced by use of a primer complementary to the T7 polymerase promoter in the polylinker of pGEM2.

A second deletion series was produced by use of a plasmid in which a strong transcriptional terminator ($Tl_4$) was positioned immediately upstream of the *aat* gene, thereby reducing the probability that a cryptic transcript originating in the vector DNA would complicate attempts to determine the minimum upstream regions required for the expression of L/F-transferase. Plasmid pAS44 was constructed by inserting a 2,967-bp *Sca*I-*Eco*RI fragment from plasmid pRS415 (39) into ScaI-EcoRI-cut pGEM2. This fragment contained four tandem copies of the strong transcriptional terminator from the *E. coli* rRNA operon (9). The pAS45 vector used in the second deletion series was constructed by inserting the *Bgl*II-*Ssp*I fragment that contained the *aat* gene (Fig. 2) into *Bam*HI-*Pvu*II-cut pAS44. For the deletion series, pAS45 was digested with *Sac*I and *Ava*I, creating a linear fragment with a 3′ overhang immediately downstream of the $Tl_4$ terminator and a 5′ overhang 389 bp upstream of the *aat* start codon. Deletions from the 5′ overhang were made, and the resulting fragments were recircularized as described above. Ligation products were transformed into TS351 containing pUBP-R-βgal as described above.

Plasmids pAS31 and pAS32 were constructed by inserting the *Pvu*II (a site within pGEM2)-*Ssp*I (position 2,791 in Fig. 1) fragment from deletion clones W1 and W3 (see Fig. 4A) into the *Eco*RV site of pBR322. Plasmid pAS25 was constructed from pAS17 by deleting the 346-bp *Kpn*I fragment that spans the region from the 3′ end of the *aat* gene to the 5′ end of the *infA* gene (see Fig. 4A).

**N-terminal sequence analysis of an N-end rule substrate.** Isolates of *E. coli* JT111 (*lacI*$^q$ *clpA14*::mini*tet*) were cotransformed with a pKKUb-X-βgal plasmid (where X is any amino acid) that expressed a Ub-X-βgal fusion from the powerful $P_{trc}$ promoter and with plasmid pJT184, which encoded the Ub-specific protease Ubp1 of *S. cerevisiae*. X-βgal proteins (where X is any amino acid) were purified by APTG-Sepharose chromatography as described previously (19). The purified X-βgal proteins were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and transferred to ProBlot membranes (Applied Biosystems) in accordance with the manufacturer's recommendations. N-terminal sequencing of X-βgal proteins (five cycles of Edman degradation) was performed by use of an Applied Biosystems model 477 protein sequencer at the Massachusetts Institute of Technology Biopolymers Laboratory.

**Nucleotide sequence accession number.** The nucleotide sequence reported in this paper has been submitted to the GenBank-EMBL data bank and assigned accession number L10383.

## RESULTS

**Screen for mutations in the N-end rule pathway.** We used transposon mutagenesis to identify genes whose inactivation perturbs the N-end rule pathway in *E. coli*. Mutagenesis was performed with a strain that expressed both the Ub-Arg-βgal fusion protein and Ubp1, one of the Ub-specific processing proteases of *S. cerevisiae* (3, 43) (see Materials and Methods). When expressed in *E. coli*, the yeast Ubp1 protease efficiently and precisely deubiquitinates Ub-Arg-βgal and the other Ub-X-βgal fusions, Ub-Pro-βgal being the single exception (43). These Ub fusions were not deubiquitinated in *E. coli* lacking yeast Ubp1, and the unprocessed fusions were invariably long-lived. In contrast, the X-βgal proteins produced from Ub-X-βgal fusions in Ubp1-expressing *E. coli* were either long-lived or unstable, depending on the identity of their N-terminal residues. These measurements, carried out with 19 Ub-X-βgal fusions, yielded the N-end rule in *E. coli* (42).

Arg-βgal bears one of the two secondary destabilizing N-terminal residues in the *E. coli* N-end rule (secondary destabilizing residues require N-terminal conjugation of a primary destabilizing residue prior to substrate degradation). Therefore, Arg-βgal could serve as a reporter for detecting changes in the activity of the N-end rule pathway toward substrates bearing either secondary or primary destabilizing N-terminal residues. Transposon insertions into the *E. coli* chromosome that inhibited the degradation of Arg-βgal were generated by Tn*10* mutagenesis by the λ-hop method (46). In this procedure, cells are infected with a nonreplicating, nonintegrating λ phage carrying the Tn*10* transposase and a short (~2-kb), Tn*10*-derived transposable element (mini*tet*) containing the tetracycline resistance gene of Tn*10* (46). A hop of mini*tet* results in a tetracycline-resistant ($Tet^r$) colony. Among such $Tet^r$ cells may be mini*tet* insertion mutants unable to degrade Arg-βgal. Such mutants are identifiable as those that form blue colonies on plates containing X-Gal, a chromogenic indicator of βgal activity. Cells that degrade Arg-βgal efficiently do not accumulate significant steady-state levels of βgal and therefore form white colonies on X-Gal plates (5, 42).

Approximately 1 in 1,500 mini*tet* insertions rendered *E. coli* MC1061 unable to degrade Arg-βgal. Examination of the ability of the mutant to degrade Leu-βgal, which bears one of the primary destabilizing N-terminal residues (requiring no modification prior to substrate degradation) divided these mini*tet* insertions into two classes. Class I mutations failed to degrade both types of N-end rule substrates, represented by Arg-βgal and Leu-βgal, while class II mutants were unable to degrade Arg-βgal but were unimpaired in the degradation of Leu-βgal.

**Insertions that stabilize both Arg-βgal and Leu-βgal reside in the *clpA* gene.** Genomic DNA fragments containing a mini*tet* insertion from class I mutant TS357 were isolated and sequenced as described above. A comparison of the *E. coli* DNA adjacent to mini*tet* in mutant TS357 with sequences in data bases showed that the transposon resided within previously analyzed *clpA*. The *clpA* gene encodes one of the two subunits of ATP-dependent protease Clp (26, 27). We conclude, in agreement with earlier findings (42), that the Clp protease is essential for the degradation of N-end rule substrates.

**Insertions that stabilize Arg-βgal but not Leu-βgal reside in the *aat* gene.** A mini*tet* insertion from mutant TS351 that stabilized Arg-βgal but not Leu-βgal was found to reside within a previously sequenced region of the *E. coli* chromosome; we also sequenced this region, as well as an adjoining, previously unsequenced ~1.5-kb stretch (see Materials and Methods, Fig. 1, and reference 12). Our earlier work (42) suggested that the insertion in TS351 inactivated a gene for L/F-transferase, a biochemically characterized enzyme that conjugates Leu or Phe to N termini of proteins or peptides bearing N-terminal Arg or Lys (40). Indeed, an *E. coli* mutant (MS845 *aat-1*) that lacked L/F-transferase activity (41) was recently shown to have an N-end rule defect indistinguishable from that in the mini*tet*-generated class II mutant in the present work (stabilization of Arg-βgal but not of Leu-βgal) (42).

Mapping experiments with both Hfr mating and P1 transduction demonstrated that the sites of our class II mini*tet* insertions, such as the one in TS351, were indeed extremely close to the site of an *aat* mutation identified earlier (MS845 *aat-1* produced by chemical mutagenesis) in a gene termed *aat* by Soffer and Savage (41). These sites and the *clpA* gene both mapped at ~19 min in the *E. coli* genome (see Materials and Methods). Southern hybridization to genomic DNA from class I (TS354 and TS357) and class II (TS351 and TS353) mutants with mini*tet* as a probe indicated that all insertions resided in the same ~10-kb *Bam*HI fragment (data not shown). Furthermore, polymerase chain reaction mapping with genomic DNA isolated from the TS351 and TS353 (class II) mutants and oligonucleotide primers complementary to the 3′ end of *clpA* and to the *tet* gene of mini*tet* (see Materials and Methods) demonstrated that the transposon insertions of TS351 and TS353 were within ~1.5 kb of the 3′ end of *clpA* (data not shown).

**aat gene.** DNA fragments encompassing *clpA* and the downstream region of the *E. coli* genome were isolated from phages 212 and 213 of the Kohara library (30), and an ~3.5-kb region downstream of *clpA* was sequenced. The mini*tet* insertion in mutant TS351 that resulted in the inhibition of Arg-βgal degradation resided within an ORF of a gene (*aat*) that encoded a neutral (calculated pI, 7.2) 234-residue protein (Fig. 1) whose molecular mass was close to the reported molecular mass of the isolated L/F-transferase (40). The position of the start (ATG) codon in *aat* was inferred so as to yield the longest ORF. This start codon was preceded by a moderately strong Shine-Dalgarno sequence, GGAG (Fig. 1).

The deduced amino acid sequence of the Aat protein lacks significant similarities to the sequence of Ate1, the catalytically similar R-transferase of *S. cerevisiae* shown previously to be a component of the N-end rule pathway in eukaryotes (4). As expected for an enzyme that interacts with RNA, the *aat*-encoded L/F-transferase contains arginine-rich regions; however, it lacks known RNA-binding motifs, such as those seen in transcriptional termination factors (31). In addition, the sequence of Aat lacks "signature" sequences of aminoacyl-tRNA synthetases (10) and also is not similar to other sequences in the current data bases. Thus, it is likely that L/F-transferase recognizes tRNA moieties of Leu-$tRNA^{Leu}$ and Phe-$tRNA^{Phe}$ in a novel manner. As described below, the ~1-kb DNA fragment containing exclusively the *aat* ORF complemented the Arg-βgal degradation defect in both the TS351 mutant (mini*tet* insertion) and the original *aat* mutant (produced by chemical mutagenesis) (41). This re-

sult, together with indistinguishable map positions of the original *aat* mutation and mini*tet* insertions in our class II mutants (see above), indicated that the *aat* gene of Soffer and Savage (41) and the gene whose inactivation yielded class II mutants in the present work are one and the same gene.

**aat operon.** Although *clpA* and *aat* are located less than 1 kb apart and encode components of the same pathway, they were found to reside in different operons. As shown in Fig. 2, *clpA* and *aat* are separated by previously characterized gene *infA*, which encodes translational initiation factor Infl, and by the *serW* gene, which encodes $tRNA^{Ser}$. There are putative transcriptional terminators between the 3′ end of *infA* and the 5′ end of *serW* (12) and between the 3′ end of *serW* and the 3′ end of the convergently transcribed *clpA* gene (21). Thus, the two N-end rule genes, *aat* and *clpA*, while nearly adjacent, are oriented in opposite directions and are not coupled transcriptionally.

A 1.7-kb region immediately upstream of the *aat* gene was sequenced and found to contain an ORF encoding a 573-residue protein. The amino acid sequence of this protein is similar to the sequences of Mdr P glycoproteins found in mammalian cells and also to the sequences of other members of the ATP-binding cassette family of membrane transporters found in other species. A comparison between the deduced amino acid sequences of this *E. coli* protein (tentatively named MdrH, for Mdr homolog) and the CyaB protein from *Bordetella pertussis* (17) is shown in Fig. 3. *B. pertussis* is the etiologic agent of whooping cough; the *cyaB* gene encodes an Mdr homolog required for export of the calmodulin-sensitive adenylate cyclase-hemolysin protein. This bifunctional secreted protein (encoded by the *cyaA* gene) contributes to elevated levels of intracellular cyclic AMP found in susceptible eukaryotic cells exposed to *B. pertussis*. The deduced amino acid sequences of the *E. coli* MdrH and *B. pertussis* CyaB proteins show significant similarities in their C-terminal halves (Fig. 3). In addition, the N-terminal half of each protein contains several stretches of hydrophobic residues that are likely to be membrane-spanning segments. The ATP-binding motif (23) resides in the C-terminal half of each protein. The sequence of MdrH is more similar to that of CyaB than it is to the sequences of bacterial importers, such as histidine permease HisP or oligopeptide permease OppF (24), suggesting that MdrH is involved in peptide or protein export from *E. coli*. Transcriptional coupling (see below) of the genes encoding L/F-transferase and the putative transporter MdrH is consistent with the possibility that L/F-transferase may also have a function outside the N-end rule pathway.

The *mdrH* and *aat* genes are separated by only 42 bp (Fig. 1). This intergenic region does not contain palindromic DNA sequences capable of forming relatively stable hairpins characteristic of transcriptional terminators. Furthermore, the only run of T residues in this intergenic region, a second common feature of transcriptional terminators (48), is present between the Shine-Dalgarno sequence and the start codon of the *aat* gene, an unlikely position for a transcriptional terminator. Deletion analyses, described below, confirmed that this region allows readthrough transcription and corroborated earlier data from S1 nuclease mapping experiments carried out during studies of the transcriptional regulation of the *infA* gene (12), which is located between *aat* and *clpA* (Fig. 2). These studies demonstrated that the *aat* gene (referred to as ORF-27 in reference 12) was transcriptionally coupled to an upstream region that contained the *mdrH* ORF of the present work. In contrast, the *infA* gene was shown to be expressed from two moderately strong promoters located downstream of *aat*. As a result, the *infA* gene is transcribed at a much higher level than its upstream neighbors (12). We conclude that the *aat* gene is the last gene of a new operon that contains at least one other gene.

**Evidence for an intraoperon promoter upstream of the *aat* gene.** The transcriptional coupling of *aat* to *mdrH* suggests that the physiologically relevant promoter for L/F-transferase expression resides far upstream of the *aat* gene. However, ColE1 plasmid-based clones containing either the *Bgl*II (1366)-*Sca*I (3077) or the *Bgl*II (1366)-*Ssp*I (2791) fragment and encoding the entire *aat* gene but only a fraction of the upstream *mdrH* gene were found to complement the Arg-βgal degradation defect of TS351 cells. Furthermore, the sequence immediately upstream of the *aat* ORF (Fig. 1) resembles the consensus sequence of a $\sigma^{70}$-dependent *E. coli* promoter (25), suggesting that *aat* could be expressed from either a polycistronic mRNA (see above) or a monocistronic message initiated immediately upstream of *aat*. We used exonuclease III digestion to define the sequences that are essential for the above-described complementation.

The extent of exonuclease III-mediated deletions and the appearance of colonies of TS351(pUBP-R-βgal) cells transformed with the recircularized deletion products are shown in Fig. 4. A reduced colony size was observed for cells transformed with the majority of the clones produced in this deletion series (Fig. 4A). The growth defect resulted from the inclusion of a complete copy of the *infA* gene on a multicopy plasmid (see below). Curiously, previous studies done with a different genetic background revealed no growth defect in cells that overexpressed the Infl protein (12). The origin of this discrepancy remains to be determined.

The results from the initial deletion series were unexpected in that clones W2 and W3, which complemented fully TS351 cells for Arg-βgal degradation, and clones M7 to M12, which complemented partially, were all shorter than the complete *aat* ORF (Fig. 4A). A resolution of this paradox is suggested by noting the reading frames of active or partially active clones in Fig. 4A. Deletion clones W2, M7, M8, M9, and M11 all started in the reading frame identical to the *aat* OFR, while clones W3, M10, and M12 started in the −1 reading frame. Conversely, clone B13, which showed no complementation, contained more upstream DNA than the partially complementing clones but started in the +1 reading frame. We therefore suspected that the expression of functionally active L/F-transferase from these truncated *aat* fragments was dependent on plasmid-encoded transcription and translation signals and that a cryptic Shine-Dalgarno sequence and start codon existed only in the 0 and −1 reading frames.

The results of other *aat* expression tests were consistent with this explanation. Specifically, the *Pvu*II (pGEM2 polylinker)-*Ssp*I (2791) fragments from clones W1 and W3 were isolated and cloned into the *Eco*RV site of plasmid pBR322. This approach has the potential to position the *aat* ORF in either orientation inside the *tet* gene of pBR322. However, no clones in which the *aat* gene was positioned in the sense orientation with respect to the *tet* gene were recovered for either fragment. Isolation and retransformation of the resulting plasmids into TS351 cells that carried pUBP-R-βgal resulted in white colonies for the W1-derived pAS31 vector (complementation of the Arg-βgal degradation defect) and blue colonies for the W3-derived pAS32 vector (lack of complementation; see Materials and Methods). We conclude that the entire *aat* ORF and more than 30 bp of upstream DNA are required for L/F-transferase expression in this

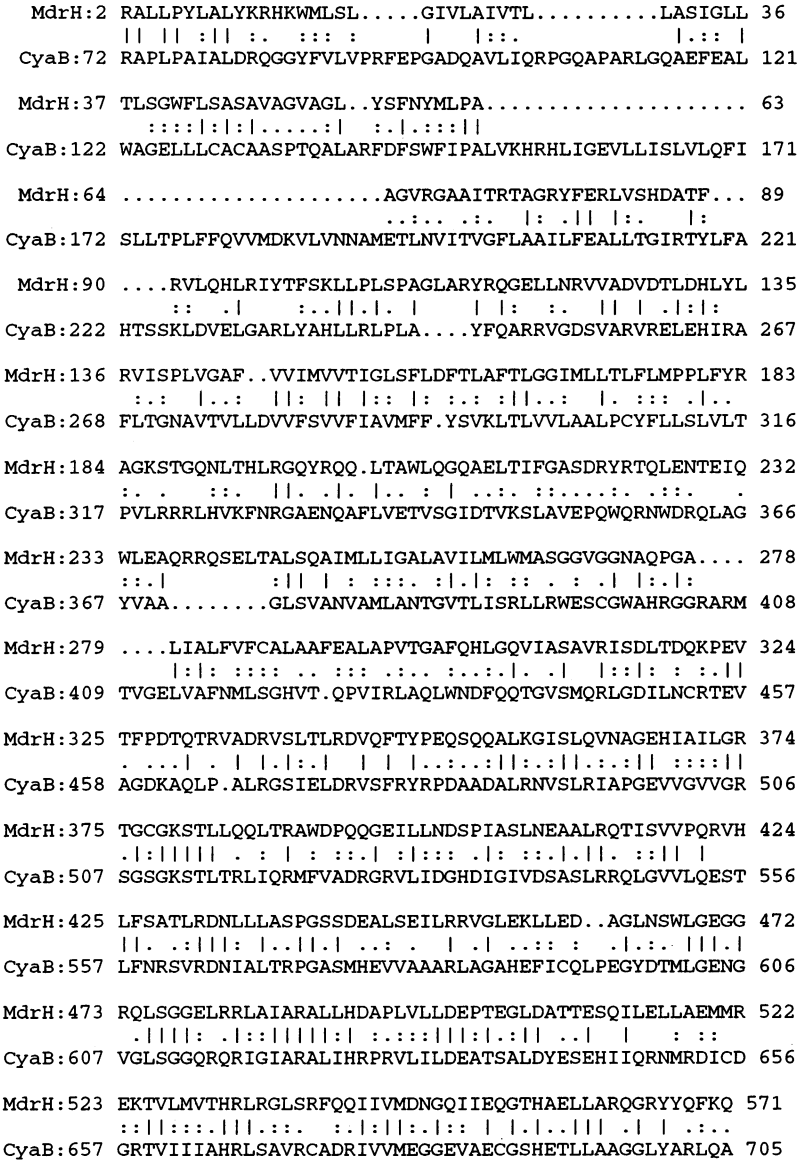


FIG. 3. Comparison of the deduced amino acid sequence of the MdrH protein (this work) with the deduced sequence of the CyaB protein from *B. pertussis* (17). The sequence of the MdrH protein contains several highly hydrophobic stretches (residues 16 to 57, 96 to 112, 137 to 158, 160 to 182, and 249 to 272) characteristic of membrane-spanning regions. The ATP-binding motif (LSGGXXXRLXIA) (23) is present in the C-terminal portion of MdrH (residues 475 to 486).

environment. Furthermore, the complete absence of clones in which the *aat* gene was positioned in the same orientation as the disrupted *tet* gene (data not shown) suggested that overexpression of L/F-transferase from multiple copies of the pBR322 $P_{tet}$ promoter is deleterious to the cell. Interestingly, the ability of plasmids expressing the M7 or M11 fragment of *aat* to complement partially the Arg-βgal degradation defect in TS351 cells (Fig. 4A) indicated that a deletion of the N-terminal ~15% of L/F-transferase does not completely abolish its enzymatic activity.

A second deletion series was created to define more precisely the sequences required for L/F-transferase expression. In this series, four copies of the transcriptional terminator from the *E. coli* rRNA operon (9) were positioned immediately upstream of the deletion clones to reduce the influence of the plasmid-based expression signals that complicated interpretation of the data obtained with the first deletion series. As shown in Fig. 4B, clones as long as $Tl_4$-2 and $Tl_4$-6 did not complement the TS351 mutant, while clone $Tl_4$-18 (which contained the complete *aat* ORF and 83 bp of the upstream sequence) resulted in full complementation, as judged by colony color on X-Gal plates. The constraints provided by these deletion data, as well as similarities between the known $\sigma^{70}$-dependent *E. coli* promoters and the ~50-bp region immediately upstream of the *aat* ORF, indicated that this region functioned as a promoter in plasmid-based constructs. The relative contribution of this intraoperon promoter to the in vitro expression of L/F-transferase remains to be determined.

**The 3′ end of *aat* and the *infA* gene.** Transformants carrying

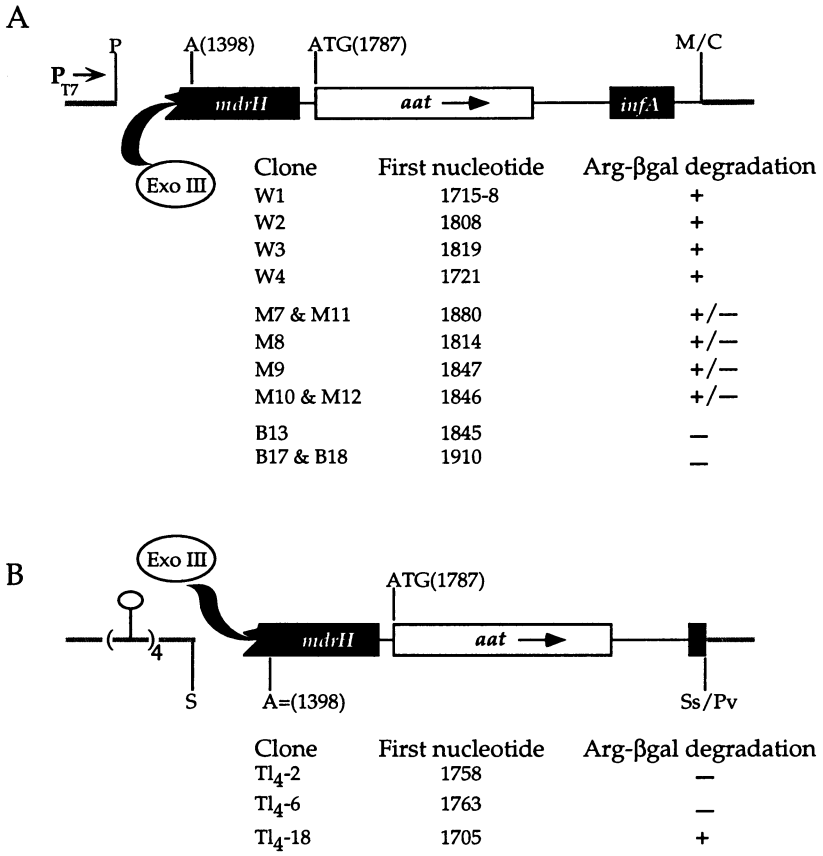


FIG. 4. Deletion constructs used to determine the sequences required for the expression of L/F-transferase. The initial plasmids were pAS17 (A) and pAS45 (B). For each deletion product, the first nucleotide of the *E. coli* DNA insert not deleted by exonuclease III (Exo III) and the ability of the recircularized plasmid to complement the Arg-βgal degradation defect in TS351 cells are indicated. In each case, treatments with exonuclease III deleted sequences upstream of the *aat* gene, while the opposite end of the fragment was protected by its 3′ extension. Abbreviations: P, *Pst*I; A, *Ava*I; ATG, start codon of the *aat* gene; M, *Sma*I; C, *Sca*I; Ss, *Ssp*I; Pv, *Pvu*II.

the plasmids generated in the first deletion series (see above and Fig. 4A) formed smaller colonies than transformants carrying either the plasmids generated in the second deletion series or plasmid pAS31 (or pAS32), which had *aat* inserted within the *tet* gene. The growth deficiency caused by the first series of plasmids stemmed from the presence of the *infA* gene in the *Bgl*II (1366)-*Sca*I (3077) fragment (Fig. 1) used to isolate the deletion fragments shown in Fig. 4A. Indeed, plasmids such as pAS45, derived from the *Bgl*II (1366)-*Ssp*I (2791) fragment, which encoded only an N-terminal region of the Infl protein, did not confer a slow-growth phenotype (data not shown). Furthermore, transformation of TS351 (*aat-2*::mini*tet*) with pAS25, which was produced from pAS17 (see Materials and Methods) by deletion of its 346-bp *Kpn*I (2458 to 2806) fragment (Fig. 1), yielded large blue colonies on X-Gal plates. The latter deletion removed the sequences encoding the last 8 residues of L/F-transferase through those encoding the first 11 residues of Infl (Fig. 1 and data not shown). Thus, multiple copies of the *infA* gene appear to be deleterious, at least in the backgrounds of the tested *E. coli* strains. In addition, the lack of complementation by plasmid pAS25 showed that even small deletions at the C terminus of L/F-transferase abolish its activity.

**In vivo modification of Arg-βgal.** Previous evidence that the degradation of the engineered N-end rule substrate Arg-βgal proceeds through a (Phe or Leu)-Arg-βgal intermediate was largely indirect. A functional *aat* gene is required for the degradation of Arg-βgal and Lys-βgal in *E. coli*, whereas the degradation of Phe-βgal and Leu-βgal is independent of *aat* (42). To test directly for the N-terminal modification of an N-end rule substrate, Arg-βgal was overexpressed in and purified from *E. coli* JT111 cells (see Materials and Methods). This strain is *aat*$^+$ but contains a *clpA14*::mini*tet* mutation that blocks the degradation of all N-end rule substrates (42) and thus allows their overexpression. The results of N-terminal sequencing of the isolated Arg-βgal confirmed the role of L/F-transferase in the in vivo degradation of Arg-βgal but were also unexpected in several respects. In the presence of isopropyl-β-D-thiogalactopyranoside (IPTG), the initial fusion, Ub-Arg-βgal, was expressed to near toxic levels from the $P_{trc}$ promoter. Nonetheless, the bulk of Ub-Arg-βgal was deubiquitinated by the Ubp1 protease (expressed from plasmid pJT184; 43). Moreover, ~70% of the resulting Arg-βgal was modified by L/F-transferase, a nonabundant enzyme (14) produced from a scarce mRNA (12). In contrast to the results of in vitro studies with partially purified L/F-transferase, which catalyzed the efficient conjugation of either Phe or Leu (and the less efficient conjugation of Met) to the N termini of test substrates (14), the N-terminal modification of Arg-βgal in vivo was found to involve exclusively the conjugation of Leu (Materials and Methods and data not shown). Thus, the in

vivo selectivity of L/F-transferase is greater than that observed with the isolated enzyme.

## DISCUSSION

We have extended our previous observation that L/F-transferase is a component of the N-end rule pathway in *E. coli* (42). The involvement of L/F-transferase in proteolysis mirrors a similar involvement of R-transferase in the N-end rule pathway of eukaryotic cells. In either pathway, the N-terminal conjugation of a primary destabilizing amino acid serves as a signal for the degradation of the modified protein.

Although each of these aminoacyl-tRNA-protein transferases catalyzes the posttranslational N-terminal aminoacylation of acceptor proteins, their amino acid sequences are dissimilar from one another and from other sequences in current data bases. Furthermore, bacterial L/F-transferase is about half the size of yeast R-transferase (234 versus 503 residues) (4); it is also considerably smaller than the smallest aminoacyl-tRNA synthetase (10). The small size of *E. coli* L/F-transferase is particularly noteworthy in light of the demonstrated ability of the enzyme to bind specific aminoacyl-tRNAs, to discriminate between cognate and noncognate aminoacyl residues in a mischarged aminoacyl-tRNA, to recognize specific N-terminal residues in substrates of L/F-transferase, and to catalyze peptide bond formation (40). On the basis of our deletion analysis, the enzymatic activity of L/F-transferase depends only weakly on the presence of the N-terminal ~15% of its sequence, whereas even small C-terminal deletions abolish the activity.

The in vivo specificity of L/F-transferase differs from that determined previously with a partially purified enzyme or whole-cell extracts. In vitro, L/F-transferase catalyzes the conjugation of Phe, Leu, or (at lower levels) Met residues to the N termini of proteins bearing N-terminal Lys or Arg (40). In contrast, when Arg-βgal was expressed from the $P_{trc}$ promoter in the presence of the wild-type chromosomal copy of the *aat* gene, the resulting Arg-βgal was found either to be completely unmodified or to contain an N-terminal Leu residue; no conjugation of Phe or Met could be detected in vivo (see Results). This greater specificity in vivo presumably reflects a higher stringency for the recognition of aminoacyl-tRNAs and suggests that another protein, missing in an active form from cell extracts, may act to increase the specificity of L/F-transferase. Earlier biochemical evidence that the isolated L/F-transferase catalyzes the complete conjugation reaction (40) suggests that additional subunits would be limited to a regulatory role. Consistent with the existence of additional subunits are biochemical data indicating that L/F-transferase activity fractionates as a high-molecular-mass complex (≥25 kDa) during gel chromatography of crude *E. coli* extracts (14).

The greater specificity of L/F-transferase in vivo (compared with the findings obtained with cell-free enzymatic assays) was also suggested by our previous in vivo findings (42). In in vitro studies, the dipeptides Arg-Ala, Lys-Ala, and (at higher concentrations) His-Ala were found to inhibit the aminoacylation of α-casein catalyzed by L/F-transferase (α-casein contains an N-terminal Arg residue) (40). No other dipeptide of the form X-Ala could inhibit this reaction. These results were interpreted to mean that N-terminal His could act as a substrate for N-terminal aminoacylation by L/F-transferase. However, the in vivo half-life of His-βgal in *E. coli* was found to be indistinguishable from those of the other long-lived X-βgal proteins (42), indicating that L/F-transferase-mediated aminoacylation of His-βgal (and its subsequent degradation by the Clp protease) was negligible in vivo.

The possibility that the presently known set of components of the N-end rule pathway in *E. coli* is incomplete is consistent with our inability to reconstruct the N-end rule pathway in vitro. Assays with purified Clp protease (a gift from M. Maurizi, National Institutes of Health) and purified X-βgal proteins yielded little degradation of X-βgal proteins and no differential degradation that depended on the identity of an N-terminal residue (data not shown). This result may have been due to the absence of a putative *N*-recognin-like component in these assays. In the eukaryotic N-end rule pathway, an *N*-recognin (also called E3) binds to a primary destabilizing N-terminal residue of an N-end rule substrate and initiates a cascade of reactions that result in substrate degradation (5).

We have screened for mutations that stabilize Arg-βgal in *E. coli* by using both transposon mutagenesis and the chemical mutagen ethyl methanesulfonate. To date, these screens have identified only the *clpA* and *aat* genes as components of the N-end rule pathway. Although this result would normally suggest that a screen is saturated, these screens have failed to yield mutations in *clpP*, which encodes the second subunit of the Clp protease (33). A constructed *E. coli* strain in which the central 55% of the *clpP* gene was replaced with a kanamycin resistance marker yielded blue colonies when transformed with either pUBP-R-βgal or pUBP-L-βgal, demonstrating that *clpP* is also essential for the degradation of N-end rule substrates in *E. coli*. The failure of our screens to identify an *E. coli N*-recognin might have resulted from this protein having an additional function that is essential for cell viability (inactivation of the N-end rule pathway is not lethal; 42) or from the existence of multiple, functionally overlapping *N*-recognins in *E. coli*. Recently, a selection-based screen in which the modified Tn*5* Kan protein, which specifies kanamycin resistance, was used as an N-end rule reporter substrate was developed. This screen, which did reveal an N-end rule-inactivating mutation in the *clpP* gene, is being used to search for other components of the N-end rule pathway (31a).

The *aat* gene maps to 19 min of the *E. coli* chromosome. This definitive position differs from a previous tentative mapping of *aat* to ~55 min (41). We have shown that *aat* is the last gene of a new operon that contains at least one additional ORF, termed *mdrH*. The deduced amino acid sequence encoded by the *mdrH* ORF is highly similar to the sequences of mammalian Mdr P glycoproteins. Other examples of *mdrH* homologs from bacteria include the *hylB* gene, required for hemolysin transport in uropathogenic *E. coli* strains (7), and the *cyaB* gene, implicated in the export of the calmodulin-sensitive adenylate cyclase-hemolysin bifunctional protein of *B. pertussis* (17). *E. coli* MdrH is also homologous to a family of bacterial membrane importers. These proteins, which include OppF and HisP, share a putative ATP-binding motif (23) with genes from the Mdr family (23). The sequence of MdrH is more similar to that of CyaB than it is to the sequences of bacterial importers involved in the membrane transport of small metabolites, suggesting that MdrH mediates the translocation of (unknown) peptides or proteins across an *E. coli* membrane (23).

The transcriptional coupling of *mdrH* and *aat* suggests a functional connection between the enzymes encoded by these genes. This possibility is particularly interesting in light of recently published mapping data (13). The genes *cydC* and *cydD* have been localized on the *E. coli* chromo-

some to the region immediately adjacent to the *infA* gene, which encodes translational initiation factor Infl. In the published map (13), the positions of *cydC* and *cydD* (13) appear to be identical to the positions demonstrated here to contain, respectively, *aat* and *mdrH* (Fig. 2). The *cydC* and *cydD* genes do not encode known components of the cytochrome *d* enzyme but are nonetheless essential for the expression of a functional cytochrome *d* (16, 35). Cytochrome *d* is a terminal oxidase that mediates *E. coli* respiration under conditions of low oxygen tension. Interestingly, at high oxygen tensions, this enzyme is replaced by the cytochrome *o* terminal oxidase. Thus, the replacements of specific terminal oxidases in response to changing oxygen tensions might involve L/F-transferase-dependent degradation of an old cytochrome or L/F-transferase-dependent transport of a new cytochrome to its position in the inner membrane. However, recent tests of these possibilities indicated that *aat* is distinct from *cydD*, because a mutant bearing a precise deletion of the *aat* gene did not have the phenotypes characteristic of mutations in *cydD* or functionally related genes (31a).

## ACKNOWLEDGMENTS

We are grateful to our colleagues whose names are cited in this paper and to B. Bachman (Yale University) for the gifts of strains and plasmids. We thank O. Lomovskaya for helpful discussions and comments on the manuscript. We also thank D. Siegle and R. Kolter for communicating their DNA sequence data prior to publication and for providing phages from their copy of the Kohara library.

This work was supported by grants to A.V. from the National Institutes of Health (GM31530 and AG08991). T.E.S. is a Merck Fellow of the Helen Hay Whitney Foundation.

## REFERENCES

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. G. Smith, and K. Struhl.** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Bachmair, A., D. Finley, and A. Varshavsky.** 1986. In vivo half-life of a protein is a function of its amino-terminal residue. Science **234**:179–186.
3. **Baker, R. T., J. W. Tobias, and A. Varshavsky.** 1992. Ubiquitin-specific proteases of *Saccharomyces cerevisiae*. Cloning of *UBP1* and *UBP2* and functional analysis of the *UBP* gene family. J. Biol. Chem. **267**:23364–23375.
4. **Balzi, E., M. Choder, W. Chen, A. Varshavsky, and A. Goffeau.** 1990. Cloning and functional analysis of the arginyl-tRNA-protein transferase gene *ATE1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **265**:7464–7471.
5. **Bartel, B., I. Wünning, and A. Varshavsky.** 1990. The recognition component of the N-end rule pathway. EMBO J. **9**:3179–3189.
6. **Biek, D. P., and S. N. Cohen.** 1986. Identification and characterization of *recD*, a gene affecting plasmid maintenance and recombination in *Escherichia coli*. J. Bacteriol. **167**:594–603.
7. **Blight, M. A., and I. B. Holland.** 1990. Structure and function of haemolysin B, P-glycoprotein and other members of a novel family of membrane translocators. Mol. Microbiol. **4**:873–880.
8. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. W. Heyneker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95.
9. **Brosius, J., T. J. Dull, D. D. Sleeter, and H. F. Noller.** 1981. Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. J. Mol. Biol. **148**:107–127.
10. **Burbaum, J. J., and P. Schimmel.** 1991. Structural relationships and the classification of aminoacyl-tRNA synthetases. J. Biol. Chem. **266**:16965–16968.
11. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction of the pACYC184 cloning vector. J. Bacteriol. **134**:1141–1156.
12. **Cummings, H. S., J. F. Sands, P. C. Foreman, J. Fraser, and J. W. B. Hershey.** 1991. Structure and expression of the *infA* operon encoding translational initiation factor INF1. J. Biol. Chem. **266**:16491–16498.
13. **Delaney, J. M., and C. Georgopoulos.** 1992. Physical map locations of the *trxB*, *htrD*, *cycC*, and *cydD* genes of *Escherichia coli*. J. Bacteriol. **174**:3824–3825.
14. **Deutch, C. E.** 1984. Aminoacyl-tRNA-protein transferases. Methods Enzymol. **106**:198–205.
15. **Finley, D., and V. Chau.** 1991. Ubiquitination. Annu. Rev. Cell Biol. **7**:25–70.
16. **Georgiou, C. D., H. Fang, and R. B. Gennis.** 1987. Identification of the *cydC* locus required for the expression of the functional form of the cytochrome *d* terminal oxidase complex in *Escherichia coli*. J. Bacteriol. **169**:2107–2112.
17. **Glaser, P., H. Sakamoto, J. Bellalou, A. Ullmann, and A. Danchin.** 1988. Secretion of cyclolysin, the calmodulin-sensitive adenylate cyclase-haemolysin bifunctional protein of *Bordetella pertussis*. EMBO J. **7**:3997–4004.
18. **Goldberg, A.** 1992. The mechanism and functions of ATP-dependent proteases in bacterial and animal cells. Eur. J. Biochem. **203**:9–23.
19. **Gonda, D. K., A. Bachmair, I. Wunning, J. W. Tobias, W. S. Lane, and A. Varshavsky.** 1989. Universality and structure of the N-end rule. J. Biol. Chem. **264**:16700–16712.
20. **Gottesman, S.** 1989. Genetics of proteolysis in Escherichia coli. Annu. Rev. Genet. **23**:163–198.
21. **Gottesman, S., W. P. Clark, and M. R. Maurizi.** 1990. The ATP-dependent Clp protease of *Escherichia coli*: sequence of *clpA* and identification of a Clp-specific substrate. J. Biol. Chem. **265**:7886–7893.
22. **Hershko, A.** 1991. The ubiquitin pathway for protein degradation. Trends Biochem. Sci. **16**:265–268.
23. **Higgins, C. F., S. C. Hyde, M. M. Mimmack, U. Gileadi, D. R. Gill, and M. P. Gallagher.** 1990. Binding protein-dependent transport systems. J. Bioenerg. Biomembr. **22**:571–592.
24. **Hiles, I. D., M. P. Gallagher, D. J. Jamieson, and C. F. Higgins.** 1987. Molecular characterization of the oligopeptide permease of *Salmonella typhimurium*. J. Mol. Biol. **195**:125–142.
25. **Hoopes, B. C., and W. R. McClure.** 1987. Strategies in regulation of transcriptional initiation, p. 1231–1240. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
26. **Hwang, B. J., K. M. Woo, A. L. Goldberg, and C. H. Chung.** 1988. Protease Ti, a new ATP-dependent protease in *Escherichia coli*, contains protein-activated ATPase and proteolytic functions in distinct subunits. J. Biol. Chem. **263**:8727–8734.
27. **Katayama, Y., S. Gottesman, J. Pumphrey, S. Rudikoff, W. P. Clark, and M. R. Maurizi.** 1988. The two-component, ATP-dependent Clp protease of *Escherichia coli*. J. Biol. Chem. **263**:15226–15236.
28. **Klausner, R. D., and R. Sitia.** 1990. Protein degradation in the endoplasmic reticulum. Cell **62**:611–614.
29. **Kleckner, N., J. Bender, and S. Gottesman.** 1991. Uses of transposons with emphasis on Tn*10*. Methods Enzymol. **204**: 139–180.
30. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole E. coli chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
31. **Lazinski, D., E. Grzadzielska, and A. Das.** 1989. Sequence-specific recognition of RNA hairpins by bacteriophage antiterminators requires a conserved arginine-rich motif. Cell **59**:207–218.

31a. **Lomovskaya, O., and A. Varshavsky.** Unpublished data.

32. **Lopilato, J., S. Bortner, and J. Beckwith.** 1986. Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. Mol. Gen. Genet. **205**:285–290.
33. **Maurizi, M. R., W. P. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers, and S. Gottesman.** 1990. Sequence and structure of ClpP, the proteolytic component of the ATP-

dependent Clp protease of *Escherichia coli*. J. Biol. Chem. **265:**12536–12545.
34. **Olson, T. S., and J. F. Dice.** 1989. Regulation of protein degradation rates in eukaryotes. Curr. Opin. Cell Biol. **1:**1194–1200.
35. **Poole, R., H. Williams, A. Downie, and F. Gibson.** 1989. Mutations affecting the cytochrome *d*-containing oxidase complex of *Escherichia coli* K12: identification and mapping of a fourth locus, *cydD*. J. Gen. Microbiol. **135:**1865–1874.
36. **Rechsteiner, M.** 1991. Natural substrates of the ubiquitin proteolytic pathway. Cell **66:**615–618.
37. **Shrader, T. E., and D. M. Crothers.** 1990. Effects of DNA sequence and histone-histone interactions on nucleosome placement. J. Mol. Biol. **216:**69–84.
38. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Simmons, R. W., F. Houman, and N. Kleckner.** 1987. Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. Gene **53:**85–96.
40. **Soffer, R. L.** 1974. Aminoacyl-tRNA-protein transferases. Adv. Enzymol. **40:**91–139.
41. **Soffer, R. L., and M. Savage.** 1974. A mutant of *Escherichia coli* defective in leucyl, phenylalanyl-tRNA-protein transferase. Proc. Natl. Acad. Sci. USA **71:**1004–1007.
42. **Tobias, J. W., T. E. Shrader, G. Rocap, and A. Varshavsky.** 1991. The N-end rule in bacteria. Science **254:**1374–1377.
43. **Tobias, J. W., and A. Varshavsky.** 1991. Cloning and functional analysis of the ubiquitin-specific protease gene *UBP1* of *Saccharomyces cerevisiae*. J. Biol. Chem. **266:**12021–12028.
44. **Varshavsky, A.** 1991. Naming a targeting signal. Cell **64:**13–15.
45. **Varshavsky, A.** 1992. The N-end rule. Cell **69:**725–735.
46. **Way, J. C., M. A. Davis, D. Morisato, D. E. Roberts, and N. Kleckner.** 1984. New Tn*10* derivatives for transposon mutagenesis and for construction of *lacZ* operon fusions by transposition. Gene **32:**369–379.
47. **Way, J. C., and N. Kleckner.** 1984. Essential sites at transposon Tn*10* termini. Proc. Natl. Acad. Sci. USA **81:**3452–3456.
48. **Yager, T. D., and P. H. V. Hippel.** 1987. Transcript elongation and termination in *Escherichia coli*, p. 1241–1275. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.