THE JOURNAL OF BIOLOGICAL CHEMISTRY Vol. 271, No. 45, Issue of November 8, pp. 28521–28532, 1996
Printed in U.S.A.

A Mouse Amidase Specific for N-terminal Asparagine

THE GENE, THE ENZYME, AND THEIR FUNCTION IN THE N-END RULE PATHWAY*

(Received for publication, July 9, 1996)

Sergei Grigoryev‡§, Albert E. Stewart¶‖, Yong Tae Kwon‡, Stuart M. Arfin¶, Ralph A. Bradshaw¶, Nancy A. Jenkins, Neal G. Copeland**, and Alexander Varshavsky‡ ‡‡**

*From the ‡Division of Biology, California Institute of Technology, Pasadena, California 91125, the ¶Department of Biological Chemistry, College of Medicine, University of California, Irvine, California 92717, and the **ABL Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702*

The N-end rule relates the *in vivo* half-life of a protein to the identity of its N-terminal residue. In both fungi and mammals, the tertiary destabilizing N-terminal residues asparagine and glutamine function through their conversion, by enzymatic deamidation, into the secondary destabilizing residues aspartate and glutamate, whose destabilizing activity requires their enzymatic conjugation to arginine, one of the primary destabilizing residues. We report the isolation and analysis of a mouse cDNA and the corresponding gene (termed *Ntan1*) that encode a 310-residue amidohydrolase (termed Nt^N-amidase) specific for N-terminal asparagine. The ~17-kilobase pair *Ntan1* gene is located in the proximal region of mouse chromosome 16 and contains 10 exons ranging from 54 to 177 base pairs in length. The ~1.4-kilobase pair *Ntan1* mRNA is expressed in all of the tested mouse tissues and cell lines and is down-regulated upon the conversion of myoblasts into myotubes. The *Ntan1* promoter is located ~500 base pairs upstream of the *Ntan1* start codon. The deduced amino acid sequence of mouse Nt^N-amidase is 88% identical to the sequence of its porcine counterpart, but bears no significant similarity to the sequence of the *NTA1*-encoded N-terminal amidohydrolase of the yeast *Saccharomyces cerevisiae*, which can deamidate either N-terminal asparagine or glutamine. The expression of mouse Nt^N-amidase in *S. cerevisiae nta1*Δ was used to verify that Nt^N-amidase retains its asparagine selectivity *in vivo* and can implement the asparagine-specific subset of the N-end rule. Further dissection of mouse *Ntan1*, including its null phenotype analysis, should illuminate the functions of the N-end rule, most of which are still unknown.

Features of proteins that confer metabolic instability are called degradation signals, or degrons (1). The essential component of one degradation signal, termed the N-degron, is a destabilizing N-terminal residue of a protein (2). A set of N-degrons containing different destabilizing residues yields a rule, termed the N-end rule, that relates the *in vivo* half-life[1] of a protein to the identity of its N-terminal residue (reviewed in Refs. 3 and 4). The N-end rule pathway has been found in all species examined, including the eubacterium *Escherichia coli* (5), the yeast *Saccharomyces cerevisiae* (6), and mammalian cells (7, 8). The N-end rules of these organisms are similar but distinct (4).

The understanding of the functions of the N-end rule is sketchy. In particular, the N-end rule pathway has been shown to be required for the import of peptides in *S. cerevisiae* (9); it may also be involved in the control of signal transduction and cell differentiation (10, 11). Physiological substrates of the N-end rule pathway include Sindbis virus RNA polymerase (12) and yeast Gα, the *GPA1*-encoded α subunit of the heterotrimeric G protein (13).

In eukaryotes, the N-degron comprises at least two determinants: a destabilizing N-terminal residue and an internal lysine (or lysines) of a substrate (6, 14–16). The Lys residue is the site of formation of a multiubiquitin chain (17–20). Ubiquitin (Ub)[2] is a 76-residue protein whose covalent conjugation to other proteins is involved, directly or by way of regulation, in a multitude of processes, including cell growth and differentiation, signal transduction, DNA repair, and the transport of substances across membranes. In many of these processes, Ub acts through routes that involve protein degradation (reviewed in Refs. 4 and 21–26).

The N-end rule is organized hierarchically. In eukaryotes such as the yeast *S. cerevisiae*, Asn and Gln are tertiary destabilizing N-terminal residues (denoted as $N\text{-}d^t$) in that they function through their conversion, by N-terminal amidohydro-

* This work was supported by National Institutes of Health Grant DK39520 (to A. V.) and Grant DK32461 (to R. A. B and S. M. A.) and by NCI under contract with ABL. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) U57690 (mouse $E2_{14K}$ cDNA), U57691 (mouse Ntan1 genomic DNA), and U57692 (mouse Ntan1 cDNA).

§ Present address: Dept. of Biology, Morrill Science Center, University of Massachusetts, Amherst, MA 01003.

‖ Present address: Dept. of Crystallography, ICRF Structural Molecular Biology Unit, Birkbeck College, University of London, Malet St., London WC1E 7HX, UK.

‡‡ To whom correspondence should be addressed: Div. of Biology, 147-75, Caltech, 391 South Holliston Ave., Pasadena, CA 91125. Tel.: 818-395-3785; Fax: 818-440-9821; E-mail: varshavskya@starbase1.caltech.edu.

[1] The term "half-life" is used in this paper as a rough approximation since the *in vivo* degradation of many short-lived proteins, including the engineered N-end rule substrates, deviates from a first-order kinetics. A terminology for describing nonexponential decay was proposed in Ref. 8.

[2] The abbreviations used are: Ub, ubiquitin; $N\text{-}d^t$, a tertiary destabilizing N-terminal residue; $N\text{-}d^s$, a secondary destabilizing N-terminal residue; $N\text{-}d^p$, a primary destabilizing N-terminal residue; Nt-amidase, amidase specific for N-terminal Asn and Gln; Nt^N-amidase, amidase specific for N-terminal Asn; Nt^Q-amidase, amidase specific for N-terminal Gln; R-transferase, Arg-tRNA-protein transferase; E2, ubiquitin-conjugating enzyme; $E2_{14K}$, a 14-kDa ubiquitin-conjugating enzyme; kb, kilobase pair(s); bp, base pair(s); ORF, open reading frame; βgal, *E. coli* β-galactosidase; RACE, rapid amplification of cDNA ends; PCR, polymerase chain reaction; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; DHFR, mouse dihydrofolate reductase; IEF, isoelectric focusing.

lase (Nt-amidase), into the secondary destabilizing N-terminal residues Asp and Glu (denoted as N-d^s), whose destabilizing activity requires their conjugation, by Arg-tRNA-protein transferase (R-transferase), to Arg, one of the primary destabilizing N-terminal residues (denoted as N-d^p) (6, 27, 28). The N-d^p residues are bound directly by N-recognin (also called E3), the recognition component of the N-end rule pathway (4). In *S. cerevisiae*, N-recognin is a 225-kDa protein (encoded by *UBR1*) that selects potential N-end rule substrates by binding to their N-d^p residue Phe, Leu, Trp, Tyr, Ile, Arg, Lys, or His (29–31). This binding is followed by the formation of a substrate-linked multi-Ub chain in a reaction mediated by the Ub-conjugating enzyme (E2) Ubc2p, one of at least 11 distinct E2 enzymes in *S. cerevisiae* (31, 32). The substrate is then processively degraded by the 26 S proteasome, an ~2-MDa, ATP-dependent, multisubunit protease (33–38). The four "upstream" components of the *S. cerevisiae* N-end rule pathway (Nta1p (Nt-amidase), Ate1p (R-transferase), Ubc2p (a Ub-conjugating enzyme), and Ubr1p (N-recognin)) are physically associated in a targeting complex (27, 31).[3]

The *NTA1*-encoded Nt-amidase of *S. cerevisiae* can deamidate N-terminal Asn or Gln (3, 27). Stewart *et al.* (39, 40) reported the purification of a porcine Nt-amidase and the cloning of its ~1.3-kb cDNA. In contrast to the yeast Nt-amidase (Nta1p), this Nt-amidase, termed Nt^N-amidase, can deamidate N-terminal Asn but not Gln. Described below is the isolation and characterization of *Ntan1*, an ~17-kb gene[4] that encodes a mouse homolog of the porcine enzyme, termed Nt^N-amidase (the superscript "N" and the second "N" in the gene's name refer to the Asn specificity of this Nt-amidase).

Both Asn and Gln are destabilizing residues in the mammalian N-end rule (7, 8). Furthermore, both N-terminal Asn and Gln of test proteins are deamidated in mammalian cell extracts (7).[5] Since Nt^N-amidase deamidates N-terminal Asn but not Gln, there must exist yet another mammalian Nt-amidase, which can deamidate N-terminal Gln. The resulting bifurcation in the structure of the mammalian N-end rule is illustrated in Fig. 1. *Ntan1* is the first cloned mammalian gene whose function appears to be confined to the N-end rule pathway.

EXPERIMENTAL PROCEDURES

Bacterial and Yeast Strains and Genetic Techniques—The *E. coli* strains MC1061, DH5α, and JM101 were grown in Luria broth containing relevant antibiotics and were transformed using the calcium chloride method (41). *S. cerevisiae* was grown on rich (YPD) or synthetic medium containing 2% glucose (SD medium) or 2% galactose (SG medium) (42). Unless stated otherwise, the transformation of *S. cerevisiae* was carried out as described (43). The *S. cerevisiae* strains were YPH500 (*MATα ura3-52 lys2-801 ade2-101 trp1-Δ63 his3-Δ200 leu2-Δ1*) (44) and SGY4, an *nta1-Δ2::LEU2* derivative of YPH500. SGY4 was constructed by transforming YPH500 (using the procedure of Ref. 45) with an *Xho*I-*Xba*I fragment of the plasmid pSG8 that bore the *nta1-Δ2::LEU2* deletion/disruption allele and selecting for Leu^+ cells. Southern hybridization analysis of Leu^+ transformants, using restriction endonuclease cuts diagnostic of the transplacement (46), was used to verify the predicted structure of the integrated *nta1-Δ2::LEU2* allele. The plasmid pSG8 was derived from a plasmid (based on pRS316 (41)) that contained *S. cerevisiae NTA1* between *Xho*I and *Xba*I sites (27). Replacement of the *Spe*I-*Eco*RI fragment of the *NTA1* ORF with *LEU2* yielded pSG8.

β-Galactosidase Assays—Colony assays for *E. coli* β-galactosidase (βgal) in *S. cerevisiae* were carried out by overlaying yeast colonies on SG plates with 0.5% agarose containing 0.1% SDS, 4% dimethylformamide, and a 0.1 mg/ml concentration of the chromogenic βgal substrate 5-bromo-4-chloro-3-indolyl β-D-galactoside, followed by incubation for a few hours at 37 °C and examination of the colony color (white to blue, depending on the level of βgal). Quantitative assays for βgal activity were carried out with *S. cerevisiae* whole cell extracts using another βgal substrate, chlorophenol red β-D-galactopyranoside. Cells in a 1-ml culture (A_{600} ~ 1) were pelleted by centrifugation and resuspended in 1 ml of buffer K (1 mM $MgCl_2$, 0.1 M potassium phosphate, pH 7.0). After determining the A_{600} of the suspension, 50- or 100-μl samples were diluted to 1 ml with buffer K. 0.1% SDS (20 μl) and $CHCl_3$ (50 μl) were then added; the suspension was vortexed for 20 s and incubated for 15 min at 30 °C, followed by the addition of 20 μl of chlorophenol red β-D-galactopyranoside (34 mg/ml in buffer K) and further incubation at 30 °C. The A_{574} of a sample was measured at different incubation times (after clarification by a brief centrifugation in a microcentrifuge). The chlorophenol red β-D-galactopyranoside units (U_{CPRG}) of βgal activity were calculated as follows: $U_{CPRG} = A_{574}/t{\cdot}v{\cdot}A_{600}$, where t and v are the time of incubation (min) and the sample volume (ml), respectively.

Mammalian Cell Lines—The mouse Friend erythroleukemia cell line MEL-C19 was provided by Dr. L. A. Brents (NCI, National Institutes of Health, Bethesda, MD). The mouse lymphoma (EL4.IL-2) and myoblast (C_2C_{12}) cell lines were purchased from the American Type Culture Collection (Rockville, MD). MEL and EL4 cells were grown in Dulbecco's modified Eagle's/Ham's F-12 medium (Mediatech, Herndon, VA). C_2C_{12} cells were grown in Dulbecco's modified Eagle's medium plus 4.5 g/liter glucose (Mediatech). All media were supplemented with 10% fetal bovine serum, glutamine, penicillin, streptomycin, and pyruvate (47). MEL cells were induced to differentiate in RPMI 1640 medium (Mediatech) containing 10% fetal bovine serum by adding *N,N′*-hexamethylenebisacetamide to a final concentration of 3 mM and incubating a culture for another 5 days. C_2C_{12} cells were induced to differentiate by replacing fetal bovine serum with 10% horse serum and incubating a confluent culture for 4–5 days.

Cloning of the Mouse Ntan1 cDNA—Stewart *et al.* (39, 40) have purified an amidohydrolase, termed Nt^N-amidase, from porcine liver that is specific for N-terminal Asn in polypeptides; they also determined the N-terminal sequence of porcine Nt^N-amidase (peptide 1) and several sequences of internal peptides (including peptide 2) produced by fragmentation with cyanogen bromide (see Fig. 2*A*). Stewart *et al.* (40) employed "reverse-translated" oligonucleotides corresponding to the sequences of peptides 1 and 2 to isolate a 632-bp porcine cDNA fragment whose 5′- and 3′-regions encoded the sequences of peptides 1 and 2, respectively.

To clone a mouse cDNA (termed *Ntan1* cDNA) that encoded a homolog of the porcine Nt^N-amidase, we used cDNA libraries (CLONTECH, Palo Alto, CA) prepared from mouse liver (in λgt11) and MEL-C19 cells (in λgt10). A low-stringency hybridization screening (carried out as recommended by CLONTECH) of the mouse liver cDNA library utilized the above 632-bp porcine cDNA fragment (at the time, only that fragment of the subsequently cloned porcine *Ntan1* cDNA had been isolated by Stewart *et al.* (40)). A positive clone contained a 1560-bp mouse DNA insert; a 114-bp region of this insert was 85% identical to the nucleotide sequence of a region of the porcine *Ntan1* mRNA. This and other evidence (see "Results") suggested that the 1560-bp mouse DNA fragment was derived from an unspliced or partially spliced *Ntan1* pre-mRNA.

Since a Northern hybridization with the 1560-bp fragment as a probe suggested that the abundance of mouse *Ntan1* mRNA in MEL cells was significantly higher than in the liver, the 1560-bp fragment was used to screen a λgt10-based cDNA library prepared from MEL cells. Two positive clones (out of ~5×10^6 clones screened) containing mouse cDNA inserts of 1030 and 605 bp were obtained. The amino acid sequence encoded by the larger insert (which encompassed most of the 605-bp insert) was 83% identical to the deduced sequence of porcine Nt^N-amidase (40); the region of similarity began 33 nucleotides downstream of the ATG start codon in the porcine *Ntan1* ORF. In the 605-bp mouse cDNA insert, the TAA stop codon of the putative mouse *Ntan1* ORF was followed by 157 nucleotides of untranslated region linked to a poly(A) tract. The 1030-bp mouse cDNA insert ended 9 bp upstream of the poly(A) addition site.

On the assumption that the 1030-bp mouse cDNA insert lacked a sequence corresponding to the 5′-coding region of the mouse *Ntan1* mRNA, we employed the 5′-RACE-PCR technique (48, 49) to amplify the missing region, using an oligonucleotide primer specific for the 1030-bp clone and a primer complementary to an *in vitro* produced 5′-oligo(dC) tract. Specifically, poly(A)$^+$ mRNA was isolated from ~4×10^8 MEL cells using the Fast Track mRNA isolation kit (Invitrogen, San Diego, CA) and dissolved in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0, to an

[3] M. Ghislain, A. Webster, and A. Varshavsky, unpublished data.

[4] The names of mouse genes are in italics, with the first letter uppercase. The names of the corresponding mouse proteins are Roman, all upper-case. The names of *S. cerevisiae* genes are in italics, all uppercase. The names of the corresponding yeast proteins are Roman, with the first letter upper-case and an extra lower-case "p" at the end (86).

[5] S. Grigoryev and A. Varshavsky, unpublished data.

A_{260} of 5. 1.5 μl of this sample were diluted with 12 μl of diethyl pyrocarbonate-treated water, incubated at 70 °C for 5 min, and cooled on ice. To this sample were added 20 pmol of a primer corresponding to the antisense strand of the mouse *Ntan1* ORF between nucleotides +700 and +725 (GenBank™/EMBL accession number U57692), 2 μl of 10 × PCR buffer (Perkin-Elmer), 2 μl of 0.1 M dithiothreitol, 1 μl of 10 mM dNTPs, 1 μl of 25 mM $MgCl_2$, and 1 μl of bovine serum albumin (2 mg/ml). The sample was incubated at 42 °C for 2 min, followed by the addition of 1 μl of Superscript II reverse transcriptase (Life Technologies, Inc.) and incubation at 42 °C for another 40 min. The temperature was increased to 55 °C, followed by the addition of 1 μl of RNase H (Life Technologies, Inc.; 2 units/ml) and incubation for 20 min. The resulting cDNA products were purified with Glass Max (Life Technologies, Inc.) using three standard washes plus a wash with cold 70% ethanol and were eluted with 50 μl of water.

To produce a cDNA-linked 5′-oligo(dC) extension, 5 μl of purified cDNA were diluted with 11.5 μl of water; incubated at 70 °C for 5 min; cooled on ice; and then mixed with 1 μl of 10 × PCR buffer (to a final concentration of 0.5×), 1 μl of 25 mM $MgCl_2$, 0.5 μl of bovine serum albumin (2 mg/ml), 0.5 μl of 10 mM dCTP, and 0.25 μl of terminal transferase (Boehringer Mannheim). After incubation at 37 °C for 5 min, the enzyme was inactivated by heating the sample at 65 °C for 10 min. The first round of RACE-PCR amplification was carried out in a 100-μl sample containing 10 μl of 10 × PCR buffer, 2.5 mM $MgCl_2$, 5 μl of cDNA linked to oligo(dC), 20 pmol of a primer corresponding to the antisense strand of the *Ntan1* ORF between nucleotides +283 and +310 (GenBank™/EMBL accession number U57692), and 20 pmol of G-anchor primer (5′-AGGCCACGCGTCGACTAGTAC$(G)_{17}$-3′). The sample was incubated at 94 °C for 5 min and then at 57 °C for 8 min, followed by the addition of AmpliTaq DNA polymerase (Perkin-Elmer) and incubation at 72 °C for 8 min to produce the strand complementary to cDNA. Thereafter, 35 cycles of a three-step PCR amplification were carried out. Each step involved consecutive incubations for 30 s at 94 °C, for 1 min at 57 °C, and for 2 min at 72 °C. 2 μl of the first-round PCR product were used for the second round of PCR that utilized, in the same total volume, 0.4 nmol of G-adapter primer (same as the G-anchor primer but lacking G_{17}) and 0.4 nmol of a primer corresponding to the antisense strand of the *Ntan1* ORF between nucleotides +232 and +259.

The PCR-produced DNA fragments were inserted into pCRII (Invitrogen). Four of the resulting clones were sequenced; all of them contained an ATG codon preceded by one and the same 34-bp sequence and followed by a region encoding an amino acid sequence similar to the directly determined N-terminal sequence of the purified porcine Nt^N-amidase (40). An in-frame stop codon 6 bp upstream of the above ATG codon suggested that the latter was indeed the *in vivo* start codon of the *Ntan1* ORF. To construct the full-length *Ntan1* cDNA, PCR was carried out using the above 1030-bp *Ntan1* cDNA fragment as a template, the primer 5′-GTAGGTACCGCCTTTGCCAAAAATAAGATTTTATTTTG--3′ (which was complementary to the 3′-end of *Ntan1* cDNA and contained a *Kpn*I site), and the primer 5′-ATCCTCGAGCATATGCCACTGCTGGTGGATGGGCAGCGCGTCCGTCTGCCACGGTCCGC-3′ (which was complementary to the 5′-end of the 1030-bp fragment and contained an *Xho*I site as well as the RACE-PCR-derived 5′-proximal region of *Ntan1* cDNA that was absent from the 1030-bp fragment). The PCR product was digested with *Kpn*I and *Xho*I; the resulting fragment was inserted into *Kpn*I/*Xho*I-cut λTRP vector (a gift from Dr. S. Elledge, Baylor College of Medicine, Houston, TX), a variant of the vector λYES (50) that contained *S. cerevisiae TRP1* (instead of *URA3*) as a selectable marker. The resulting plasmid, pSG61, was used to express the mouse NTAN1 protein in *S. cerevisiae* from the P_{GAL1} promoter.

Cloning and Sequencing of the Mouse Ntan1 Gene—A genomic DNA library produced from mouse embryonic stem cells (strain C129) and carried in the P1 phage (51) was used. Screening of the library (164,000 P1 clones arranged in 400 pools) was carried out by Genome Systems Inc. (St. Louis, MO) as a custom service, using PCR and synthetic primers derived from the sequence of the mouse *Ntan1* cDNA (nucleotides +641 to +663 and nucleotides +730 to +752 (GenBank™/EMBL accession number U57692)) (see Fig. 2*B*). Four positive clones were identified; one of them (clone 1798) was used for further analysis. To amplify and isolate a P1 plasmid containing the insert 1798, 0.3 ml of an overnight *E. coli* NS3529 culture bearing the plasmid was inoculated into 30 ml of Luria broth containing kanamycin at 25 μg/ml. Following a 30-min incubation at 37 °C in a rotary shaker, isopropylthiogalactoside was added to a final concentration of 1 mM (to induce the lytic operon of P1), and the culture was incubated at 37 °C until saturation (~5 h). The P1 plasmid was isolated using the alkaline lysis method (41). The plasmid was digested with various restriction endonucleases; the resulting fragments were subcloned into pUC18 (41) or Bluescript II SK(+) (Stratagene, La Jolla, CA) and propagated in *E. coli* DH5α. Genomic DNA inserts containing sequences of *Ntan1* cDNA were identified by colony hybridization using the *Ntan1* cDNA insert of pSG61 (see above) as a probe. The regions of *Ntan1* that encompassed exons I–X and the regions of introns adjacent to exons were sequenced using the chain termination method (41) and exon-specific primers. The physical map of the *Ntan1* locus (see Fig. 3) was produced initially by restriction mapping of *Ntan1*-specific genomic DNA fragments and by PCR-based analyses of DNA that spanned the borders of sequenced regions. Subsequently, the entire ~17-kb *Ntan1* gene was sequenced using a similar strategy: identifying a set of restriction fragments of P1 clone 1798 that encompassed the *Ntan1* locus, subcloning these fragments to a size of ~1.5 kb into Bluescript II SK(+), and sequencing the inserts from both ends. The remaining regions were sequenced using primers specific for the junctions to already sequenced fragments. Most (>75%) of the *Ntan1* fragments were sequenced on both strands. These nucleotide sequences have been submitted to the GenBank™/EMBL Data Bank with accession numbers U57692 (*Ntan1* cDNA) and U57691 (*Ntan1* genomic DNA).

Cloning of Mouse cDNA Encoding the $E2_{14K}$ Ub-conjugating Enzyme—We screened the λgt10-based cDNA library prepared from MEL cells (see above) with the rabbit $E2_{14K}$ cDNA (a gift from Dr. S. Wing) (52) as a probe. In this library, the $E2_{14K}$ probe detected an ~30-fold higher number of clones than the *Ntan1* probe, consistent with a lower *in vivo* level of *Ntan1* mRNA in comparison with $E2_{14K}$ mRNA (see "Results"). All four of the analyzed mouse $E2_{14K}$ cDNAs contained identical sequences encompassing the $E2_{14K}$ ORF (GenBank™ accession number U57690), which was 93% identical to the rabbit $E2_{14K}$ ORF and which encoded an amino acid sequence 100% identical to that of rabbit $E2_{14K}$. However, while three of the mouse $E2_{14K}$ cDNAs contained a poly(A) sequence 174 bp downstream of the $E2_{14K}$ ORF, in the fourth cDNA, the poly(A) sequence was located 500 bp farther downstream (data not shown), consistent with the presence of two $E2_{14K}$ mRNA species in the Northern hybridization patterns (see "Results").

5′-Mapping of Ntan1 mRNA—Total cytoplasmic RNA was isolated from mouse MEL-C19 cells as described (41). The primer extension mapping was carried out as described (41) using 50 μg of RNA/assay. The Superscript II polymerase (Life Technologies, Inc.) was used for the reverse transcription step, which employed the following synthetic primers: nucleotides 985-1014 and 1485–1511, corresponding to the antisense DNA strand of *Ntan1* (see Fig. 2*B*). Reaction products were analyzed by polyacrylamide gel electrophoresis on a 9% sequencing gel.

Chromosome Mapping of the Mouse Ntan1 Gene—The chromosomal position of *Ntan1* was determined using the interspecific backcross analysis. Interspecific backcross progeny were generated by mating (C57BL/6J × *Mus spretus*)F1 females and C57BL/6J males as described (53). A total of 205 N_2 mice were used to map *Ntan1*. DNA isolation, digestion with restriction endonucleases, gel electrophoresis, and Southern hybridization were performed essentially as described (54). All blots were prepared with Hybond-N^+ nylon membranes (Amersham Corp.). The probe, an ~2.3-kb *Eco*RI fragment of mouse genomic DNA that encompasses exons II–IV of *Ntan1*, was labeled with [α-^{32}P]dCTP using a nick translation labeling kit (Boehringer Mannheim); washing was carried out to a final stringency of 0.1 × SSCP, 0.1% SDS at 65 °C. An ~15-kb fragment and an ~4-kb fragment were detected with this probe in *Hin*dIII-digested C57BL/6J DNA and *M. spretus* DNA, respectively. The presence or absence of the ~4-kb *M. spretus*-specific *Hin*dIII fragment was followed in DNA of backcross mice.

A description of the probes and restriction fragment length polymorphisms for the loci linked to *Ntan1*, including the loci encoding protamine-1 (*Prm1*), CCAAT/enhancer-binding protein-δ (*Cebpd*), and the immunoglobulin λ chain (*Igl*), has been reported (55). Recombination distances were calculated as described (56) using the program SPRETUS MADNESS. Gene order was determined by minimizing the number of recombination events required to explain the allele distribution patterns.

Southern and Northern Hybridizations—For Southern hybridization, DNA was isolated from mouse liver, digested with restriction endonucleases, fractionated on 1% agarose gels, and blotted onto Hybond-N^+ membranes as described (41). Prehybridization was carried out at 42 °C for 5 h in a buffer containing 2 × Denhardt's solution, 1 M NaCl, 0.5% SDS, 1 mM EDTA, 50 mM Tris-HCl, pH 7.5, and 0.1 mg/ml sheared denatured calf thymus DNA. Hybridization was performed at 37 °C for 40 h in hybridization buffer (0.5 × Denhardt's solution, 30% formamide, 1% SDS, 5 mM EDTA, 1 M NaCl, 50 mM sodium phosphate, pH 7.2, and 0.1 mg/ml sheared denatured calf thymus DNA). The probe DNA was labeled to ~10^9 cpm/μg using [^{32}P]dCTP and the Random Primer labeling kit (DuPont NEN). The final concentration of labeled,

heat-denatured probe DNA was ~5 ng/ml. After hybridization, the membranes were washed under low-stringency conditions (2 h at 25 °C in 0.4 M NaCl, 1% SDS, 5 mM EDTA, and 5 mM sodium phosphate, pH 7.2). For washes of increasing stringency, carried out with shaking for 30 min, the concentration of NaCl was decreased in 0.1 M increments; at 0 M NaCl, the temperature of washes was increased, in 10 °C increments, to 65 °C.

For Northern hybridization, the poly(A)$^+$ RNA was isolated from mouse MEL-C19 cells, mouse lymphoma EL4.IL-2 cells, and mouse C_2C_{12} myoblasts using the Fast Track mRNA isolation kit (Version 3.5). The mouse skeletal muscle poly(A)$^+$ RNA was purchased from CLONTECH. RNA was fractionated by electrophoresis on 1% agarose-formaldehyde gels and blotted onto Hybond-N$^+$ membranes as described (41). Northern hybridization was carried out as recommended by CLONTECH (protocol c, for hybridization with DNA probes). The probes (labeled with ^{32}P as described above) were mouse β-actin cDNA (CLONTECH) and DNA inserts of the plasmids pSG61 (mouse *Ntan1* cDNA; see above), pSG72 (mouse $E2_{14K}$ cDNA; see above), and pSG76 (mouse glyceraldehyde-3-phosphate dehydrogenase (GAPDH) cDNA). The GAPDH cDNA was produced using reverse transcription-PCR (41), the oligo(dT)-primed MEL-C19 cDNA library (see above), and the GAPDH-specific primers 5′-ATGGTGAAGGTCGGTGTGAACGGA-3′ and 5′-TTACTCCTTGGAGGCCATGTAGGC-3′, derived from the sequence of GAPDH cDNA (57).

X-DHFR Test Proteins—ORFs encoding Ub-*X*-dihydrofolate reductase fusion proteins (Ub-*X*-DHFR proteins) (construct I of Ref. 6; *X* = Asn, Asp, or Gln) were used to replace ORFs encoding Ub-*X*-βgal in pKKUB-*X*-βgal plasmids (5). The resulting plasmids, pSG4, pSG41, and pSG44, expressed Ub-Asn-DHFR, Ub-Asp-DHFR, and Ub-Gln-DHFR, respectively, from the P_{trc} promoter. *E. coli* JM101 was transformed with one of these plasmids and also with pJT184, which expressed Ubp1p, a Ub-specific protease of *S. cerevisiae* (5). A 50-ml culture was grown at 37 °C to an A_{600} of ~0.9 in Luria broth plus ampicillin (40 μg/ml) and chloramphenicol (20 μg/ml). Unless stated otherwise, all subsequent procedures were carried at 4 °C. The culture was chilled on ice for 15 min; centrifuged at 3000 × *g* for 5 min; washed twice with M9 medium (41); and resuspended in 50 ml of M9 medium supplemented with glucose (0.2%), thiamine (2 μg/ml), ampicillin (40 μg/ml), 1 mM isopropylthiogalactoside, and methionine assay medium (Difco). The suspension was shaken for 1 h at 37 °C, followed by the addition of 1 mCi of Tran^{35}S-label (ICN, Costa Mesa, CA) and further incubation for 1 h at 37 °C. Unlabeled L-methionine was then added to 1 mM, and the shaking was continued for another 10 min. The cells were harvested by centrifugation; washed twice with M9 medium; and resuspended in 0.5 ml of 25% (w/v) sucrose, 50 mM Tris-HCl, pH 8.0. Freshly prepared egg white lysozyme (0.1 ml of a 10 mg/ml solution in 0.25 M Tris-HCl, pH 8.0) was then added, and the suspension was incubated for 5 min at 0 °C, followed by the addition of 0.1 ml of 0.5 M EDTA, pH 8.0, and a 5-min incubation at 0 °C. The suspension was transferred to a centrifuge tube containing 1 ml of 65 mM EDTA, 50 mM Tris-HCl, pH 8.0, and protease inhibitors (25 μg/ml each antipain, chymostatin, leupeptin, pepstatin, and aprotinin; Sigma). Triton X-100 (10 μl of a 10% (v/v) solution) was then added and dispersed by pipetting. The lysate was centrifuged at 40,000 × *g* for 30 min in a TL-100 ultracentrifuge (Beckman Instruments). The supernatant containing ^{35}S-*X*-DHFR (the Ub moiety of Ub-*X*-DHFR was removed *in vivo* by the Ubp1p protease) was frozen in liquid N_2 and stored at −80 °C. *X*-DHFR (*X* = Asn, Asp, or Glu) was purified on a methotrexate affinity column (Pierce; 0.5-ml bed volume) that had been equilibrated in 50 ml of 50 mM Tris-HCl, pH 7.2. The thawed ^{35}S-supernatant was clarified by centrifugation at 12,000 × *g* for 1 min and applied to the column, which was then washed with 10 ml of 1 M KCl, 20 mM Tris-HCl, pH 8.0. *X*-DHFR was eluted with 2 mM folic acid, 50 mM Tris-HCl, pH 8.0. The elution buffer was applied in 1-ml samples, and 1-ml fractions were collected. Pooled fractions containing the peak of eluted ^{35}S were dialyzed for 20 h against two changes of 50% glycerol, 1 mM $MgCl_2$, 0.1 mM EDTA, and 40 mM HEPES, pH 7.5, and stored at −20 °C. ^{35}S-*X*-DHFR proteins were examined by SDS-polyacrylamide gel electrophoresis and fluorography and found to be >95% pure.

Isoelectric Focusing Assay for Amidase Activity—30 ml of an *S. cerevisiae* culture (A_{600} ~ 1) were collected; washed once with 40 ml of water; pelleted at 3000 × *g*; washed once with 40 ml of buffer A (125 mM KCl, 15 mM $MgCl_2$, and 70 mM Tris-HCl, pH 8.0); and resuspended in 1 ml of buffer A containing 1 mM dithiothreitol, 0.2 mM phenylmethylsulfonyl fluoride, and the above mixture of protease inhibitors. 0.4 ml of 0.5-mm glass beads was added, and the suspension was vortexed at the maximal setting (three times for 1 min, with 2-min incubations on ice in between). The lysate was centrifuged at 12,000 × *g* for 5 min, and the supernatant was used for the amidase assay either immediately or after dilution with buffer A containing 1% bovine serum albumin.

In most assays, 5 μl of ^{35}S-*X*-DHFR (0.5 mg/ml in storage buffer) were mixed with 20 μl of yeast extract, incubated for 20 min at 30 °C, and thereafter chilled on ice. Samples (5 μl) were applied onto isoelectric focusing (IEF) polyacrylamide plates, pH 3.5–9.5 (Pharmacia Biotech Inc.), precooled to 10 °C. IEF was carried out for 80 min at 30 watts in a cooled IEF apparatus (Hoefer Scientific Instruments, San Francisco, CA). The plates were soaked in 100 ml of 10% CCl_3COOH, 5% 5-sulfosalicylic acid; stained with Coomassie Blue to detect IEF markers (Pharmacia Biotech Inc.); and autoradiographed. ^{35}S in the bands of more acidic (deamidated) and more basic (initial) *X*-DHFR species was quantified using a PhosphorImager (Molecular Dynamics, Inc., Sunnyvale, CA).

Pulse-Chase Analysis—*S. cerevisiae nta1*Δ-carrying plasmids that expressed specific *X*-βgal (Ub-*X*-βgal) test proteins and either mouse NTAN1 (pSG61; see above) or *S. cerevisiae* Nta1p (pRB201 (27); a gift from Dr. R. T. Baker) were labeled with Tran^{35}S-label for 5 min, followed by a chase in the presence of cycloheximide, preparation of extracts, immunoprecipitation of *X*-βgal, SDS-polyacrylamide gel electrophoresis, autoradiography, and quantitation essentially as described (27–30, 58, 59).

RESULTS

Cloning and Analysis of the Mouse cDNA Encoding Nt^N-amidase—Stewart *et al.* (39) have purified an amidohydrolase, termed Nt^N-amidase (also called protein-N-terminal asparagine amidohydrolase), from porcine liver that specifically deamidates N-terminal Asn in polypeptides, but does not deamidate free Asn. Stewart *et al.* (40) have also determined the amino acid sequence of several Nt^N-amidase fragments and used it to clone an ~1.3-kb porcine cDNA (*Ntan1* cDNA) encoding this enzyme. To isolate a mouse *Ntan1* cDNA, we screened a λgt11-based mouse liver cDNA library with a porcine cDNA-derived probe and low-stringency hybridization (see "Experimental Procedures"). The initially isolated 1560-bp mouse DNA insert contained a 114-bp stretch that was 85% identical to a region of the porcine *Ntan1* mRNA; an even greater similarity between this region and its porcine counterpart was apparent at the level of deduced amino acid sequences. This 114-bp region was flanked by consensus sequences for mammalian intron/exon junctions (60), suggesting that the library's 1560-bp mouse DNA insert was derived from an unspliced or partially spliced precursor of *Ntan1* mRNA. It is shown below that the 114-bp region of the 1560-bp insert is exon IX of the mouse *Ntan1* gene.

Northern hybridization with the 1560-bp fragment as a probe detected an ~1.4-kb mRNA whose relative abundance in several mouse cell lines, including Friend erythroleukemia (MEL) cells, was significantly higher than in mouse liver (see below). Therefore, we screened a cDNA library from mouse MEL-C19 cells with the 1560-bp DNA fragment as a probe. This screen yielded a 1030-bp mouse DNA insert containing a 897-bp region that encoded a protein whose deduced amino acid sequence was 83% identical to the deduced sequence of porcine Nt^N-amidase, but lacked a region homologous to the sequence of the N-terminal 11 residues of the porcine enzyme. The method of 5′-RACE-PCR (see "Experimental Procedures") was used to isolate the missing mouse cDNA sequence.

The resulting mouse *Ntan1* cDNA clone contained a 930-bp ORF encoding a 310-residue (35 kDa) NTAN1 protein (Nt^N-amidase) with a calculated pI of 6.1, whose deduced amino acid sequence (Fig. 2*A*) was 88% identical to the deduced sequence of porcine Nt^N-amidase (40). The recently determined (deduced) sequence of human Nt^N-amidase is 92 and 91% identical to the deduced sequences of mouse and porcine Nt^N-amidases, respectively.[6] The sequences of these Nt^N-amidases lack signif-

[6] S. Yu, A. E. Stewart, S. M. Arfin, and R. A. Bradshaw, manuscript in preparation.

icant similarities to the sequences of other known proteins, including other amidotransferases. In particular, the sequence of the 35-kDa mouse NTAN1 protein was dissimilar to the sequence of the 52-kDa *S. cerevisiae* amidase (Nt-amidase) encoded by the *NTA1* gene. (In contrast to Nt^N-amidase, the *S. cerevisiae* Nta1p Nt-amidase (a component of the yeast N-end rule pathway) deamidates N-terminal Asn or Gln (Fig. 1).) The deduced sequence of mouse NTAN1 (Fig. 2*A*) lacks motifs resembling membrane-spanning regions, signal sequences, or nuclear localization signals. Since Nt^N-amidase is likely to be a part of a multiprotein targeting complex analogous to the one observed in yeast (4), the absence of nuclear localization signal-like sequences in NTAN1 does not, by itself, preclude the possibility that Nt^N-amidase is located in both the cytosol and the nucleus.

A partially translated 208-bp region that begins at nucleotide 896 of the 930-bp mouse *Ntan1* ORF (nucleotide 17031 in Fig. 2*B*) and encompasses the *Ntan1* mRNA poly(A) addition site (Fig. 2*B*) is 98.6% identical to a 206-bp segment within the 3′-flanking untranslated region of the mouse *Il2* gene that encodes interleukin-2. The *Il2*-homologous region of *Ntan1* contains the *Ntan1* stop codon and the poly(A) addition site (Fig. 2*B*); neither of these functions is performed by the corresponding sequences of the *Il2* locus. The near identity of sequences in these two regions of the mouse *Ntan1* and *Il2* loci and the absence of significant sequence similarities between the corresponding regions of human *NTAN1* and *IL2* (data not shown) suggest a recombination event in the mouse lineage that involved *Ntan1* and *Il2* and that occurred after separation of the mouse and human lineages.

Cloning and Analysis of the Mouse Ntan1 Gene—A low-stringency Southern hybridization analysis of the BALB/c mouse kidney DNA digested with *Eco*RI, *Bam*HI, or *Hin*dIII, using the 1560-bp fragment containing a putative 114-bp *Ntan1* exon (see above) as a probe, revealed several hybridizing restriction fragments. A screening of a phage P1-based mouse genomic DNA library (derived from embryonic stem cells) for an insert containing the 114-bp *Ntan1* exon was carried out (see "Experimental Procedures"), yielding several P1 clones bearing ~100-kb mouse DNA inserts. Three of these inserts were found to contain the entire *Ntan1* locus; one of them (P1 clone 1798) was used for nucleotide sequencing and exon mapping.

The mouse *Ntan1* gene was found to span ~17 kb of DNA and to contain 10 exons whose length ranged from 54 to 177 bp (GenBank™/EMBL accession number U57691) (Figs. 2*B* and 3). The 79-bp exon I, encoding the first 26 residues of NTAN1, was located in an ~1.7-kb *Bam*HI fragment, which also contained the putative promoter region of *Ntan1* and the 5′-untranslated leader sequence (Figs. 2*B* and 3). The 177-bp exon X encoded the last 59 residues of NTAN1 and was followed by the 138-bp 3′-untranslated region containing the poly(A) addition site, which was also the boundary of a region homologous to the locus that encodes interleukin-2 (Figs. 2*B* and 3).

Among the intron/exon junctions, all of them except those of introns I and V contained the GT and AG consensus dinucleotides characteristic of mammalian nuclear pre-mRNA splice sites (60, 61). However, intron I contained TG (instead of consensus GT) and AC (instead of consensus AG) at the 5′- and 3′-splice sites, respectively, while intron V contained CG (instead of consensus AG) at the 3′-splice site (Fig. 2*B*). A GT consensus dinucleotide was present 1 bp upstream of the predicted 5′-splice site in intron I, while another consensus dinucleotide, AG, was located at the beginning of the coding region of exon II, 2 bp downstream of the expected splice site (Fig. 2*B*). If these consensus sequences were the actual splice junctions,

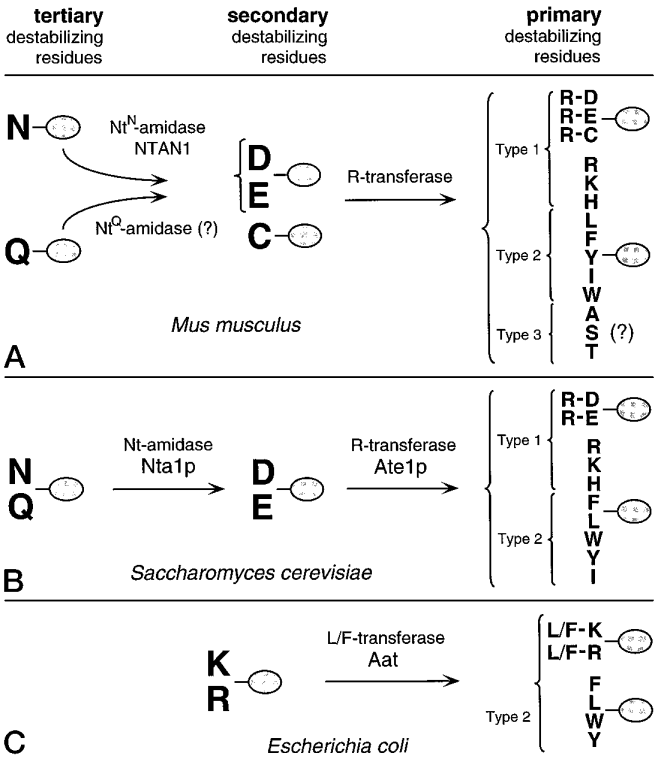

FIG. 1. **Comparison of enzymatic reactions that underlie the activity of N-d^t and N-d^s residues in the N-end rule pathways of different organisms.** *A*, mammalian cells: rabbit reticulocytes and mouse (*M. musculus*) L-cells (7, 8). *B*, the yeast *S. cerevisiae* (6). *C*, the eubacterium *E. coli* (5). The *E. coli* N-end rule lacks N-d^t residues. The postulated mammalian Nt^Q-amidase (a *question mark* in *A*) remains to be identified. It is also unknown whether an N-terminal Cys residue is arginylated by the same species of mammalian R-transferase that arginylates N-terminal Asp or Glu or whether a Cys-specific R-transferase is involved. A *question mark* adjacent to the Ser residue in *A* refers to the fact that Ser is a destabilizing residue in reticulocyte extract (7), but a stabilizing residue in fibroblast-like mouse L-cells *in vivo* (8). *L/F-transferase*, Leu/Phe-tRNA-protein transferase.

the resulting DNA equivalent of the *Ntan1* mRNA sequence in this region would have been CTGGAAAGA rather than CTGGAGGAAAGA actually present in the cloned *Ntan1* cDNA. The latter sequence occurred not only in the *Ntan1* cDNA derived from a MEL cell cDNA library, but also in the products of reverse transcription-PCR (see "Experimental Procedures") carried out using MEL cell mRNA and two independent sets of primers (data not shown). Moreover, the homologous region of porcine cDNA also contained "nonconsensus" splice junctions (40). We conclude that the usually strict constraints on the sequences of splice sites (61, 62) are bypassed at the sites that define intron V and especially intron I of the mouse *Ntan1* gene.

The Ntan1 Promoter—To identify the promoter of mouse *Ntan1*, we first analyzed a region immediately upstream of exon I within the ~1.7-kb genomic *Bam*HI fragment (Figs. 2*B* and 3). Nucleotide sequencing indicated the presence of multiple binding sites for the transcription factor Sp1 (63) immediately upstream of the *Ntan1* ATG start codon. This area (Fig. 2*B*) resembled the GC-rich promoter regions of several weakly transcribed housekeeping genes (64); the similarities included the absence of TATA and CCAAT consensus sequences characteristic of many other eukaryotic promoters. However, given the ~1.4-kb size of *Ntan1* mRNA (see below), the distance between the *Ntan1* start codon and the poly(A) addition site (1068 bp), and the presumed size of the poly(A) tail (100–200 nucleotides), the *Ntan1* promoter was expected to be present ~200 bp upstream of the start codon. The mapping of *Ntan1*

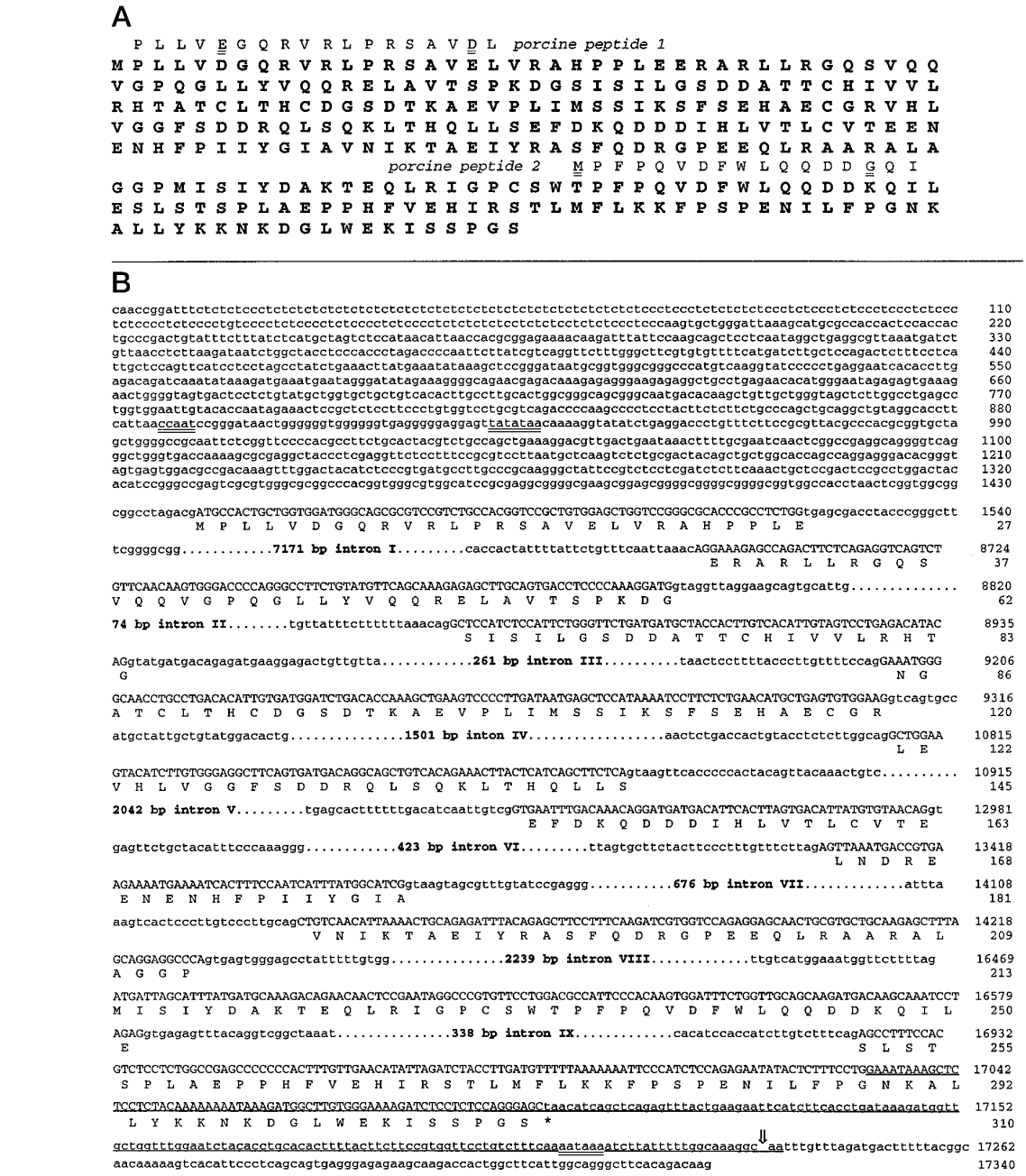

FIG. 2. *A*, the deduced amino acid sequence of mouse Nt^N-amidase (NTAN1). The directly determined sequences of two peptides produced from porcine Nt^N-amidase (40) are shown above the corresponding regions of mouse Nt^N-amidase; divergent residues are *doubly underlined*. *B*, nucleotide sequence of the mouse *Ntan1* gene upstream of its start codon, nucleotide sequences (and the corresponding amino acid sequences) of *Ntan1* exons, and the adjacent intron sequences. The putative promoter consensus elements CCAAT and TATAA are *doubly underlined*. A 208-bp region that encompasses a 3′-region of the *Ntan1* ORF and that is 98.6% identical to a 3′-flanking untranslated region of the mouse *Il2* gene (see "Results") is *singly underlined*. A *double arrow* indicates the location of the poly(A) addition site in the *Ntan1* cDNA (data not shown). A sequence matching the consensus sequence that precedes poly(A) addition sites in mammalian genes is *doubly underlined*. Most (>75%) of the ~17-kb *Ntan1* locus was sequenced on both strands. Nucleotide sequences reported in this paper have been submitted to the GenBank™/EMBL Data Bank with accession numbers U57691 (mouse *Ntan1* genomic DNA) and U57692 (mouse *Ntan1* cDNA).

mRNA by primer extension (see below) also failed to indicate the presence of *Ntan1* mRNAs whose 5′-ends were located within 100 bp of the *Ntan1* start codon.

Nucleotide sequencing revealed a putative TATA box consensus sequence 510 bp upstream of the *Ntan1* start codon. A CCAAT consensus sequence was present 45 bp upstream of this TATA box (Fig. 2*B*). In a typical TATA/CCAAT-containing eukaryotic promoter, transcription starts 16–35 bp downstream of TATA (65, 66). Indeed, a primer extension analysis, using *Ntan1* mRNA as a template and a synthetic oligonucleotide primer matching the 3′-end of the sequence of exon I, mapped a set of major 5′-ends of *Ntan1* mRNA to the region ~18 to ~32 bp downstream of the above TATA box, with the 5′-ends at positions 959 and 960 (Fig. 2*B*) being the most frequent ones (data not shown). On the assumption that the 5′-ends of *Ntan1* mRNA defined by primer extension were the actual transcription start sites (rather than sites resulting from a 5′-processing of initial transcription products), these data indicated that the synthesis of *Ntan1* pre-mRNA begins ~480 bp upstream of the *Ntan1* start codon. Two other TATA box consensus sequences

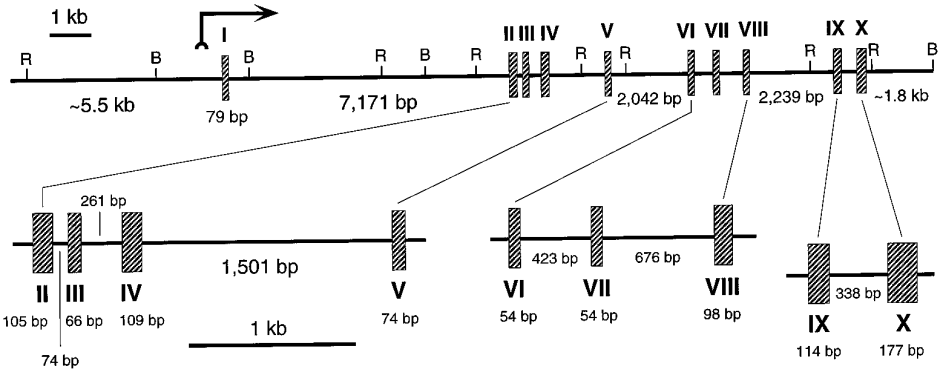

FIG. 3. **The mouse *Ntan1* gene.** The *horizontal line* represents genomic DNA, and the *rectangles* denote exons. The sizes of exons and introns are indicated. An *arrow* upstream of exon I indicates the position of the *Ntan1* transcription start site, ~500 bp from the *Ntan1* start codon, as determined by the primer extension technique. *Bam*HI (*B*) and *Eco*RI (*R*) restriction sites are also indicated. See Fig. 2*B* for the putative promoter consensus elements and "Results" for a partial list of putative binding sites for transcription factors.

were present in the genomic DNA upstream of the sites corresponding to the mapped 5′-ends of *Ntan1* mRNAs (Fig. 2*B*), but these putative TATA boxes were not accompanied by CCAAT boxes.

The ~480-bp distance between *Ntan1* transcription start sites and the *Ntan1* start codon predicts the *Ntan1* mRNA size of 1.7 kb (assuming the mean length of the *Ntan1* mRNA poly(A) tail to be 0.15 kb), which is larger than the apparent size of *Ntan1* mRNA in the Northern experiments (~1.4 kb; see below). This discrepancy may be due to an anomalous electrophoretic mobility of *Ntan1* mRNA, to its cleavage during isolation, or, less likely, to the presence of an untranslated intron in the 5′-region of *Ntan1* pre-mRNA (no evidence for such an intron has been found).

In addition to the TATA and CCAAT sequences, the apparent promoter region ~0.5 kb upstream of the *Ntan1* start codon (Fig. 2*B*) contains putative binding sites for "general" transcription factors such as CP-1, AP-1, E2A, ATF, and NF-I (63, 67, 68) as well as putative sites for other transcription factors such as c/EBP, AP-2, HNF-1, SDR, PEA3, Ets1, H-2DIIBP, NFuE1, NFuE2, PU.1, and HC3 (67, 69–79). The roles, if any, of these transcription factors in the expression of *Ntan1* remain to determined.

Chromosome Mapping of Ntan1—The chromosomal location of *Ntan1* was determined by interspecific backcross analysis using DNA of progeny derived from matings of ((C57BL/6J × *M. spretus*)F1 × C57BL/6J) mice. This interspecific backcross mapping panel has been typed over 2000 loci distributed among all of the mouse 19 autosomes as well as the X chromosome (53). C57BL/6J and *M. spretus* DNAs were digested with several restriction endonucleases and analyzed by Southern hybridization for informative restriction fragment length polymorphisms using a mouse *Ntan1* genomic probe. The ~4-kb *Hin*dIII *M. spretus*-specific restriction fragment length polymorphism (see "Experimental Procedures") was used to follow the segregation of *Ntan1* in backcross mice. The results of this mapping indicated that *Ntan1* is located in the proximal region of mouse chromosome 16 and is linked to *Prm1*, *Cebpd*, and *Igl* (see "Experimental Procedures") (Fig. 4). Although 160 mice were analyzed for every marker to produce the segregation analysis data in Fig. 4, up to 185 mice were typed for some pairs of markers. Each locus was analyzed in pairwise combinations for recombination frequencies using these additional data. The ratios of the total number of mice exhibiting recombinant chromosomes to the total number of mice analyzed for each pair of loci and the most likely gene order were as follows: centromere-*Prm1*-(4:168)-*Ntan1*-(1:166)-*Cebpd*-(1:185)-*Igl*.
The recombination frequencies (expressed as genetic distances in centimorgans ± S.E.) were as follows: *Prm1*-(2.4 ± 1.2)-*Ntan1*-(0.6 ± 0.6)-*Cebpd*-(0.5 ± 0.5)-*Igl*.

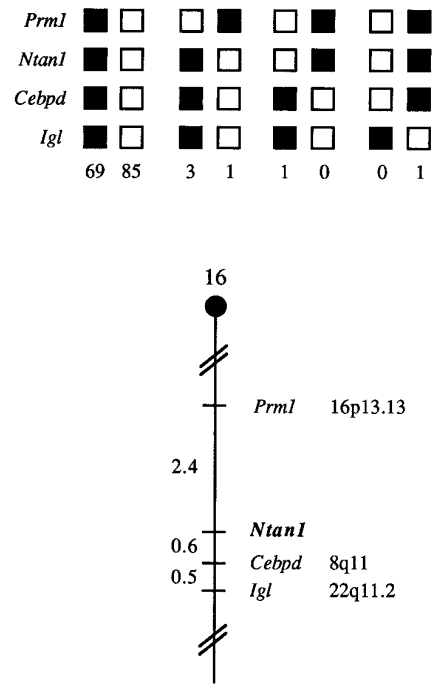

FIG. 4. ***Ntan1* is located in the proximal region of mouse chromosome 16.** The chromosomal position of *Ntan1* was determined using interspecific backcross analysis (see "Experimental Procedures"). The segregation patterns of *Ntan1* and flanking loci in 160 backcross animals that were typed for all loci are shown at the top. For individual pairs of loci, >160 animals were typed (see "Results"). Each column represents the chromosome identified in the backcross progeny that was inherited from the (C57BL/6J × *M. spretus*)F1 parent. The *filled boxes* denote the presence of a C57BL/6J allele. The *unfilled boxes* denote the presence of a *M. spretus* allele. The number of offspring inheriting each type of chromosome is listed at the bottom of each column. A partial chromosome 16 linkage map, showing the location of *Ntan1* in relation to linked genes, is shown at the bottom. Recombination distances between loci (in centimorgans) are shown to the left of the chromosome. The positions of homologous loci in human chromosomes, where known, are shown on the right. References for the human map positions of loci cited here can be obtained from Genome Data Base, a computerized data base of human linkage information maintained by the William H. Welch Medical Library of the Johns Hopkins University (Baltimore, MD).

We have compared our interspecific map of chromosome 16 with a composite mouse linkage map that reports the map location of many uncloned mouse mutations (the mouse genome data base at the Jackson Laboratory, Bar Harbor, ME). *Ntan1* is located in a region of the composite map that is

relatively devoid of uncloned mouse mutations (data not shown). The proximal region of mouse chromosome 16 shares regions of homology with human chromosomes 16p, 8q, and 22q (summarized in Fig. 4) (80). Indeed, recent analyses, using fluorescent *in situ* hybridization and phage P1 clones containing human *NTAN1* genomic DNA, localized the human *NTAN1* gene to chromosome 16p (a region of homology to the chromosomal location of mouse *Ntan1*).[7]

Differential Regulation of mRNAs Encoding NTAN1 and a Specific E2 Enzyme—Northern hybridization was used to compare the levels of *Ntan1* mRNA among mouse tissues and cell lines. In addition to an *Ntan1*-specific probe, these comparisons included a probe for $E2_{14K}$, a mouse E2 enzyme whose sequence is similar to that of *S. cerevisiae* Ubc2p. The latter E2 enzyme (one of at least 11 distinct E2 enzymes in *S. cerevisiae*) is a component of the yeast N-end rule pathway (31, 32). We employed the rabbit $E2_{14K}$ cDNA (52) to isolate a mouse cDNA that encodes $E2_{14K}$ (see "Experimental Procedures"). This cDNA was used as a probe in the Northern hybridizations (Fig. 5, *B* and *D*). The control probes in these analyses were cDNAs encoding mouse β-actin and GAPDH (see "Experimental Procedures") (57). The level of *Ntan1* mRNA in mouse tissues was ~100-fold lower than the level of β-actin mRNA (data not shown).

A Northern blot containing approximately equal amounts of poly(A)$^+$ RNA from various mouse tissues was hybridized with *Ntan1* cDNA. This probe detected a diffuse ~1.4-kb band whose relative abundance (quantified using a PhosphorImager) varied at most 2-fold among different tissues (Fig. 5*A*). The testis-derived *Ntan1* mRNA existed as two species: the minor one comigrated with the ~1.4-kb *Ntan1* mRNA of the other tissues, while the major one had the apparent size of ~1.1 kb (Fig. 5*A*). It is unknown whether the testis-specific *Ntan1* Northern pattern was entirely the result of RNA degradation during isolation. The two species of $E2_{14K}$ mRNA (Fig. 5*B*) were apparently derived from two distinct primary transcripts (see "Experimental Procedures"). In contrast to the relative constancy of *Ntan1* mRNA levels among different tissues, the levels of $E2_{14K}$ mRNA varied significantly, with the highest expression in skeletal muscle (Fig. 5, *A* and *B*). Consistent with this observation was the presence of several putative binding sites for MyoD, a muscle-specific transcription factor (81), in a region upstream of the start codon of the rat $E2_{14K}$ ORF (GenBank™/EMBL accession number U04303) (data not shown); no putative MyoD-binding sites were detected in the *Ntan1* promoter.

A Northern analysis of *Ntan1* and $E2_{14K}$ mRNAs was also carried out with poly(A)$^+$ RNA prepared from Friend erythroleukemia (MEL-C19) and myoblast (C_2C_{12}) cell lines (Fig. 5, *C* and *D*). The relative amount of *Ntan1* mRNA in MEL and C_2C_{12} cells before differentiation was ~3- and ~20-fold higher, respectively, than the amount of *Ntan1* mRNA in skeletal muscle (Fig. 5 and data not shown). Poly(A)$^+$ RNA was also isolated from MEL cells that were induced to differentiate under conditions that resulted in 95% of the cells ceasing growth and producing hemoglobin (see "Experimental Procedures"). The relative levels of *Ntan1* mRNA were approximately the same in the growing, undifferentiated MEL cells and in their differentiated counterparts (Fig. 5*C*). Similar results were obtained with the mouse lymphoma EL4 cells that have been induced by a phorbol ester to differentiate and express interleukin-2 (data not shown). By contrast, the conversion of C_2C_{12} myoblasts (which contained the highest levels of

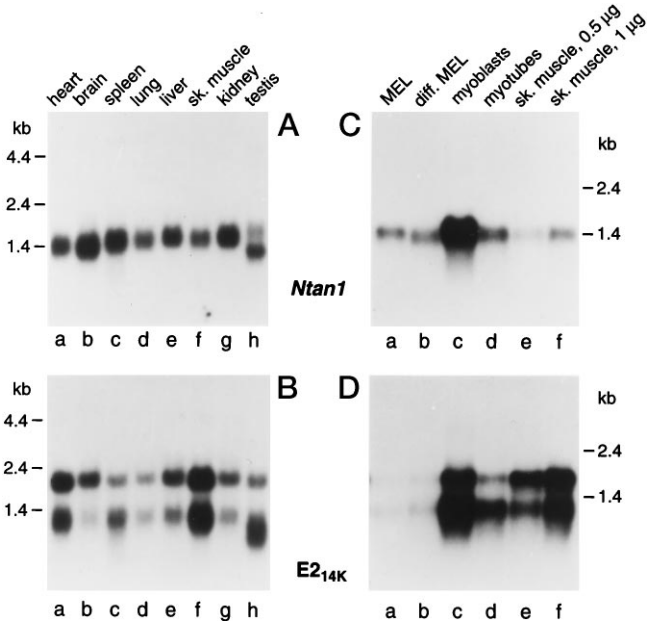

FIG. 5. ***Ntan1* and $E2_{14K}$ mRNAs in mouse tissues and cell lines.** Shown is the Northern hybridization analysis of RNA from the indicated tissues and cell lines, using DNA probes specific for *Ntan1* mRNA (*A* and *C*) and $E2_{14K}$ mRNA (*B* and *D*) (see "Experimental Procedures").

Ntan1 mRNA among the mouse tissues and cell lines examined) into myotubes (see "Experimental Procedures") was accompanied by an ~4-fold decrease in the level of *Ntan1* mRNA, whose relative amount in myotubes was slightly higher than in MEL cells, but significantly higher than in skeletal muscle (Fig. 5*B*).

Mouse NTAN1 Is an N-terminal Asparagine-specific Amidase—The 88% similarity between the sequences of porcine Nt^N-amidase (40) and the mouse NTAN1 protein (Fig. 2*A*) strongly suggested that NTAN1 is also an Nt^N-amidase. To verify this conjecture, we asked whether extracts prepared from *S. cerevisiae nta1*Δ cells that lacked the endogenous yeast Nt-amidase (encoded by *NTA1* (27)) but expressed mouse NTAN1 possess the expected Nt^N-amidase activity. The amidase assay employed ^{35}S-labeled derivatives of mouse DHFR that bore different N-terminal residues (see "Experimental Procedures"). A ^{35}S-*X*-DHFR (*X* = Asn, Gln, or Asp) was incubated in an *S. cerevisiae* extract and subjected to IEF on a urea-containing polyacrylamide gel. This procedure fractionates a protein according to its pI (82). Since deamidation of Asn or Gln converts a neutral side chain into a negatively charged one, this modification should be detectable through a distinct shift in pI. Indeed, while the pI positions of Asn-DHFR and Gln-DHFR were indistinguishable, the pI position of otherwise identical Asp-DHFR that bore N-terminal Asp was shifted toward a lower pI (Fig. 6, *lanes j–m*; and data not shown).

The pI of ^{35}S-Asn-DHFR was not altered after its incubation with an extract from *S. cerevisiae nta1*Δ that lacked the *NTA1*-encoded yeast Nt-amidase (Fig. 6, *lanes j* and *m*). However, the incubation of ^{35}S-Asn-DHFR with an extract from otherwise identical *nta1*Δ cells that expressed mouse NTAN1 resulted in a pI shift of the DHFR substrate (Fig. 6, *lanes g–i* and *p*). In contrast to the yeast *NTA1*-encoded Nt-amidase, which deamidated either Asn-DHFR or (more slowly) Gln-DHFR (Fig. 6, *lanes a–f* and *n–o*), no deamidation of Gln-DHFR was observed with *S. cerevisiae nta1*Δ extracts containing mouse NTAN1 (Fig. 6, *lane q*), strongly suggesting that NTAN1 is specific for N-terminal Asn (see below for independent *in vivo* evidence bearing on this conclusion). In addition, the absence of charge modifications in the case of Gln-DHFR (but not in the case of

[7] U. Bengtsson, A. E. Stewart, R. A. Bradshaw, and S. M. Arfin, manuscript in preparation.

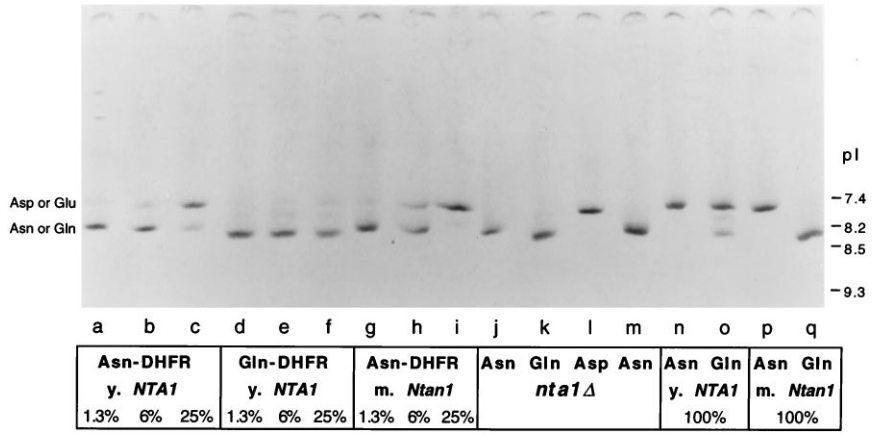

FIG. 6. **Purified, ^{35}S-labeled *X*-DHFR test proteins bearing either Asn, Gln, or Asp at their N termini were incubated with the indicated extract from *S. cerevisiae* and thereafter fractionated by IEF (see "Experimental Procedures")**. The control extract was prepared from *S. cerevisiae nta1*Δ. This extract was tested directly ("*nta1*Δ") or mixed with extracts prepared from otherwise identical *nta1*Δ strains that also carried plasmids expressing either *S. cerevisiae NTA1* ("*y. NTA1*") or mouse *Ntan1* ("*m. Ntan1*"). Percent values shown refer to the relative content of a given extract in a total sample, the rest of it being the extract from *S. cerevisiae nta1*Δ. The IEF positions of *X*-DHFR proteins bearing N-terminal Asp or Glu *versus* Asn or Gln are shown on the left.

Asn-DHFR) that has been incubated with NTAN1 (Fig. 6, *lanes p* and *q*) strongly suggested that the activity of NTAN1 is confined to N-terminal residues (no deamidation of internal Asn or Gln in Gln-DHFR). These findings confirmed and extended the observations made by Stewart *et al.* (39) with porcine Nt^N-amidase, using peptide-size amidase substrates and high pressure liquid chromatography fractionation of reaction products.

Mouse Nt^N-amidase Can Implement the Asparagine-specific Subset of the Yeast N-end Rule—We asked whether the mouse *Ntan1*-encoded Nt^N-amidase (NTAN1) could function in the yeast N-end rule pathway. In particular, we wished to determine whether the confinement of the *in vitro* deamidating activity of mouse Nt^N-amidase (Fig. 7) to N-terminal Asn is retained under *in vivo* conditions. The test proteins were derivatives of *E. coli* βgal. In eukaryotes, Ub fusions such as Ub-*X*-βgal (*X* = a residue at the Ub-βgal junction) are rapidly deubiquitylated by Ub-specific proteases to yield *X*-βgal test proteins (2). In contrast to the function of Ub in protein degradation, the role of Ub in these engineered Ub fusions is to allow the *in vivo* generation of otherwise identical proteins bearing different N-terminal residues. Metabolic stabilities of *X*-βgal proteins could be compared either indirectly (by determining the intracellular concentrations of *X*-βgal proteins) (Fig. 7) or directly (in pulse-chase experiments) (Fig. 8). Previous work (27–32) has shown that the enzymatic activity of an *X*-βgal protein in yeast cells is a sensitive indicator of its metabolic stability.

An *nta1*Δ strain of *S. cerevisiae* that lacked the yeast *NTA1*-encoded Nt-amidase and expressed specific *X*-βgal (Ub-*X*-βgal) proteins such as Met-βgal, Asp-βgal, Asn-βgal, or Gln-βgal was transformed with plasmids expressing either *S. cerevisiae* Nta1p or mouse NTAN1, and the relative metabolic stabilities of *X*-βgal test proteins were determined for each transformant by measuring the activity of βgal (Fig. 7). In wild-type (*NTA1*) *S. cerevisiae*, the level of Met-βgal ($t_{1/2}$ > 20 h (6)) was much higher than the levels of Asp-βgal, Asn-βgal, and Gln-βgal ($t_{1/2}$ = ~3, ~3, and ~10 min, respectively (6)), with the level of Gln-βgal being slightly higher than the levels of the two other normally short-lived *X*-βgal proteins (Fig. 7). The same measurements in congenic *nta1*Δ cells showed that the level of Asp-βgal (bearing a secondary destabilizing N-terminal residue) remained unchanged (in comparison to *NTA1* cells), whereas the levels of Asn-βgal and Gln-βgal (bearing the tertiary destabilizing N-terminal residues) became nearly equal to the level of the long-lived Met-βgal (Fig. 7; see also Ref. 6).

As expected, the expression of a plasmid-borne *S. cerevisiae NTA1* gene in *nta1*Δ cells restored the low levels of Asn-βgal and Gln-βgal without altering the levels of Met-βgal or Asp-βgal (Fig. 7). However, the expression of NTAN1 in *S. cerevisiae nta1*Δ resulted in low levels of Asn-βgal, but did not restore the low levels of Gln-βgal, in contrast to the effect of the yeast Nta1p Nt-amidase (Fig. 7). These findings were confirmed by pulse-chase analyses of Asn-βgal, Gln-βgal, and Arg-βgal degradation in yeast *nta1*Δ cells transformed with a plasmid expressing mouse NTAN1 (Fig. 8). We conclude that the *in vitro* selectivity of mouse Nt^N-amidase for N-terminal Asn was retained *in vivo*.

Although Nt-amidase activity (with Asn-DHFR as a substrate) in extracts from *S. cerevisiae nta1*Δ that expressed NTAN1 was significantly higher than the same activity in extracts from wild-type (Nta1p-containing) *S. cerevisiae* (data not shown), Asn-βgal was more unstable in wild-type yeast than in congenic cells that lacked Nta1p and expressed mouse NTAN1 (Fig. 7). The apparently lower *in vivo* efficiency of the mouse Nt^N-amidase (in comparison to the yeast Nta1p Nt-amidase) in conferring a short half-life on a test protein such as Asn-βgal may be caused by the likely absence of specific interactions between the mouse Nt^N-amidase and the targeting complex of the *S. cerevisiae* N-end rule pathway.

DISCUSSION

Stewart *et al.* (39, 40) reported the purification of a porcine Nt^N-amidase and the cloning of its ~1.3-kb cDNA. Described in the present paper is a further study of this enzyme, carried out with its mouse counterpart. We report the following results.

1) The mouse *Ntan1* gene spans ~17 kb of genomic DNA and contains 10 exons ranging from 54 to 177 bp in length. The ~1.4-kb *Ntan1* mRNA encodes a 310-residue (35 kDa) protein, NTAN1 or Nt^N-amidase.

2) The mapping of *Ntan1*, carried out using interspecific backcross analysis, located it in the proximal region of mouse chromosome 16, in an area devoid of uncloned mouse mutations.

3) The deduced amino acid sequence of mouse Nt^N-amidase lacks motifs that resemble membrane-spanning regions, signal sequences, or nuclear localization signals. The sequence of mouse Nt^N-amidase is 88 and 92% identical to the deduced sequences of porcine and human Nt^N-amidases, respectively. However, there are no significant sequence similarities between Nt^N-amidases and other known amidotransferases, including the *NTA1*-encoded Nt-amidase of the yeast *S. cerevisiae*. The 52-kDa Nta1p protein, a component of the *S. cerevisiae* N-end rule pathway, can deamidate N-terminal Asn or Gln (27), whereas the 35-kDa mammalian Nt^N-amidase can deamidate N-terminal Asn but not Gln (Fig. 1) (39).

4) The putative promoter region of *Ntan1* is located ~500 bp upstream of the *Ntan1* start codon. In addition to the TATA box and CCAAT consensus sequences, this region contains putative

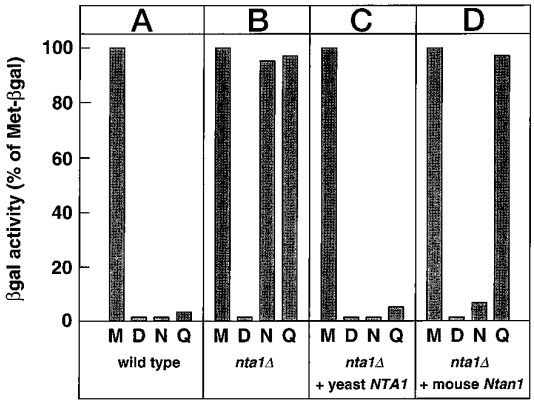

FIG. 7. **Mouse Nt^N-amidase can implement the asparagine-specific subset of the yeast N-end rule.** *A*, levels of βgal activity in wild-type (*NTA1*) *S. cerevisiae* (strain YPH500) expressing *X*-βgal (Ub-*X*-βgal) test proteins, where *X* = Met (*M*), Asp (*D*), Asn (*N*), or Gln (*Q*) (see "Experimental Procedures"). *B*, same as *A*, but in a congenic *nta1*Δ strain. *C*, same as *B*, but *nta1*Δ cells carried a high-copy plasmid expressing the yeast *NTA1* gene from the P_{ADH1} promoter in pRB201 (27). *D*, same as *B*, but *nta1*Δ cells carried a high-copy plasmid expressing the mouse *Ntan1* cDNA from the P_{GAL1} promoter in pSG61 (see "Experimental Procedures"). Values shown are the means from duplicate measurements, which yielded results within 15% of each other. The absolute levels of Met-βgal activity in different *S. cerevisiae* strains were within 20% of each other (data not shown).

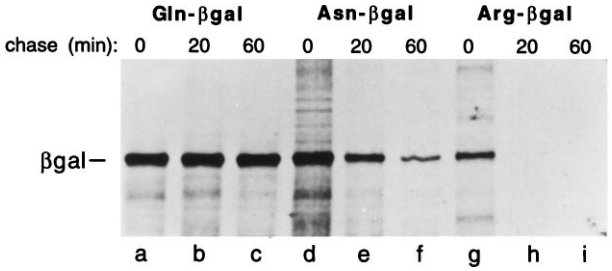

FIG. 8. **Metabolic stabilities of Gln-βgal, Asn-βgal, and Arg-βgal in *S. cerevisiae nta1*Δ that expressed mouse Nt^N-amidase.** Cells were labeled with [^{35}S]methionine/cysteine for 5 min, followed by a chase for 0, 20, and 60 min in the presence of cycloheximide; preparation of extracts; immunoprecipitation of *X*-βgal; and SDS-polyacrylamide gel electrophoresis (see "Experimental Procedures").

binding sites for many of the known transcription factors, including CP-1, AP-1, AP-2, E2A, c/EBP, SDR, PEA3, NFuE1, NFuE2, HNF-1, Ets1, H-2DIIBP, PU.1, and HC3.

5) Northern analyses of the ~1.4-kb *Ntan1* mRNA showed it to be present in roughly equal amounts in all of the mouse tissues examined, including skeletal muscle, brain, and liver. The examined mouse cell lines expressed *Ntan1* more strongly than did mouse tissues. While the *in vitro* differentiation of either Friend or lymphoma cells did not lead to a significant alteration in the amount of *Ntan1* mRNA, its level was found to be decreased ~4-fold upon the conversion of C_2C_{12} myoblasts into myotubes.

6) An isoelectric focusing assay was used to compare the *in vitro* enzymatic specificities of the mouse Nt^N-amidase and the yeast Nt-amidase. These experiments, using ^{35}S-*X*-DHFR test proteins and extracts from *S. cerevisiae nta1*Δ that either lacked or expressed mouse Nt^N-amidase, have shown that Nt^N-amidase deamidates N-terminal Asn but not Gln.

7) The expression of mouse NTAN1 in *S. cerevisiae nta1*Δ was found to restore the degradation of Asn-βgal but not Gln-βgal, indicating that mouse Nt^N-amidase can implement the Asn-specific subset of the yeast N-end rule and also showing that the *in vitro* selectivity of this enzyme for N-terminal Asn is retained *in vivo*.

The hierarchic structure of the N-end rule "disperses" the domains that recognize specific destabilizing N-terminal residues among several proteins such as Nt-amidase, R-transferase, and N-recognin (see the Introduction and Fig. 1). The regulatory possibilities of this arrangement are apparent, but physiological substrates of Nt-amidase and R-transferase remain unknown in both fungi and metazoans. Part of the difficulty in identifying these substrates stems from the fact that although a number of apparently cytosolic or nuclear proteins bear, at least initially, the N-terminal sequence Met-Asn or Met-Gln (39), the known Met-aminopeptidases are incapable of removing N-terminal Met if it is followed by an Asn or a Gln residue (4). In addition, at least the Met-Asn sequence is often acetylated at the N-terminal Met residue (83). It is unknown whether any of the resulting Ac-Met-Asn-containing N-terminal regions are processed *in vivo* to yield N-terminal Asn. Recently, certain sequences bearing N-terminal Met-Gln have been found to yield N-terminal Gln *in vivo* (84). The corresponding processing protease(s) remains to be identified. The isolation of *Ntan1* and the resulting possibility of producing mouse *ntan1*Δ cells through targeted mutagenesis may provide additional routes to physiological substrates of Nt^N-amidase, for instance, through a search for proteins whose metabolic stabilities (or isoelectric points) differ between *Ntan1* and *ntan1*Δ cells.

The possibility that certain cell types in a multicellular organism may lack Nt^N-amidase, yielding, in these cells, an N-end rule pathway that spares normally short-lived proteins bearing N-terminal Asn, has been made less likely by the finding that *Ntan1* is expressed in all of the mouse tissues and cell lines examined. However, the Northern-based *Ntan1* expression assays used thus far (Fig. 5) have been crude, their results being still compatible with the possibility of strong local differences in the levels of *Ntan1* expression within specific tissues. Furthermore, if a binding site for Nt^N-amidase in a targeting complex of the mammalian N-end rule pathway overlaps with a binding site for the postulated Nt^Q-amidase, the competition between these enzymes for binding to the targeting complex and the likely presence of substrate channeling in the N-end rule pathway (27) may strongly influence the degradation of N-end rule substrates bearing N-terminal Asn *versus* those bearing N-terminal Gln. This hypothesis may account for the paradoxically slow degradation of Gln-βgal (but not Asn-βgal) in C_2C_{12} myoblasts, which exhibited the highest level of *Ntan1* mRNA among the tested mouse tissues and cell lines, but contained the postulated Nt^Q-amidase as well, as indicated by *in vitro* deamidation tests with extracts from C_2C_{12} cells.[5]

The apparent absence of *Ntan1* homologs in the mouse genome (those that can be detected by a low-stringency Southern hybridization) (data not shown) suggests that the postulated Nt^Q-amidase might be encoded by a differentially spliced mRNA derived from the *Ntan1* gene. Indeed, although no differential splicing of mouse *Ntan1* pre-mRNA has been detected thus far, the GenBank™/EMBL Data Bank was found to contain partial nucleotide sequences of apparent human *NTAN1* cDNAs (accession numbers H27710, H28425, R10897, N98532, H44229, N93132, N80205, and N69930), one of which (H28425), but not the others, *lacks* exon II of *Ntan1* (data not shown). In addition, intron I (Figs. 2*B* and 3), which abuts exon II, contains noncanonical dinucleotide sequences at both 5′- and 3′-splice sites (see "Results"). This circumstantial evidence for the hypothesis of a single gene encoding both Nt^N-amidase and Nt^Q-amidase remains to be verified directly.

The hierarchic organization of N-end rules, with their tertiary, secondary, and primary destabilizing residues (N-d^t, N-d^s, and N-d^p, respectively), is a feature that is more conserved in evolution than the Ub dependence of an N-end rule pathway or

the identity of enzymatic reactions that mediate the hierarchy of destabilizing residues (4). For example, in a bacterium such as *E. coli*, which lacks the Ub system, the N-end rule has both N-d^s and N-d^p residues (it lacks N-d^t residues) (Fig. 1) (5, 85). The identities of N-d^s residues in *E. coli* (Arg and Lys) are different from those in eukaryotes (Asp and Glu in yeast; Asp, Glu, and Cys in mammals). Bacterial and eukaryotic enzymes that implement the coupling between N-d^s and N-d^p residues are also different: Leu/Phe-tRNA-protein transferase in *E. coli* and R-transferase in eukaryotes. Note, however, that bacterial Leu/Phe-tRNA-protein transferase and eukaryotic R-transferase catalyze reactions of the same type (conjugation of an amino acid to an N-terminal residue of a polypeptide) and use the same source of activated amino acid (aminoacyl-tRNA) (Fig. 1).

The apparent confinement of R-transferase to eukaryotes and of Leu/Phe-tRNA-protein transferase to prokaryotes (Fig. 1) suggests that secondary destabilizing residues were recruited late in the evolution of the N-end rule, after the divergence of prokaryotic and eukaryotic lineages. The lack of sequence similarity between the yeast Nt-amidase and the mammalian Nt^N-amidase and the more narrow specificity of the mammalian enzyme (Fig. 1) suggest that tertiary destabilizing residues (Asn and Gln) became a part of the N-end rule pathway much later yet, possibly after the divergence of metazoan and fungal lineages. If so, the N-end rule pathway may be an especially informative witness of evolution: the ancient origins of this proteolytic system, the simplicity and discreteness of changes in the rule books of N-end rules among different species, and the diversity of proteins that either produce or target the N-degron should facilitate phylogenetic deductions once the components of this pathway become characterized across a broad range of organisms.

Although the properties of the *Ntan1*-encoded mouse Nt^N-amidase are fully consistent with its being a component of the mouse N-end rule pathway and although mouse Nt^N-amidase was shown to be capable of implementing the Asn-specific subset of the yeast N-end rule *in vivo*, a rigorous test of the presumed function of this enzyme requires the production of mouse cell lines and/or whole animals that lack *Ntan1*. Work in this direction is under way.

Acknowledgments—We thank the colleagues whose names are cited in the text for gifts of cell lines, strains, and plasmids. S. G. is grateful to R. J. Dohmen and K. Madura for the introduction to methods of yeast genetics and to L. Larson and F. Lévy for advice on cell culture techniques. N. G. C. and N. A. J. thank D. J. Gilbert for excellent technical assistance.

REFERENCES

1. Varshavsky, A. (1991) *Cell* **64,** 13–15
2. Bachmair, A., Finley D., and Varshavsky, A. (1986) *Science* **234,** 179–186
3. Varshavsky, A. (1992) *Cell* **69,** 725–735
4. Varshavsky, A. (1996) *Cold Spring Harbor Symp. Quant. Biol.* **60,** 461–478, and references therein
5. Tobias, J. W., Shrader, T. E., Rocap, G., and Varshavsky, A. (1991) *Science* **254,** 1374–1377
6. Bachmair, A., and Varshavsky, A. (1989) *Cell* **56,** 1019–1032
7. Gonda, D. K., Bachmair, A., Wünning, I., Tobias, J. W., Lane, W. S., and Varshavsky, A. (1989) *J. Biol. Chem.* **264,** 16700–16712
8. Lévy, F., Johnsson, N., Rümenapf, T., and Varshavsky, A. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 4907–4912
9. Alagramam, K., Naider, F., and Becker, J. M. (1995) *Mol. Microbiol.* **15,** 225–234
10. Hondermarck, H., Sy, J., Bradshaw, R. A., and Arfin, S. M. (1992) *Biochem. Biophys. Res. Commun.* **189,** 280–288
11. Ota, I. M., and Varshavsky, A. (1993) *Science* **262,** 566–569
12. deGroot, R. J., Rümenapf, T., Kuhn, R. J., Strauss, E. G., and Strauss, J. H. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 8967–8971
13. Madura, K., and Varshavsky, A. (1994) *Science* **265,** 1454–1458
14. Johnson, E. S., Gonda, D. K., and Varshavsky, A. (1990) *Nature* **346,** 287–291
15. Hill, C. P., Johnston, N. L., and Cohen, R. E. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 4136–4140
16. Dohmen, R. J., Wu, P., and Varshavsky, A. (1994) *Science* **263,** 1273–1276
17. Chau, V., Tobias, J. W., Bachmair, A., Marriott, D., Ecker, D. J., Gonda, D. K., and Varshavsky, A. (1989) *Science* **243,** 1576–1583
18. Arnason, T. A., and Ellison, M. J. (1994) *Mol. Cell. Biol.* **14,** 7876–7883
19. Spence, J., Sadis, S., Haas, A. L., and Finley, D. (1995) *Mol. Cell. Biol.* **15,** 1265–1273
20. Scherer, D., Brockman, J. A., Chen, Z., Maniatis, T., and Ballard, D. W. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 11259–11263
21. Hershko, A., and Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–807
22. Jentsch, S. (1992) *Annu. Rev. Genet.* **26,** 179–207
23. Vierstra, R. D. (1993) *Annu. Rev. Plant Physiol.* **44,** 385–410
24. Parsell, D. A., and Lindquist, S. (1993) *Annu. Rev. Genet.* **27,** 437–496
25. Ciechanover, A. (1994) *Cell* **79,** 13–21
26. Deshaies, R. J. (1995) *Trends Cell Biol.* **5,** 428–434
27. Baker, R. T., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 12065–12074
28. Balzi, E., Choder, M., Chen, W., Varshavsky, A., and Goffeau, A. (1990) *J. Biol. Chem.* **265,** 7464–7471
29. Bartel, B., Wünning, I., and Varshavsky, A. (1990) *EMBO J.* **9,** 3179–3189
30. Baker, R. T., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1090–1094
31. Madura, K., Dohmen, R. J., and Varshavsky, A. (1993) *J. Biol. Chem.* **268,** 12046–12054
32. Dohmen, R. J., Madura, M., Bartel, B., and Varshavsky, A. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 7351–7355
33. Rechsteiner, M., Hoffman, L., and Dubiel, W. (1993) *J. Biol. Chem.* **268,** 6065–6068
34. Pühler, G., Pitzer, F., Zwickl, P., and Baumeister, W. (1994) *Syst. Appl. Microbiol.* **16,** 734–741
35. Hochstrasser, M. (1995) *Curr. Biol.* **7,** 215–223
36. Jentsch, S., and Schlenker, S. (1995) *Cell* **82,** 881–884
37. Rubin, D. M., and Finley, D. (1995) *Curr. Biol.* **5,** 854–858
38. Hilt, W., and Wolf, D. H. (1996) *Trends Biochem. Sci.* **21,** 96–102
39. Stewart, A. E., Arfin, S. M., and Bradshaw, R. A. (1994) *J. Biol. Chem.* **269,** 23509–23517
40. Stewart, A. E., Arfin, S. M., and Bradshaw, R. A. (1995) *J. Biol. Chem.* **270,** 25–28
41. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (1992) *Current Protocols in Molecular Biology*, Wiley-Interscience, New York
42. Sherman, F., Fink, G. R., and Hicks, J. B. (1986) *Methods in Yeast Genetics*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
43. Dohmen, R. J., Strasser, A. W. M., Honer, C., and Hollenberg, C. P. (1991) *Yeast* **7,** 691–692
44. Sikorski, R. S., and Hieter, P. (1989) *Genetics* **122,** 19–27
45. Gietz, D., St. Jean, A., Woods, R. A., and Schiestl, R. H. (1992) *Nucleic Acids Res.* **20,** 1425–1426
46. Rothstein, R. (1991) *Methods Enzymol.* **194,** 281–301
47. Jacoby, W. B., and Pastan, I. H. (eds) (1988) in *Methods in Enzymology: Cell Culture,* Vol. 58, Academic Press, San Diego, CA
48. Frohman, M. A., Dush, M. K., and Martin, G. R. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 8998–9002
49. Schuster, D. M., Buchman, G. W., and Rashtchian, A. (1992) *Focus* (*Idaho*) **14,** 46–52
50. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M., and Davis, R. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 1731–1735
51. Pierce, J. C., Sternberg, N., and Sauer, B. (1992) *Mamm. Genome* **3,** 550–558
52. Wing, S., Dumas, F., and Banville, D. (1992) *J. Biol. Chem.* **267,** 6495–6501
53. Copeland, N. G., and Jenkins, N. A. (1991) *Trends Genet.* **7,** 113–118
54. Jenkins, N. A., Copeland, N. G., Taylor, B. A., and Lee, B. K. (1982) *J. Virol.* **43,** 26–36
55. Jenkins, N. A., Gilbert, D. J., Cho, B. C., Strobel, M. C., Williams, S. C., Copeland, N. G., and Johnson, P. F. (1995) *Genomics* **28,** 333–336
56. Green, E. L. (1981) *Genetics and Probability in Animal Breeding Experiments*, pp. 77–113, Oxford University Press, New York
57. Sabath, D. E., Broome, H. E., Prystowsky, M. B. (1990) *Gene* (*Amst.*) **91,** 185–191
58. Johnson, E. S., Bartel, B., Seufert, W., and Varshavsky, A. (1992) *EMBO J.* **11,** 497–505
59. Johnson, E. S., Ma, P. C. M., Ota, I. M., and Varshavsky, A. (1995) *J. Biol. Chem.* **270,** 17442–17456
60. Shapiro, M. B., and Senapathy, P. (1987) *Nucleic Acids Res.* **15,** 7155–7174
61. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., and Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150
62. Mount, S. M. (1996) *Science* **271,** 1690–1692
63. Mitchell, P. J., and Tjian, R. (1989) *Science* **245,** 371–378
64. Sehgal, A., Patil, N., and Chao, M. (1988) *Mol. Cell. Biol.* **8,** 3160–3167
65. Bucher, P., and Trifonov, E. (1986) *Nucleic Acids Res.* **14,** 10009–10026
66. Buratowski, S., and Sharp, P. (1992) in *Transcriptional Regulation* (McKnight, S. L., and Yamamoto, K. R., eds) pp. 227–246, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
67. Johnson, P. F., and McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58,** 799–839
68. Faisst, S., and Meyer, S. (1992) *Nucleic Acids Res.* **20,** 3–26
69. Umek, R. M., Friedman, A. D., and McKnight, S. L. (1991) *Science* **251,** 288–292
70. Tronche, F., and Yaniv, M. (1992) *BioEssays* **14,** 579–587
71. Osborne, T. F., Gil, G., Goldstein, J. L., and Brown, M. S. (1988) *J. Biol. Chem.* **263,** 3380–3387
72. Xin, J. H., Cowie, A., Lachance, P., and Hassel, J. A. (1992) *Genes Dev.* **6,** 481–496
73. Wasylyk, B., Hahn, S. L., and Giovane, A. (1993) *Eur. J. Biochem.* **211,** 7–18
74. Augereau, P., and Chambon, P. (1986) *EMBO J.* **5,** 1791–1797
75. Romano-Spica, V., Georgiou, P., Suzuki, H., Papas, T. S., and Bhat, N. K. (1995) *J. Immunol.* **154,** 2724–2732
76. Thompson, C. B., Wang, C. Y., Ho, I. C., Bohjanen, P. R., Petryniak, B., June, C. H., Miesfeldt, S., Zhang, L., Nabel, G. J., Karpinski, B., and Leiden, J. M.

(1992) *Mol. Cell. Biol.* **12,** 1043–1053
77. Schwarzenbach, H., Newell, J. W., and Matthias, P. (1995) *J. Biol. Chem.* **270,** 898–907
78. Sawyers, C. L., and Denny, C. T. (1994) *Cell* **77,** 171–173
79. Lenardo, M., Pierce, J. W., and Baltimore, D. (1987) *Science* **236,** 1573–1577
80. Copeland, N. G., Jenkins, N. A., Gilbert, D. J., Eppig, J. T., Maltais, L. J., Miller, J. C., Dietrich, W. F., Weaver, A., Lincoln, S. E., Steen, R. G., Stein, L. D., Nadeau, J. H., and Lander, E. S. (1993) *Science* **262,** 57–66
81. Weintraub, H., Dwarki, V. J., Verma, I., Davis, R., Hollenberg, S., Snider, L., Lassar, A., and Tapscott, S. J. (1991) *Genes Dev.* **5,** 1377–1386
82. Righetti, P. G., Gianazza, E., Gelfi, C., and Chiari, M. (1990) in *Gel Electrophoresis of Proteins* (Hames, B. D., and Rickwood, D., eds) 2nd Ed., pp. 149–216, IRL Press, Oxford
83. Arfin, S. M., and Bradshaw, R. A. (1988) *Biochemistry* **27,** 7979–7984
84. Ghislain, M., Dohmen, R. J., Lévy, F., and Varshavsky, A. (1996) *EMBO J.*, in press
85. Shrader, T. E., Tobias, J. W., and Varshavsky, A. (1993) *J. Bacteriol.* **175,** 4364–4374
86. Stewart, A. (1995) *Trends in Genetics Genetic Nomenclature Guide*, Elsevier Science Ltd., Cambridge, United Kingdom, and references therein