THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 271, No. 5, Issue of February 2, pp. 2789–2794, 1996
Printed in U.S.A.

Identification of a Novel Family of Ubiquitin-conjugating Enzymes with Distinct Amino-terminal Extensions*

(Received for publication, September 11, 1995, and in revised form, November 15, 1995)

Kai Matuschewski‡, Hans-Peter Hauser§¶, Mathias Treier¶‖, and Stefan Jentsch‡‖**

From the ‡Zentrum für Molekulare Biologie der, Universität Heidelberg, Im Neuenheimer Feld 282, D-69120 Heidelberg, Federal Republic of Germany, §Behring Werke, Postfach 1140, 3550 Marburg, Federal Republic of Germany, the ‖Howard Hughes Medical Institute, University of California, San Diego, La Jolla, California 92093-0648, and the ‖Friedrich Miescher Laboratory, Max Planck Society, 72076 Tübingen, Federal Republic of Germany

The ubiquitin/proteasome system is the main eukaryotic nonlysosomal protein degradation system. Substrate selectivity of this pathway is thought to be mediated in part by members of a large family of ubiquitin-conjugating (E2) enzymes, which catalyze the covalent attachment of ubiquitin to proteolytic substrates. E2 enzymes have a conserved ~150-residue so-called UBC domain, which harbors the cysteine residue required for enzyme-ubiquitin thioester formation. Some E2 enzymes possess additional carboxyl-terminal extensions that are involved in substrate specificity and intracellular localization of the enzyme. Here we describe a novel family of E2 enzymes from higher eukaryotes (*Drosophila*, mouse, and man) that have amino-terminal extensions but lack carboxyl-terminal extensions. We have identified four different variants of these enzymes that have virtually identical UBC domains (94% identity) but differ in their amino-terminal extensions. In yeast, these enzymes can partially complement mutants deficient in the UBC4 E2 enzyme. This indicates that members of this novel E2 family may operate in UBC4-related proteolytic pathways.

In eukaryotes, selective protein degradation is largely mediated by the ubiquitin/proteasome system (for reviews, see Refs. 1–5). Degradation by this system was recently found to be instrumental in a variety of cellular functions such as DNA repair, cell cycle progression, signal transduction, transcription, and antigen presentation. Known substrates of this pathway include transcription factors (MATα2, GCN4, c-Jun, p53, NF-κB), protein kinases (Mos), cyclins, inhibitors of cyclin-dependent kinases (SIC1, p27), and subunits of trimeric G proteins (for review, see Refs. 1–5). Moreover, the ubiquitin/proteasome system also eliminates abnormal proteins, *e.g.* misfolded, mislocalized, or misassembled proteins.

Substrate recognition by this pathway involves a specialized recognition and targeting apparatus, the ubiquitin-conjugating system, which operates spatially detached from the proteasome. Proteins recognized by this system are earmarked by the covalent attachment of ubiquitin, a small and highly stable protein. In most cases, ubiquitination involves the formation of multiubiquitin chains attached to the substrate that are subsequently recognized by a specific receptor of the (26 S) proteasome. Proteins bound to the receptor are then probably unfolded and translocated into the central cavity of the proteasome where they are degraded to small polypeptides. Ubiquitin chains are released from substrates and recycled to single ubiquitin moieties (for review, see Ref. 5).

Ubiquitin conjugation involves a reaction cascade (1, 3, 6, 7). Initially, ubiquitin-activating (E1)[1] enzyme hydrolyses ATP and forms a thioester bond between itself and ubiquitin. Ubiquitin is then passed on to ubiquitin-conjugating (E2) enzymes and often subsequently to ubiquitin ligases (E3). Each step involves the formation of a thioester-linked ubiquitin-enzyme (E1, E2, or E3) intermediate (7). E2 and/or E3 enzymes finally catalyze isopeptide formation between the carboxyl terminus of ubiquitin and ϵ-amino groups of internal lysine residues of target proteins. Both E2 and E3 enzymes exist as protein families, and diverse combinations of E2·E3 enzyme complexes are thought to define the substrate specificity of the conjugation system.

In the yeast *Saccharomyces cerevisiae*, 12 different genes for ubiquitin-conjugating enzymes (UBC genes) have been detected to date (3, 6). Genetic studies revealed that the encoded enzymes mediate strikingly diverse functions such as DNA repair (8), sporulation (3, 8), cell cycle progression (9, 10), peroxisome biogenesis (11), membrane-protein degradation (12), heat shock resistance (13), and cadmium tolerance (15). One of the most prominent E2 enzymes from yeast is UBC4 (13). A principal function of UBC4 and the highly related UBC5 enzyme appears to be the degradation of abnormal proteins as indicated by the sensitivity of *ubc4 ubc5* double mutants to heat shock, canavanine (an arginine analog), and cadmium (13, 15). In addition, UBC4/5-mediated proteolysis is important for some regulatory processes. One example is the UBC4/UBC5-mediated degradation of the yeast transcription factor MATα2 involved in mating type control (16). UBC4/UBC5 homologs have been described from several organisms including *Drosophila* (*UbcD1*; Ref. 17), *Caenorhabditis elegans* (*ubc-2*; Ref. 18) and man (*UbcH5*; Ref. 19). The function of these enzymes in these organisms is not known, but, in the cases tested, the respective genes can fully complement yeast *ubc4 ubc5* mutants. Interestingly, vertebrate UBC4 homologs can mediate p53 (19) and cyclin (20) ubiquitination *in vitro*, suggesting that UBC4/5-mediated degradation may be of central regulatory importance.

Here we report the identification of a novel family of ubiq-

* This work was supported by grants (to S. J.) from Deutsche Forschungsgemeinschaft, German-Israeli Foundation of Research and Development, Human Frontier Science Program, and Fonds der chemischen Industrie. 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™/EMBL Data Bank with accession number(s) X92663, X92664, and X92665.

** To whom correspondence should be addressed. Tel.: 6221-56-8416; Fax: 6221-56-5891; E-mail: jentsch@sun0.urz.uni-heidelberg.de.

[1] The abbreviations used are: E1, ubiquitin activating enzyme; E2, ubiquitin-conjugating enzyme; E3, ubiquitin ligase; PCR, polymerase chain reaction; kb, kilobase pair(s); bp, base pair(s).

uitin-conjugating enzymes from higher eukaryotes that are related in function and sequence to yeast UBC4/5. In contrast to UBC4 (and its homologs from higher eukaryotes) these enzymes have amino-terminal extensions. The UBC domain of these enzymes from *Drosophila*, mouse, and man is exceptionally highly conserved, exhibiting 94% sequence identity over a stretch of 149 amino acid residues. In contrast, the amino-terminal extensions of four different members of this family exhibit little sequence similarity and may define substrate specificity or are involved in the regulation of these enzymes. In each species, probably all four (or possibly more) different members of this UBC subfamily are expressed, suggesting that each one of these enzymes fulfils special tasks.

EXPERIMENTAL PROCEDURES

Bacterial and Yeast Strains, Media—Growth and handling of *Escherichia coli* were by standard techniques (21). Strain XL1 was used as a host for plasmids. Transformation was carried out using the electroporation procedure. λ phage was propagated on strain C600*hfl*. *S. cerevisiae* cultures were grown in rich (YP) or synthetic (S) media containing either 2% glucose (YPD or SD media), 2% raffinose (YPRaf or SRaf media), or 2% galactose (YPGal or SGal media) as carbon sources. Yeast transformation was carried out using standard protocols (21). The *ubc4 ubc5* yeast strain is a haploid *MAT***a** derivative of YWO23 (14). Growth at 30 and 37 °C was assayed by streaking transformed *ubc4 ubc5* mutant cells pregrown on SGal plates at 23 °C onto YPGal plates and incubating at either 30 or 37 °C for 5 or 7 days, respectively.

Isolation and Sequencing of UBC cDNAs—Plasmid isolation, cloning, and screening methods were according to standard protocols (21). New *UBC* genes were identified with degenerate oligonucleotide primer A (GGGAATTCGGICC(A/T)ICIC(A/G)I(A/T)CICCITA(T/C)(G/A)(A/C)IG(G/A)(A/T/C)GG), primer B (ATTCTAGAGGTGGI(G/T)(A/T)I(A/T)(A/T)IGG(A/G)(A/T)A(A/C/G/T)TC), and primer C (ATTCTAGA(G/C)(A/G)T(T/C)(T/C)TTIA(G/A)IAT(G/A)TCIA(G/A)(G/A)CA) (I, inosine) corresponding to amino-acid sequence GP*X*(A/G)TPY*X*(G/D)G, the amino acid sequence (D/E)YP*X*(S/K)PP (*X*, any residue) and the region adjacent to the active cysteine, conserved in many yeast UBC proteins, respectively. PCR reactions were carried out as described previously (17). To clone *Drosophila melanogaster UBC* genes, primers A and C and genomic *D. melanogaster* Oregon P2 DNA as template were used. Fragments of mouse *UBC* genes were amplified with primers B and C and genomic mouse DNA as template. Aliquots of the genomic DNA were digested with *Eco*RI, *Sal*I, or *Xba*I, pooled, and 2 μg per reaction was used. Reaction products were separated on 6% polyacrylamide gels and DNA from bands of the correct size were eluted. The PCR-generated fragments were digested with *Eco*RI and *Xba*I and subcloned into M13mp18/19 vectors (21). *Drosophila* cDNA were isolated from a pNB40-based *D. melanogaster* cDNA library, carrying cDNA inserts from oligo(dT)-primed *Drosophila* RNA from 0–24-h-old embryos (22), using radiolabeled PCR-generated fragment as a probes. Positive clones were purified, and the 1.25-kb *Eco*RI-*Hin*dIII-digested cDNA insert was subcloned into the M13mp18/19 vectors. *UbcM2* cDNA was isolated by screening a λgt10-based cDNA library, carrying cDNA inserts from oligo(dT)-primed mouse RNA from embryonic mouse fibroblasts (kindly provided by Dr. P. Ekblom, Uppsala, Sweden), with the PCR-generated fragments as a radiolabeled probe. Positive clones were purified and *Eco*RI digested. Due to an internal *Eco*RI restriction site within the *UbcM2* cDNA, we obtained two fragments (700 and 900 bp), which were subsequently ligated to obtain the full-length clone in Bluescript plasmid (Stratagene, San Diego, USA). Cloning of *UbcM3* cDNA was done by screening a λgt10-based cDNA library, carrying cDNA inserts from random-primed mouse RNA from adult mouse Balb/c brain tissue (Clontech, Palo Alto, USA), with the PCR-generated fragments as radiolabeled probes. Positive clones were purified, and the 1.32-kb *Eco*RI-digested cDNA insert was subcloned. DNA sequences of the three cDNAs were determined after subcloning inserts of appropriate lengths with the Applied Biosystems model 373 DNA sequencer. For each cDNA, both strands were sequenced completely with sufficient overlaps. DNA and deduced amino acid sequences were compared with data bases (GenBank, EMBL, PIR, SWISS-PROT) using the BLAST algorithm (23). Alignments were carried out by the BoxAlign program (GCG package).

Northern Blot Analysis—Isolation of RNA from *D. melanogaster* embryos and mouse liver tissue was prepared using standard protocols (24, 21). Resolution of RNA in a 1.0% agarose gel following glyoxal modification was carried out as described previously (21). Separated RNA was transferred to GeneScreen (DuPont NEN) and hybridized with corresponding probes. The probe for *UbcD2* RNA was generated by subcloning the 315-bp *Pst*I-*Ssp*I fragment of the *UbcD2* cDNA (nucleotides 319–634 of Fig. 1) into Bluescript plasmid. After recovering the fragment by *Pst*I-*Kpn*I digestion, eluted DNA was used as a template for the random-primed labeling kit (Boehringer Mannheim) to generate a radiolabeled hybridization probe. Likewise, the probes for *UbcM2* and *UbcM3* were obtained by subcloning and labeling the 335-bp *Eco*RI-*Bam*HI fragment of the *UbcM2* cDNA (nucleotides 194–528 of Fig. 1) and the 240-bp *Hin*dIII-*Ssp*I-digested fragment of the *UbcM3* cDNA (nucleotides 127–364 of Fig. 1), respectively. Northern blot hybridization and detection of the hybridized probe were carried out by standard procedures (21).

Western Blot Analysis—Isolation of yeast proteins, generation of polyclonal antibodies and Western blot analysis were performed essentially as described previously (14, 21). Antibodies were raised against the *E. coli*-expressed UBC domain of UbcD2. Due to the unique conservation of this domain, the antibody reacts well with UbcM3 (see Fig. 4*B*).

Construction of Plasmids—For heterologous expression of the identified cDNAs in the yeast *ubc4 ubc5* mutant, the cDNAs were cloned into yeast expression vectors. The 1.1-kb *Asu*II-*Eco*RI fragment of *UbcD2* was cloned into a derivative of pSEY8, a 2 micron based plasmid carrying the *URA3* marker and the *UBC1* promoter (provided by J. Jungmann, Heidelberg, Germany). Yeast *UBC4* was under the control of its own promoter on pSEY8 (14). The 880-bp *Sma*I-*Stu*I fragment of *UbcM2* was subcloned into Bluescript plasmid and subsequently cloned into pSGal3T, a YEplac112 (Ref. 25) based plasmid carrying the inducible *GAL1* promoter and the transcription terminator of the *ADH* gene (provided by S. Smith, Heidelberg, Germany). *UbcM3* and truncated *UbcM3* were cloned by PCR using *UbcM3* cDNA as template. Amplification of full-length *UbcM3* was done with primer 1 (GCTCTAGATAATGTCGGATGACGATTCG) and primer 2 (CCCGTCGACCCAGTTTATGTAGCTG). The 600-bp PCR product was digested with *Xba*I and *Sal*I, cloned into Bluescript plasmid for sequence determination, and subsequently cloned into vector pSGal3T (see above). Amplification of truncated *UbcM3* was achieved with primer 3 (GCTCTAGATAATGTCTAGCGCTAAGAGGATCC) and primer 2. Primer 3 generated a new ATG start codon together with a TCT codon encoding serine (a highly conserved second amino acid of yeast UBCs) fused to the codons corresponding to the amino acid sequence SAKRI, corresponding to the residues 4–8 of the conserved UBC domain. The resulting 460-bp PCR product was digested with *Xba*I and *Sal*I and cloned into Bluescript plasmid for sequence determination before transfer into vector pSGal3T (see above).

RESULTS

Cloning of a Novel UBC Gene Family From Higher Eukaryotes—To study the function of the ubiquitin system in higher eukaryotes, we initiated a homology-based screen for *UBC* genes from organisms, which are amenable to genetic analysis such as *Drosophila*, *C. elegans*, and mouse (17, 18). Ubiquitin-conjugating enzymes are highly related proteins showing at least 30% amino acid sequence identity between different members of this enzyme family. In particular, sequences within the UBC domain, which harbors the cysteine residue required for ubiquitin-thioester formation, are highly similar. Primer pairs specific for conserved sequences were designed (see "Experimental Procedures") and used for the PCR. From *Drosophila* and mouse genomic DNA as templates, we were able to amplify fragments of sizes equivalent to the corresponding regions within yeast *UBC* genes. DNA sequence analysis indicated that the amplified fragments correspond to segments of five different genes encoding novel ubiquitin-conjugating enzymes. Interestingly, the deduced amino acid sequence of one of the PCR fragments isolated from *Drosophila* and two fragments isolated from mice were virtually identical, suggesting that we have identified three members of a family of highly related UBC genes (see below). Using the cloned five PCR fragments as probes, we isolated the complete cDNAs for the corresponding genes from a plasmid-born *Drosophila* library and a mouse cDNA library in λ phage, respectively. We named the three *Drosophila* genes *UbcD1, UbcD2*, and *UbcD3*

UbcD2

```
-291                                        GGAGTTTCTAACAAAGATAAAGTCGGCTCTGCGTCGCCTCGTAATTTGCCAACT
-237   GTAATTTAGCGAAATAACAACGATCGCGACCGTGAGCCGCCGAGGAGTTTGGAGCGCCGACACTGCCACCCACCAAGGA
-158   CGCCAGCGGAACAGCCGTGCATCCACTGCGAAGGCCGTGTGCGCGAAAGTCCGAGAGGGCAGCAGGCGACACAGG
 -79   GAGAATCATTTTCGCAAGTGTTGTGATAGAAAAAGTGAAAGAAACTGCAATAGTGAGTTCGAAGAAACGAGACCAAAA

   1   ATG TCT TCA ACC CCA GCA GCG GGC AGT GCC GCC GAG GTG GCC ACC TCC AGT GCC ACC TCA
       Met Ser Ser Thr Pro Ala Ala Gly Ser Ala Ala Glu Val Ala Thr Ser Ser Ala Thr Ser

  61   AAT GCG CCC AGT GCA CCC AGC ACA ACG GCC AGC AAC GTG AGC AAT ACC AGC CAG CCG ACG
       Asn Ala Pro Ser Ala Pro Ser Thr Thr Ala Ser Asn Val Ser Asn Thr Ser Gln Pro Thr

 121   ACT GCC GGA ACT CCG CAG GCA CGT GGC GGC AGG GGC AGC AAC GCC AAC GGT GGT GCC TCC
       Thr Ala Gly Thr Pro Gln Ala Arg Gly Gly Arg Gly Ser Asn Ala Asn Gly Gly Ala Ser

 181   GGG AGC AAT GCC GGT GGC GGA GAT GAG CCG CGA AAG GAA GCC AAA ACA ACG CCC AGG ATA
       Gly Ser Asn Ala Gly Gly Gly Asp Glu Pro Arg Lys Glu Ala Lys Thr Thr Pro Arg Ile

 241   TCC AGG GCG CTG GGC ACC TCG GCC AAG CGC ATA CAG AAG GAG CTG GCT GAG ATC ACA CTG
       Ser Arg Ala Leu Gly Thr Ser Ala Lys Arg Ile Gln Lys Glu Leu Ala Glu Ile Thr Leu

 301   GAC CCA CCG CCC AAC TGC AGT GCC GGG CCC AAG GGC GAC AAC CTG TAC GAG TGG GTA TCC
       Asp Pro Pro Pro Asn Cys Ser Ala Gly Pro Lys Gly Asp Asn Leu Tyr Glu Trp Val Ser

 361   ACC ATA CTG GGA CCA CCC GGA TCC GTC TAC GAG GGC GGC GTC TTC TTC CTC GAC ATA CAC
       Thr Ile Leu Gly Pro Pro Gly Ser Val Tyr Glu Gly Gly Val Phe Phe Leu Asp Ile His

 421   TTC TCG CCG GAG TAT CCC TTC AAG CCA CCC AAA GTC ACG TTC CGC ACG CGC ATC TAT CAC
       Phe Ser Pro Glu Tyr Pro Phe Lys Pro Pro Lys Val Thr Phe Arg Thr Arg Ile Tyr His

 481   TGC AAC ATC AAC AGC CAG GGC GTC ATT TGC CTG GAC ATA CTC AAG GAC AAC TGG TCG CCG
       Cys Asn Ile Asn Ser Gln Gly Val Ile Cys Leu Asp Ile Leu Lys Asp Asn Trp Ser Pro

 541   GCG CTG ACC ATA TCA AAG GTC TTG CTG TCA ATT TGC TCC CTG CTC ACA GAC TGT AAT CCA
       Ala Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp Cys Asn Pro

 601   GCC GAT CCG CTG GTG GGC AGC ATT GCC ACG CAA TAT TTG CAG AAT CGT GAG GAG CAC GAT
       Ala Asp Pro Leu Val Gly Ser Ile Ala Thr Gln Tyr Leu Gln Asn Arg Glu Glu His Asp

 661   CGA ATT GCG CGT CTC TGG ACA AAA AGG TAC GCA ACA TGATGCGTAAGTTTTGGCAACGTGCATGTAC
       Arg Ile Ala Arg Leu Trp Thr Lys Arg Tyr Ala Thr

 728   CACAAACCGAAAACTGCAACAACAGAAACGAAATTCCCATCTTGCAACGGAGAAAGTCTATTTTTMTAAATATATTTA
 807   AAGTACGATACTACTACTAAATTAACAGTAACAACTACCGTACACGAAAGCAACCGCCAACAGCCAGTCGATCCAATCC
 886   AATCTAAATTCAGTCTATATTTATATCACGTAATACAGAAGTTGT
```

UbcM2

```
-171                                                                    GGGCCCGGAGCCTC
-158   GGCGGCGGCTTCTTTTCCCGCCCCCCCGGTGCCCCCCCCAGGCTGCCTCCATTTCCTCAAGGAAGGTTCCTCGGCGCTC
 -79   TCCCTCCCCCACAGCGTAGCGGCGCACGAGCGGGCCGGGCGGGCGGCCGAGAAGTTTCCCAGAGATAAGTTCACCAAG

   1   ATG TCC AGT GAC AGG CAA AGG TCG GAT GAT GAG AGC CCT AGC ACC AGC AGT GGC AGT TCA
       Met Ser Ser Asp Arg Gln Arg Ser Asp Asp Glu Ser Pro Ser Thr Ser Ser Gly Ser Ser

  61   GAT GCA GAT CAG CGA GAC CCT GCC GCT CCA AAG CCC GAG GAA CAG GAA GAA AGA AAA CCT
       Asp Ala Asp Gln Arg Asp Pro Ala Ala Pro Lys Pro Glu Glu Gln Glu Glu Arg Lys Pro

 121   TCT GCC ACC CAG CAG AAG AAA AAC ACC AAA CTC TCT AGC AAA ACG ACT GCT AAG TTA TCC
       Ser Ala Thr Gln Gln Lys Lys Asn Thr Lys Leu Ser Ser Lys Thr Thr Ala Lys Leu Ser

 181   ACT AGT GCT AAA AGA ATT CAG AAG GAG CTA GCA GAG ATA ACC CTT GAT CCT CCT CCT AAC
       Thr Ser Ala Lys Arg Ile Gln Lys Glu Leu Ala Glu Ile Thr Leu Asp Pro Pro Pro Asn

 241   TGC AGT GCT GGG CCT AAA GGA GAT AAC ATT TAT GAA TGG AGG TCA ACT ATA CTT GGT CCA
       Cys Ser Ala Gly Pro Lys Gly Asp Asn Ile Tyr Glu Trp Arg Ser Thr Ile Leu Gly Pro

 301   CCA GGT TCT GTA TAT GAA GGT GGT GTT TTT TTT CTG GAT ATC ACA TTT TCT TCA GAT TAT
       Pro Gly Ser Val Tyr Glu Gly Gly Val Phe Phe Leu Asp Ile Thr Phe Ser Ser Asp Tyr

 361   CCA TTT AAG CCA CCA AAG GTT ACT TTC CGT ACC AGA ATC TAT CAC TGC AAC ATC AAC AGT
       Pro Phe Lys Pro Pro Lys Val Thr Phe Arg Thr Arg Ile Tyr His Cys Asn Ile Asn Ser

 421   CAG GGA GTC ATC TGC CTG GAT ATT CTG AAA GAC AAC TGG AGT CCT GCT TTG ACT ATT TCA
       Gln Gly Val Ile Cys Leu Asp Ile Leu Lys Asp Asn Trp Ser Pro Ala Leu Thr Ile Ser

 481   AAG GTT TTG CTC TCT ATT TGT TCC CTT TTG ACA GAT TGC AAC CCT GCG GAT CCT CTG GTC
       Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp Cys Asn Pro Ala Asp Pro Leu Val

 541   GGA AGC ATA GCC ACT CAG TAT TTG ACC AAC AGA GCA GAA CAT GAC AGG ATA GCC AGA CAG
       Gly Ser Ile Ala Thr Gln Tyr Leu Thr Asn Arg Ala Glu His Asp Arg Ile Ala Arg Gln

 601   TGG ACC AAG AGA TAC GCA ACA TAATTCACATCATTTGTATGCAGTGTGAAGGAGCAGAAGGCCTCTTCCCAC
       Trp Thr Lys Arg Tyr Ala Thr

 673   TGTGCTGCGGATCTTTATAGCCTTTACAATACGGACTTCTGTG
```

UbcM3

```
-153        GGGTAAAGTTGCTGCACTTCCGCTTCTCTCCCAGTGAGAGCGAGACACGAGTGGCCAGGCCCAGTGCCAGCCGC
 -79   AGCAGCAGCCGCCGGCGGCGGCACGGAGCAGCCAGACACAAAGAGAGGGGCCCGTTCGCCGGGCTGGGGTGGGGGTTCGCT

   1   ATG TCG GAT GAC GAT TCG AGG GCC AGC ACC AGC TCC TCG TCA TCT TCG TCC TCC AAC CAG
       Met Ser Asp Asp Asp Ser Arg Ala Ser Thr Ser Ser Ser Ser Ser Ser Ser Asn Gln

  61   CAG ACC GAG AAA GAA GGC AGC ACC CCC AAG AAG AAG GAG AGT AAA GTC AGC ATG AGC AAA
       Gln Thr Glu Lys Glu Gly Ser Thr Pro Lys Lys Lys Glu Ser Lys Val Ser Met Ser Lys

 121   AAT TCG AAG CTT CTC TCC ACC AGC GCT AAG AGG ATC CAG AAG GAG CTG GCA GAC ATC ACT
       Asn Ser Lys Leu Leu Ser Thr Ser Ala Lys Arg Ile Gln Lys Glu Leu Ala Asp Ile Thr

 181   TTA GAC CCT CCG CCA AAC TGC AGT GCT GGT CCC AAG GGT GAC AAC ATC TAT GAG TGG AGA
       Leu Asp Pro Pro Pro Asn Cys Ser Ala Gly Pro Lys Gly Asp Asn Ile Tyr Glu Trp Arg

 241   TCA ACC ATT CTG GGT CCT CCA GGG TCT GTG TAT GAA GGT GGA GTA TTC TTC CTT GAC ATT
       Ser Thr Ile Leu Gly Pro Pro Gly Ser Val Tyr Glu Gly Gly Val Phe Phe Leu Asp Ile

 301   ACT TTT ACA CCA GAG TAT CCT TTC AAG CCT CCA AAG GTT ACA TTT CGT ACA AGA ATT TAT
       Thr Phe Thr Pro Glu Tyr Pro Phe Lys Pro Pro Lys Val Thr Phe Arg Thr Arg Ile Tyr

 361   CAC TGT AAT ATT AAT AGC CAA GGA GTT ATT TGC TTG GAC ATA TTG AAA GAC AAC TGG AGC
       His Cys Asn Ile Asn Ser Gln Gly Val Ile Cys Leu Asp Ile Leu Lys Asp Asn Trp Ser

 421   CCA GCC CTA ACC ATC TCG AAA GTC CTC CTT TCT ATC TGC TCA CTC CTT ACA GAC TGC AAT
       Pro Ala Leu Thr Ile Ser Lys Val Leu Leu Ser Ile Cys Ser Leu Leu Thr Asp Cys Asn

 481   CCT GCT GAC CCT TTG GTG GGA AGT ATT GCC ACT CAG TAC ATG ACC AAC AGA GCA GAA CAC
       Pro Ala Asp Pro Leu Val Gly Ser Ile Ala Thr Gln Tyr Met Thr Asn Arg Ala Glu His

 541   GAC AGA ATG GCC AGA CAG TGG ACC AAG AGA TAC GCT ACA TAAACTGGGTTTTCAGTTCTTACATGA
       Asp Arg Met Ala Arg Gln Trp Thr Lys Arg Tyr Ala Thr

 607   TTTGTCTGTCACAGAGAGACCTGCTTATGATTTTGAAAGGGTGGGGGTGGCGTTGGTAAAGAGTAGGCTATTTCTAT
 686   ATCAGATATTATTCAGTCTTACTTCCCAGGATTTTGTTACTTAAGTATCTTGCTACAGTAGACAAACTGGTAATA
 765   GCGACTTTTAGAGCTGAAACTTTGCAAGATTAGCTGACTTTTAGTACAAAAGAAGCATCGTAGATACTTGTGTTCCA
 844   TTGATCAGCTTAAATTTGTATAGGTACATGTCATTTATGTAAAGTCCTATTTTCTTGTTCAAAAAAAATGGAATTTGTT
 923   TTTCCTTCTTAAAAGGAAACAATGGTATTTAACAAGTTTGCAAAGATGAGTTTGTTACCTCGTATATATGTAAAATGGACA
1002   CCTGGCTACAAGACACCATGTAGTTAATGTATTTTATCTTATATGCAAATTACTTCAAGGGTTGAATCACAATTTGTGG
1081   AAACCTGTAGTGAAATACCTTAAGCTGTTAACTGTAAGTGTGGAATAGGAATTGTTCAGTCAATTGGTTTAGACTTCTT
1160   ACGTCTGCATTTGTGC
```

FIG. 1

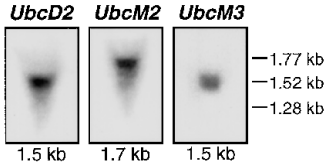

FIG. 2. **Northern blot analysis.** *UbcD2*, *UbcM2*, and *UbcM3* transcripts are shown with their respective sizes. The positions of RNA size markers are indicated on the *right*.

and the two mouse genes *UbcM2* and *UbcM3* (ubiquitin-conjugating enzymes from Drosophila or mice; the numbering reflects the order of identification). As reported previously (17), the UbcD1 enzyme is the structural and functional homolog of yeast UBC4. *UbcD3* identified by our screen is identical to *bendless*, a gene implicated in *Drosophila* nervous system development characterized while this work was in progress (26, 27). We mapped *UbcD1*, *UbcD2*, and *UbcD3* on the *Drosophila* genome to positions 88D, 32A/B, and 12D, respectively (data not shown).

A Family of Highly Related Ubiquitin-conjugating Enzymes with Distinct Amino-terminal Extensions—The DNA sequences and the deduced amino acid sequences of the three new genes, *UbcD2*, *UbcM2*, and *UbcM3* are shown in Fig. 1. The initiator methionines of the three genes were assigned to the first ATG of the cDNAs preceded by stop codons. The open reading frames of *UbcD2*, *UbcM2*, and *UbcM3* predicted proteins of 232 (24.5 kDa), 207 (22.9 kDa), and 193 (21.3 kDa) residues, respectively. By Northern analysis using total RNA isolated either from *Drosophila* or mouse, we identified with the respective probes transcripts for *UbcD2*, *UbcM2*, and *UbcM3* of ~1.5, ~1.7, and ~1.5 kb, respectively (Fig. 2). The sizes fit well to the corresponding cDNAs and indicate that the isolated cDNAs contain the complete open reading frames.

A comparison of the deduced amino acid sequences of UbcD2, UbcM2, and UbcM3 shows that they are highly related (Fig. 3). Unlike previously identified E2 enzymes (3, 6), these new enzymes possess amino-terminal extensions in addition to the UBC domain. The UBC domains of the three enzymes are almost identical in sequence (94% identity over 149 amino acid residues; between 72 and 79% at the DNA level). UbcD2 differs from the UbcD2/UbcM2/UbcM3 consensus by 6, UbcM2 by 2, and UbcM3 by 3 residues (Fig. 3). In contrast to the extreme conservation of the UBC domains, the amino-terminal extensions of the three enzymes (designated extensions *A*, *B*, and *C*; Fig. 3) show little sequence similarity among each other and differ in size (Fig. 3). The weak sequence similarity of the extensions is largely restricted to clusters of serine/threonine and basic residues. No significant sequence similarities between the extensions and known sequences in the data bases were found except for short consensus sequences for phosphorylation sites. Further data base searches detected (in addition to a human UbcM2 homolog designated UbcH9) a partial open reading frame from a human cDNA fragment (designated *UbcH8*), and this represents a probable fourth member of this enzyme family. This partial sequence exhibits a 100% match to the corresponding sequences of the UBC domains of UbcD2, UbcM2, and UbcM3 and an amino-terminal extension (designated extension *D*) nonidentical, but related to, extension B of

FIG. 1. **Nucleotide and predicted amino acid sequences of *UbcD2*, *UbcM2*, and *UbcM3* cDNAs.** Nucleotide numbers starting at the first nucleotide of the coding region are given on the *left*. In-frame stop codons in the 5′-untranslated region are *underlined*. Active site cysteine residues required for thioester formation are shown in *boldface*. The nucleotide sequences have been submitted to the Genbank/EMBL data base with the accession numbers X92663 *UbcD2*, X92664 *UbcM2* and X92665 *UbcM3*.

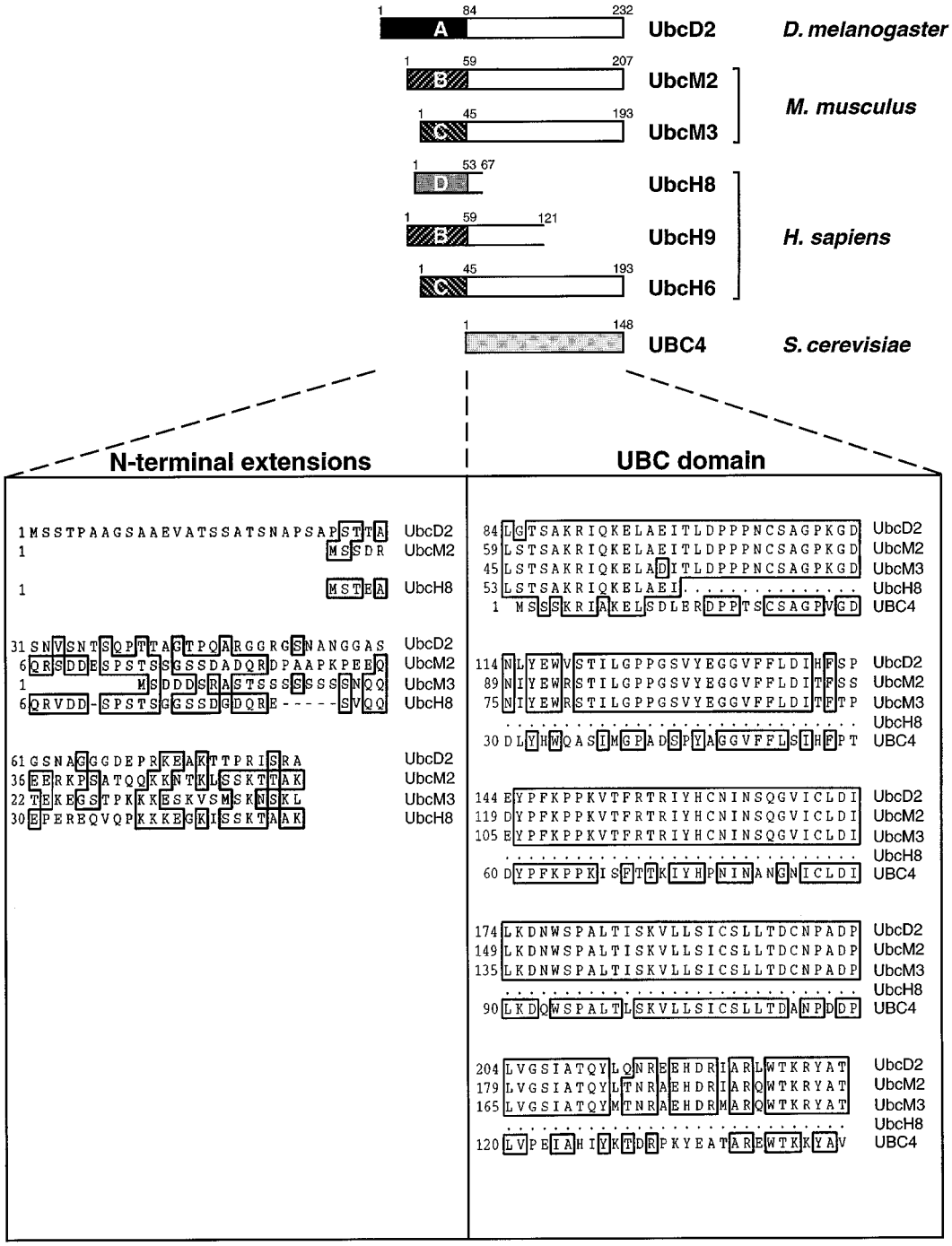

FIG. 3. **Sequence similarity between members of the novel UBC-family.** Primary sequences of UbcD2, UbcM2, UbcM3, and three additional members of this family designated UbcH8, UbcH9 (translated from partial cDNA sequences, accession numbers Z44894 and H12272) and UbcH6[2] are compared with yeast UBC4 (14). *Upper panel*, schematic diagram of the primary sequences. The highly conserved UBC domain of the new UBC family is shown as a *white box* and corresponds to the entire sequence of UBC4 (*light gray*). The four different amino-terminal extensions are shown as boxes in distinct *gray shades* and are designated *A*, *B*, *C*, and *D* extensions. Numbers above the boxes correspond to amino acid residues. *Lower left panel,* sequence comparison of the amino-terminal extensions of UbcD1, UbcM2, UbcM3, and UbcH8. Residues that are identical in at least two proteins are *boxed*. Gaps (indicated by *dashes*) were permitted to optimize alignments. Amino-terminal extensions of the homologs UbcM2 and UbcH9 (extension B) differ by only one residue (K31E) and of the homologs UbcM3 and UbcH6 (extension C) by two residues (G26T/S27N). *Lower right panel*, sequence comparison of the UBC domains of different family members with UBC4. Unavailable sequence data of the partial *UbcH8* clone is indicated by a *dotted line*. Residues that are identical in at least three proteins are *boxed*. The sequences were aligned using the BoxAlign program (GCG package). Residue numbers are given on the *left*.

UbcM2 and UbcH9 (Fig. 3). This suggests that the novel E2 enzyme family described here has at least four distinct members.

UbcD2, UbcM2, and UbcM3 Are Structurally and Functionally Related to Yeast UBC4—Computer aided sequence comparisons of the UBC domains of UbcD2, UbcM2, and UbcM3 with the UBC domains of other known ubiquitin-conjugating enzymes revealed that their closest homologs are yeast UBC4/UBC5 (an enzyme pair expressed from duplicated genes; Refs. 13 and 28) and their respective homologs from higher eukaryotes, including the *Drosophila UbcD1* (17), *C. elegans ubc-2* (18), and human *UbcH5* (19) gene products. The UBC domains of UbcD2, UbcM2 and UbcM3 share 64% sequence identity with these UBC4-like enzymes, suggesting that the

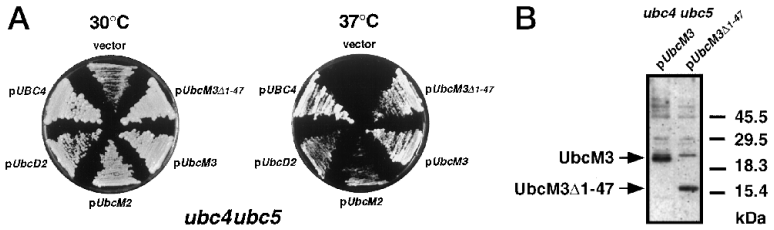

FIG. 4. **Complementation of the yeast *ubc4 ubc5* mutant by expression of *UbcD2*, *UbcM2*, *UbcM3* and an amino-terminal truncated *UbcM3*.** *A*, growth of yeast *ubc4 ubc5* double mutant on YPGal plates at normal growth temperature (30 °C) and heat shock temperature (37 °C) expressing the following UBCs (in clockwise orientation). Vector as negative control, amino-terminal truncated UbcM3 (p*UbcM3Δ1–47*), UbcM3 (p*UbcM3*), UbcM2 (p*UbcM2*), UbcD2 (p*UbcD2*), and yeast UBC4 (p*UBC4*) as positive control. *B*, Western blot analysis of total yeast proteins from *ubc4 ubc5* cells expressing *UbcM3* and *UbcM3Δ1–47*, respectively, with antibodies generated against the conserved UBC domain of UbcD2/UbcM2/UbcM3 family. Additional protein bands cross-reacting with the antiserum serve as loading control. Size references are given on the right.

newly identified enzymes may be functionally related to UBC4 (Fig. 3). Yeast UBC4/UBC5 ubiquitin-conjugating enzymes are involved in stress-related functions and in the turnover of regulatory proteins (Ref. 13; see the Introduction). Single mutations in these genes are viable and lead to only moderate mutant phenotypes, but *ubc4 ubc5* double mutants are slowly growing and inviable at elevated temperatures (13). As reported previously, *UbcD1*, the *UBC4* homolog from *Drosophila* (80% identical to yeast UBC4), can rescue the deficiencies of yeast *ubc4 ubc5* mutants (17). Complementation of UBC4/UBC5 functions by *UbcD1* was nearly complete, even when the enzyme was expressed from a single copy in the genome (17).

To study the activity of *UbcD2*, *UbcM2* and *UbcM3* in yeast, we cloned the respective reading frames into yeast high copy number, 2 micron based expression vectors. When yeast cells were transformed with these plasmids, all three genes could complement the growth deficiency and heat sensitivity of *ubc4 ubc5* double mutants (Fig. 4*A*). Although complementation was only partial as indicated by slight growth defects at 30- and, in particular, at 37 °C, UbcD2, UbcM2, and UbcM3 are likely to function in similar proteolysis pathways as UBC4 (or UbcD1). The incomplete complementation of *ubc4 ubc5* by these genes may indicate that UbcD2, UbcM2, and UbcM3 only interact with a subset of UBC4's substrates or that they may fail to collaborate with certain components of a UBC4-dependent degradation pathway (*e.g.* ubiquitin ligases, E3), or both.

In addition to a comparably weak sequence similarity to UBC4/UBC5 (identity of 64 *versus* 80% for UbcD1, Fig. 5) the three enzymes differ from UBC4 and its homologs by the presence of amino-terminal extensions. These extensions could possibly function as regulatory (either activating or repressing) or interacting domains with specific components of the ubiquitin-conjugation system. To test these possibilities, we constructed a derivative of UbcM3 lacking the amino-terminal extension (UbcM3Δ1–47). When yeast cells were transformed with *UbcM3* and truncated *UbcM3Δ1–47*, both gene products were expressed with the expected sizes to similar levels (Fig. 4*B*). Remarkably, full-length and truncated UbcM3 enzymes complemented the yeast *ubc4 ubc5* mutant to a similar extend (Fig. 4*A*). Thus the amino-terminal extension of UbcM3 (and probably also of UbcD2 and UbcM2) does not modulate the activity of the enzyme to function in UBC4-dependent pathways in yeast.

DISCUSSION

Previously identified ubiquitin-conjugating enzymes (3, 6) are small proteins (~14–32 kDa), which either consist of the UBC domain only (class I E2 enzymes) or they possess additional carboxyl-terminal extensions (class II enzymes). Here we describe a novel family of ubiquitin-conjugating enzymes from higher eukaryotes that have amino-terminal but lack carboxyl-terminal extensions (designated class III enzymes). We have cloned three members of this class and identified a fourth in the

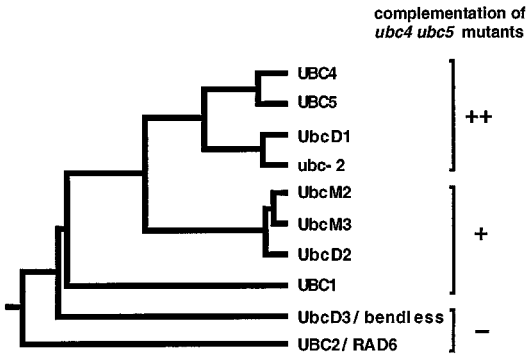

FIG. 5. **Phylogenetic tree of the UBC4-related subfamily.** Relatedness was calculated by the algorithm provided by the DNA Star package and compared with distantly related yeast UBC2/RAD6. Only the UBC domains were compared. Complementation of yeast *ubc4 ubc5* mutants (growth at 30 and 37 °C) is indicated; ++, full complementation; +, partial complementation by overexpression; –, no complementation (see text for experimental details).

data base. The UBC domain of these novel enzymes is virtually identical, but the amino-terminal extensions show limited sequence similarity. The recent identification of human homologs to UbcM2 (UbcH9; Fig. 3) and UbcM3 (UbcH6; Fig. 3),[2] which are homologous to their respective murine counterparts over their entire lengths (including the extensions), and the extreme conservation of the UBC domains of different members of this family strongly suggest that a homolog of each of these novel four UBCs may be present in each of these species, *i.e. Drosophila*, mouse, and man. We have unsuccessfully tried to identify yeast homologs to these enzymes using different PCR strategies. Thus we assume this family probably evolved relatively late in evolution and may be unique to multicellular organisms. Intriguingly, these enzymes are among to the most highly conserved proteins of these organisms. The UBC domains of UbcD2 from *Drosophila* and UbcM2, and UbcM3 from mice share 94% identical amino acid residues (homologs of other UBCs are typically 70–80% identical in sequence; Refs. 17–19, 29–32; see Fig. 5). The extreme conservation of UbcD2, UbcM2, and UbcM3 is even more remarkable given the likely possibility that the true homologs, *i.e.* the enzymes with similar extensions are yet to be identified. Proteins of similar high conservation, *e.g.* histones or ubiquitin, either have multiple interacting partners or most of their amino acid residues participate in intramolecular contacts. Both types of interactions are thought to prevent evolutionary amino acid sequence drift. We thus assume that the novel UBC enzymes interact with several proteins. Candidates for binding partners are components of the ubiquitin/proteasome system or substrates. Since overexpressed *UbcD2*, *UbcM2*, and *UbcM3* can partially sup-

[2] M. Scheffner, personal communication.

press *UBC4/UBC5* deficiency in yeast, these enzymes and UBC4/UBC5 probably have many substrates in common. However, the presence of multiple, highly conserved extensions of these enzymes suggests that they are likely to carry out specialized functions distinct from those of UBC4. What these functions are is not known at present, but the gene expression pattern of the *Drosophila* UBCs may provide some clues. Interestingly, *UbcD1*, the *UBC4* homolog, is continuously expressed throughout development consistent with a "housekeeping" function of the encoded enzyme.[3] Transcripts of *UbcD3(bendless)*, another *UBC4*-related gene (which is actually unable to rescue *ubc4 ubc5* mutants; Fig. 5)[3] can also be detected at all developmental stages of *Drosophila* development. In contrast, *UbcD2* appears to be exclusively expressed at postlarval (L3) stages, but in eggs the transcript is supplied maternally.[3] Thus the functions of these class III enzymes may be predominantly restricted to distinct tissues in pupae or adult flies.

The significance of the amino-terminal extensions is currently unclear, but their conservation between species (*e.g.* the extensions B and C; Fig. 3) indicate that they are probably relevant to their cellular functions. The carboxyl-terminal extensions of class II E2 enzymes are known either to contribute to their substrate specificity (UBC2, UBC3; Refs. 9 and 33–35) or they mediate intracellular localization (UBC6; Ref. 12). The prevalence of putative phosphorylation sites within the extensions of the UbcD2, UbcM2, and UbcM3 enzymes may indicate that the enzymatic activity or a possible interaction with other proteins is possibly controlled by enzyme phosphorylation. Alternatively, these sequences rich in serine, threonine, and basic residues may represent binding sites for specific components of the ubiquitin-conjugating system or proteolytic substrates. Class I ubiquitin-conjugating enzymes have highly conserved three dimensional structures with exposed amino termini (36). This suggests that the highly charged amino-terminal extensions of the class III enzymes described here may fold into separate domains, which are probably readily accessible to interacting partners.

Acknowledgments—We thank Petra Hubbe for technical assistance, Dr. Martin Scheffner for sharing results prior to publication, and Dr. Susan Smith for comments on the manuscript.

[3] K. Matuschewski, H.-P. Hauser, M. Treier, and S. Jentsch, unpublished results.

REFERENCES

1. Finley, D., and Chau, V. (1991) *Annu. Rev. Cell Biol.* **7,** 25–69
2. Hershko, A., and Ciechanover, A. (1992) *Annu. Rev. Biochem.* **61,** 761–8073
3. Jentsch, S. (1992) *Annu. Rev. Genet.* **26,** 179–207
4. Hochstrasser, M. (1995) *Curr. Opin. Cell Biol.* **7,** 215–223
5. Jentsch, S., and Schlenker, S. (1995) *Cell* **82,** 881–884
6. Jentsch, S. (1992) *Trends Cell Biol.* **2,** 98–103
7. Scheffner, M., Nuber, U., and Huibregtse, J. M. (1995) *Nature* **373,** 81–83
8. Jentsch, S., McGrath, J. P., and Varshavsky, A. (1987) *Nature* **329,** 131–134
9. Goebl, M. G., Yochem, J., Jentsch, S., McGrath, J. P., Varshavsky, A., and Byers, B. (1988) *Science* **241,** 1331–1335
10. Seufert, W., Futcher, B., and Jentsch, S. (1995) *Nature* **373,** 78–81
11. Wiebel, F., and Kunau, W.-H. (1992) *Nature* **359,** 73–76
12. Sommer, T., and Jentsch, S. (1993) *Nature* **365,** 176–179
13. Seufert, W., and Jentsch, S. (1990) *EMBO J.* **9,** 543–550
14. Seufert, W., McGrath, J. P., and Jentsch, S. (1990) *EMBO J.* **9,** 4535–4541
15. Jungmann, J., Reins, H.-A., Schobert, C., and Jentsch, S. (1993) *Nature* **361,** 369–371
16. Chen, P., Johnson, P., Sommer, T., Jentsch, S., and Hochstrasser, M. (1993) *Cell* **74,** 357–369
17. Treier, M., Seufert, W., and Jentsch, S. (1992) *EMBO J.* **11,** 367–372
18. Zhen, M., Heinlein, R., Jones, D., Jentsch, S., and Candido, E. P. M. (1993) *Mol. Cell. Biol.* **13,** 1371–1377
19. Scheffner, M., Huibregtse, J. M., and Howley, P. M. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 8797–8801
20. King, R. W., Peters, J.-M., Tugendreich, S., Rolfe, M., Hieter, P., and Kirschner, M. W. (1995) *Cell* **81,** 279–288
21. Ausubel, F. J., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1992) *Current Protocols in Molecular Biology,* Greene Publishing Associates and Wiley-Interscience, NY
22. Brown, N. H., and Kafatos, F. C. (1988) *J. Mol. Biol.* **203,** 425–437
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
24. Ashburner, M. (1989) *Drosophila: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25. Gietz, R. D., and Sugino, A. (1988) *Gene* (*Amst.*) **74,** 527–534
26. Muralidhar, M. G., and Thomas, J. B. (1993) *Neuron* **11,** 253–266
27. Oh, C. E., McMahon, R., Benzer, S., and Tanouye, M. A. (1994) *J. Neurosci.* **14,** 3166–3179
28. Seufert, W., and Jentsch, S. (1990) *Nucleic Acids Res.* **18,** 1638
29. Kaiser, P., Seufert, W., Höfferer, L., Kofler, B., Sachsenmaier, C., Herzog, H., Jentsch, S., Schweiger, M., and Schneider, R. (1994) *J. Biol. Chem.* **269,** 8797–8802
30. Koken, M., Reynolds, P., Bootsma, D., Hoijmakers, J., Prakash, S., and Prakash, L. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 3832–3836
31. Sullivan, M. L., and Vierstra, R. D. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 9861–9865
32. Van Nocker, S., and Vierstra, R. D. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88,** 10297–10301
33. Sung, P., Prakash, S., and Prakash, L. (1988) *Genes & Dev.* **2,** 1476–1485
34. Kolman, C. J., Toth, J., and Gonda, D. K. (1992) *EMBO J.* **11,** 3081–3090
35. Silver, E. T., Gwozd, T. J., Ptak, C., Goebl, M., and Ellison, M. J. (1992) *EMBO J.* **11,** 3091–3098
36. Cook, W. J., Jeffrey, L. C., Xu, Y., and Chau, V. (1993) *Biochemistry* **32,** 13809–13817