Mol Gen Genet (1997) 255:302–310

ORIGINAL PAPER

J. S. Chandler · B. McArdle · J. Callis

*At*UBP3 and *At*UBP4 are two closely related *Arabidopsis thaliana* ubiquitin-specific proteases present in the nucleus

Received: 14 November 1996 / Accepted: 6 February 1997

Abstract The ubiquitin-specific proteases (UBPs) are a class of enzymes vital to the ubiquitin pathway. These enzymes cleave ubiquitin at its C-terminus from two types of substrates containing (i) ubiquitin in an α-amino linkage, as found in the primary ubiquitin translation products, polyubiquitin and ubiquitin-ribosomal fusion proteins, or (ii) ubiquitin in an ε-amino linkage, as found in multiubiquitin chains either unattached or conjugated to cellular proteins. We have isolated cDNAs for two *Arabidopsis thaliana* genes, *AtUBP3* and *AtUBP4*, which encode UBPs that are 93% identical. These two cDNAs represent the only two members of this subgroup and encode the smallest UBPs described to date in any organism. Using in vivo assays in *Escherichia coli* that allow the coexpression of a UBP with a putative substrate, we have shown that *At*UBP3 and *At*UBP4 can specifically deubiquitinate the artificial substrate Ub-X-β-gal but cannot act upon the natural α-amino-linked ubiquitin fusions *Arabidopsis* Ub-CEP52 and *Arabidopsis* polyubiquitin. Affinity-purified antibody prepared against *At*UBP3 expressed in *E. coli* recognizes both *At*UBP3 and *At*UBP4. *At*UBP3 and/or *At*UBP4 are present in all *Arabidopsis* organs examined and at multiple developmental stages. Subcellular localization studies show that *At*UBP3 and/or *At*UBP4 are present in nuclear extracts. Possible physiological roles for these UBPs are discussed.

Key words *Arabidopsis* · Ubiquitin · Ubiquitin-specific protease

Communicated by H. Saedler

J. S. Chandler · B. McArdle · J. Callis
Section of Molecular and Cellular Biology,
University of California-Davis, Davis, CA 95616, USA
Tel. +1-916-752-1015; Fax +1-916-752-3085

Introduction

Ubiquitin, an abundant 76 amino acid polypeptide, is synthesized both as a fusion protein with one of two ribosomal proteins and as a polyprotein comprising an uninterrupted head-to-tail tandem array of ubiquitin units. The translation products of all ubiquitin genes are enzymatically cleaved to produce functional ubiquitin monomers, which are then covalently attached to cellular proteins via the C-terminus of ubiquitin and the ε-amino groups of lysine residues of the substrate proteins. This linkage is referred to as an ε-amino linkage or an isopeptide bond. Ubiquitination is catalyzed by three enzymes; ubiquitin-activating enzyme (E1), one of a family of ubiquitin-conjugating enzymes (E2s), and possibly one of a family of recognition proteins (E3s) (reviewed in Jentsch 1992). Ubiquitin can also become attached to one of several lysine residues on another ubiquitin to create multiubiquitin chains (Chau et al. 1989). Multiubiquitination has been hypothesized to be a necessary, but not sufficient, tag that targets most proteins for degradation by the 26S protease complex (Chau et al. 1989). In several cell types, including erythrocytes, reticulocytes, Ehrlich ascites tumor cells and HeLa cells, there are pools of both labile and stable multiubiquitinated proteins (Hershko et al. 1982; Hass and Bright 1985; Carlson et al. 1987). Why some multiubiquitinated proteins are rapidly degraded while others are stable remains unknown.

Deubiquitinating enzymes function in the ubiquitin pathway in several places (reviewed in Rose 1988; Hershko and Ciechanover 1992). Deubiquitinating enzymes catalyze the hydrolysis of two types of peptide bonds: the peptide bond between the C-terminus of ubiquitin and the α-amino group of another protein, as in the processing of primary ubiquitin translation products to monomers, or the isopeptide bond between the C-terminus of ubiquitin and the lysyl ε-amino group of another protein or peptide. A variety of isopeptide linkages are possible. The C-terminus of free ubiquitin

can become modified by cellular nucleophiles, so deubiquitinating enzymes are needed to remove these modifications. Ubiquitin monomers are recycled following degradation by the 26S protease complex; deubiquitinating enzymes disassemble the multiubiquitin chain and free the final ubiquitin monomer from the remaining residues of the degraded protein. Some deubiquitinating enzymes appear to perform an editing function to control the fidelity of the conjugation process, thus preventing inappropriate degradation of cellular proteins.

Such functional diversity suggests an equivalent diversity of the enzymes responsible. Two families of deubiquitinating enzymes have been identified on the basis of in vitro activities and/or sequence identity. The ubiquitin carboxyl-terminal hydrolases (UCHs) are small (approximately 25 kDa) thiol proteases that share amino acid sequence identity and cleave esters and amides from the C-terminus of ubiquitin (reviewed in Wilkinson 1994). Unrelated in sequence to the UCHs are the ubiquitin-specific proteases (UBPs), a large family of proteins differing greatly in size but characterized by sequence similarity in several regions: the Cys box, the His box, and two other blocks of amino acid sequence identity (Wilkinson et al. 1995).

For several of the UBPs, in vivo roles are beginning to be elucidated. Two UBPs appear to be involved in the regeneration of ubiquitin monomers. The yeast *DOA4* gene, essential for normal degradation of the MATα2 repressor, encodes a UBP. *doa4* mutants accumulate ubiquitin with small peptide fragments attached, presumably owing to an inability to free the final ubiquitin monomer following degradation by the 26S protease complex (Papa and Hochstrasser 1993). Isopeptidase T is a UBP that is able to disassemble polyubiquitin chains (Chen and Pickart 1990). Isopeptidase T requires ubiquitin with a free C-terminus so it appears that this enzyme is responsible for disassembling free multiubiquitin chains (Wilkinson et al. 1995).

Other UBPs might monitor the fidelity of the conjugation process to prevent untimely proteolysis. The *fat facets* gene, involved in the development of *Drosophila* eyes and embryos (Fischer-Vize et al. 1992), encodes a UBP (Huang et al. 1995). Because the mutant phenotype is suppressed in a genetic background that also has reduced proteasome activity, the Fat facets protein is hypothesized to act by deubiquitinating a regulatory protein, preventing its degradation (Huang et al. 1995).

Although no direct link to an inhibition of proteolysis has been established for other UBPs, some might act in a similar fashion. Alternately, some UBPs might change the ubiquitination state of a substrate protein, thus altering its activity. *Sc*UBP3 interacts with the SIR4 protein, which is required for transcriptional silencing, and *ScUBP3* deletion strains show an increase in silencing. This suggests that *Sc*UBP3 normally deubiquitinates and either stabilizes a protein necessary for the inhibition of silencing or renders the silencing complex inactive (Moazed and Johnson 1996). When overexpressed, DUB-1, a mammalian UBP, results in cell cycle arrest. This suggests that DUB-1 acts upon some type of growth factor, either inhibiting its proteolysis or influencing its activity (Zhu et al. 1996).

To begin to elucidate the roles that UBPs play in regulating the degradation or activity of proteins in higher plants, we have isolated cDNAs that encode two *Arabidopsis thaliana* UBPs, *At*UBP3 and *At*UBP4. These closely related UBPs represent a subgroup of the UBP family and are the smallest UBPs characterized to date. We have demonstrated their expression in multiple *Arabidopsis* organs and their presence in nuclear fractions by Western blotting.

Materials and methods

Strains of *Escherichia coli*

The strains of *E. coli* used were: XL-1 Blue for plasmid rescue from the λZAPII (Stratagene) cDNA expression library; BNN102 for the λGT10 library screening; DH5α and JM101 for DNA manipulation; JM101 for single-stranded DNA preparation; BW313 for single-stranded DNA preparation for mutagenesis; DH5α for expression of MBPubpΔ138-371 and GST*At*UBP4 (see below); MC4100 for in vivo UBP activity assays against Ub-R-β-gal or Ub-M-β-gal; and BL21 for in vivo UBP activity assays against Ub-CEP52 and polyubiquitin. Standard techniques were used for all work with *E. coli* (Sambrook et al. 1989).

Isolation of *AtUBP3* cDNA

Plasmids were rescued from a cDNA expression library in λZAPII of immature flowers of *Arabidopsis* Landsberg *erecta* (Weigel et al. 1992), as described by Stratagene. From 5×10^6 PFU (plaque-forming units), 30 000 colonies containing the rescued cDNA library in pBluescript SK-(Stratagene) were obtained. Cells were scraped from plates and plasmid DNA was isolated using standard methods (Sambrook et al. 1989).

E. coli containing the plasmids pACUb-R-β-gal or pACUb-M-β-gal (Baker et al. 1992) was transformed with the rescued plasmids (described above), plated on Luria Broth containing ampicillin (100 μg/ml), chloramphenicol (34 μg/ml), isopropyl β-D-thiogalactopyranoside (32 μg/ml), and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) (32 μg/ml), and incubated overnight at 37° C. Plates were stored at 4° C for 24–48 h to allow full development of blue color. The plasmid p*At*UBP3 (p8015) was isolated in this screen.

Isolation of *AtUBP4* cDNA and DNA sequencing

1.8×10^5 PFU of a cDNA library in λGT10 of immature flowers of *Arabidopsis* Landsberg *erecta* (Yanofsky et al. 1990) were screened. [α^{32}P]dCTP-labeled DNA probes were made from a 1.5 kb *Eco*RI-*Xho*I fragment of p8015 (Feinberg and Vogelstein 1983). Plaque lifts were made by standard protocols (Sambrook et al. 1989) onto Biotrace RP membrane (Gelman Sciences). Hybridization was at 65° C in $3 \times$ SSPE, $2 \times$ Denhardt's solution, 0.1% SDS, and 100 μg/ml salmon sperm DNA. Washing was done at 65° C in $3 \times$ SSPE and 0.1% SDS. Phage DNA was isolated (Sambrook et al. 1989) and a 1.35 kb *Eco*RI fragment containing the cDNA ligated into the *Eco*RI site of pBluescript KS-(Stratagene) to create p*At*UBP4 (p8041).

For all DNA sequence analysis, single-stranded DNA was produced by standard methods (Sambrook et al. 1989), and

overlapping sequence in both directions of p8015 and p8041 was determined by the dideoxynucleotide chain-termination method (Sanger et al. 1977).

Southern blot hybridization of genomic DNA

Total cellular DNA of *A. thaliana* ecotype Landsberg *erecta* was isolated as previously described (Dellaporta et al. 1983). Digested DNA (5 μg) was fractionated by electrophoresis in a 0.8% agarose gel and transferred to Hybond-N membrane (Amersham) using standard protocols (Sambrook et al. 1989). The [α-^{32}P]dCTP-labeled DNA probe used was the 1.5 kb *Eco*RI-*Xho*I fragment of p*At*UBP3 (p8015). Labeling, hybridization and washing were as described above.

Construction of the *E. coli* expression vectors pMBP*At*UBP and pGST*At*UBP4

Mutagenesis by the polymerase chain reaction (PCR) of the cDNAs p*At*UBP3 (p8015) and p*At*UBP4 (p8041) was done to introduce a *Bam*HI site (underlined below) to allow *At*UBP3 and *At*UBP4 to be encoded as fusions with the C-terminus of the maltose-binding protein (MBP) encoded by pMAL-c2 (New England Biolabs). The PCRs were done using 50 mM KCl, 10 mM TRIS-HCl, pH 9.0, 0.1% Triton X-100, 2.5 mM $MgCl_2$, 0.05 mM dNTP, approximately 25 ng plasmid DNA, 20 pmol each of the universal (−20) primer and the oligonucleotide 5′CGGGATCCATGGGCGCCGCGGG3′, and 2.5 units *Taq* DNA polymerase (Promega).

A 715 bp *Bam*HI-*Pst*I fragment of the p*At*UBP3 PCR product was ligated into the *Bam*HI/*Pst*I sites of pBluescript KS+ to create p8087. p8087 was sequenced from the newly created *Bam*HI site across 352 bp through an internal *Eco*RI site. This *Bam*HI-*Eco*RI fragment was ligated into the *Bam*HI/*Eco*RI sites of p*At*UBP3 (p8015) to create p8136. To place a *Hind*III site 3′ of the coding region, the 1.4 kb *Xba*I-*Kpn*I fragment of p8136 was ligated into the *Xba*I/*Kpn*I sites of pGEM-7Zf(+) (Promega) to create p8137. The 1.4 kb *Bam*HI-*Hind*III fragment of p8137 was ligated into the *Bam*HI/*Hind*III sites of pMAL-c2 to create pMBP*At*UBP3 (p8139).

A 700 bp *Bam*HI-*Pst*I fragment of the p*At*UBP4 PCR product was ligated into the *Bam*HI/*Pst*I sites of pBluescript KS+ to create p8088. p8088 was sequenced from the newly created *Bam*HI site across 379 bp through an internal *Bgl*II site. The 379 bp *Bam*HI-*Bgl*II fragment of p8088 was ligated into the *Bam*HI/*Bgl*II sites of p*At*UBP4 (p8040) to create p8099. p8040 is the same as p8041 except that the cDNA is in the opposite orientation. Using a *Hind*III site in the 3′ untranslated region of the cDNA, the 1.1 kb *Bam*HI-*Hind*III fragment of p8099 was ligated into the *Bam*HI/*Hind*III sites of pMAL-c2 to create pMBP*At*UBP4 (p8110).

pMBP*At*ubp3^{C32S} and pMBP*At*ubp4^{C32S} were created using the Kunkel (1985) method of oligonucleotide-mediated mutagenesis, using standard protocols (Sambrook et al. 1989). The degenerate oligonucleotide 5′CGGCAACACTAG(T/C)TACTGTAACAG3′ was used to mutagenize p8136 and p8088 (described above) in order to change amino acid 32 from cysteine to serine. The underlined base is the change introduced and the bases in parentheses are degenerate so as to facilitate use for both p8136 and p8088. The success of mutagenesis was verified by sequencing of the products and these were used to replace the corresponding regions in the pMAL expression vectors, creating pMBP *At*ubp3^{C32S} (p8161) and pMBP*At*ubp4^{C32S} (p8162).

pMBP*At*UBP3Δ138-371 was created to facilitate antigen production because UBP3 has the same molecular weight as the MBP, which makes gel purification of antigen impossible. The 710 bp *Bam*HI-*Pst*I fragment of p8136 was ligated into the *Bam*HI/*Pst*I sites of pMAL-c2 to create pMBPUBP3Δ138-371 (p8186). pGST*At*UBP4 was created so that GST*At*UBP4 could be used for the affinity purification of the anti-UBP3/4 antibodies. The 1.3 kb *Bam*HI-*Eco*RI fragment of p8099 was ligated into the *Bam*HI/*Eco*RI sites of pGEX-2T to create pGST*At*UBP4 (p8100).

Construction of *E. coli* expression vector pUB-CEP52

The 675 bp *Bgl*II-*Bam*HI fragment of p2301 (Sullivan et al. 1990), containing the T7 promoter of pET3c and the *Arabidopsis UBQ1* coding region (Callis et al. 1990), was ligated into the *Bam*HI site of pACYC184 to create pUB-CEP52 (p8185).

In vivo assays in *E. coli* for UBP activity against Ub-X-β-galactosidase (UB-X-β-gal) and UB-CEP52

For all coexpression experiments, 125 ml flasks containing 30 ml Luria Broth with ampicillin (100 μg/ml) and chloramphenicol (34 μg/ml) were inoculated with 500 μl of overnight culture and incubated at 30° C with shaking at 2000 rpm. Equal number of cells, as determined by optical density at 600 nm, were collected after 16 h for Ub-X-β-gal activity assays and after 3.5 h for UB-CEP52 activity assays, resuspended in Laemmli (1970) sample buffer, sonicated at 4° C and cleared by centrifugation. Protein extracts from volumes containing the equivalent of 5×10^8 cells (Ub-X-β-gal) and 2.5×10^8 cells (UB-CEP52) were used for SDS-polyacrylamide gel electrophoresis (SDS-PAGE).

Production and affinity purification of anti-*At*UBP3/4 polyclonal antibodies

Anti-UBP3/4 antibodies were raised in rabbits by immunization with a 137 amino acid N-terminal portion of the UBP3 protein initially expressed as a fusion with MBP (vector described above). Purification and cleavage were according to the manufacturer's instructions (New England Biolabs). The UBP3Δ138-371 antigen was purified from MBP by SDS-PAGE. GST*At*UBP4 was purified on glutathione-agarose. One milligram of GST*At*UBP4 was coupled to Affi-Gel 10 according to the manufacturer's instructions (Bio-Rad) with 90% efficiency. Antibody purification was carried out as described by Harlow and Lane (1988).

Isolation of protein and immunoblotting

The Landsberg *erecta* ecotype of *A. thaliana* was grown under conditions previously described (Beers et al. 1992). All organs were isolated from 6 week old plants except for yellow siliques, which were isolated from 8 week old plants. Tissue was ground in liquid nitrogen and proteins extracted from the frozen tissue into 50 mM TRIS-pH 8.0, 5 mM EDTA. Lysates were centrifuged for 10 min at 24 000 *g* at 4° C. Nuclear extracts were prepared as previously described (Beers et al. 1992). Contamination of the nuclear extract by cytosolic proteins was assessed by marker enzyme assays (Beers et al. 1992). The nuclear extract was assayed for the presence of a cytosolic enzyme. An aliquot of a cytosolic fraction containing the same level of activity of this cytosolic marker enzyme as found in 20 μg of nuclear extract was analyzed simultaneously for the presence of *At*UBP3 and/or *At*UBP4. Immunoblotting of SDS-PAGE-fractionated protein was carried out as previously described (Beers et al. 1992). Immunoblotting with anti-β-gal antibody (Cappel) was done using primary antibody at a 1:650 dilution for 2.5 h. For detection of *At*UBP3 and *At*UBP4 expressed in *E. coli*, the anti-*At*UBP3/4 antibody from crude serum was used at a 1:800 dilution for 2 h. For detection of *At*UBP3 and *At*UBP4 in *Arabidopsis*, the affinity-purified anti-*At*UBP3/4 antibody was used at a 1:32 dilution for 16 h, the secondary antibody used was goat anti-rabbit linked to horseradish peroxidase, and visualization was by chemiluminescence (Gallagher et al. 1994).

Results

Identification of *AtUBP3* and *AtUBP4*

To identify genes encoding UBPs from *Arabidopsis*, we used an in vivo *E. coli* screen developed previously (Baker et al. 1992). *E. coli* expressing the protein fusion Ub-R-β-gal were transformed with plasmids rescued from a cDNA λZAPII expression library made from mRNA of immature flower buds (Weigel et al. 1992). If a cDNA encodes an active UBP, ubiquitin is cleaved from the fusion protein leaving R-β-gal and, as previously described (Tobias et al. 1991), R-β-gal is rapidly degraded, resulting in a white colony on X-gal-containing plates. Conversely, if a cDNA does not encode an active UBP the Ub-R-β-gal remains intact and, because this form of β-gal is active and stable, a blue colony is produced on X-gal-containing plates.

From such a screen, cDNA-containing plasmids were isolated from white or pale blue colonies and transformed into a different strain expressing Ub-M-β-gal, and screened as described above. Nonspecific interference with β-gal activity resulted in white colonies with both Ub-R-β-gal and Ub-M-β-gal. cDNAs encoding UBPs produced blue colonies in the strain expressing Ub-M-β-gal because, in contrast to R-β-gal, M-β-gal is stable (Tobias et al. 1991). Two cDNA clones encoded enzymatic activities that, by the above criteria, were UBPs. The two cDNAs were found by restriction endonuclease analysis to be independent isolates of the same clone. The gene encoding this 1.54 kb cDNA was designated *AtUBP3*. We have designated this the third *Arabidopsis* UBP because the preliminary characterization of two *Arabidopsis* UBP cDNAs has been reported (Thoma et al. 1994; S. Thoma, Y. Ning, and R. Vierstra, personal communication).

The 1.54 kb *AtUBP3* cDNA was used as a hybridization probe with *Arabidopsis* genomic sequences to determine whether there were any other similar sequences in the *Arabidopsis* genome. Under stringency conditions that allow hybridization of sequences with 30% mismatch, hybridization was detected to a few sequences other than those accounted for by the *AtUBP3* cDNA (Fig. 1). This suggested that there might be an additional gene similar to *AtUBP3*.

The 1.54 kb *AtUBP3* cDNA was used as a hybridization probe to screen a second cDNA library and a 1.33 kb cDNA related to *AtUBP3* was isolated and has been designated *AtUBP4*. Hybridization of *AtUBP3* and *AtUBP4* sequences to Southern blots of genomic DNA indicated that these two genes account for all of the previously observed hybridizing sequences (data not shown).

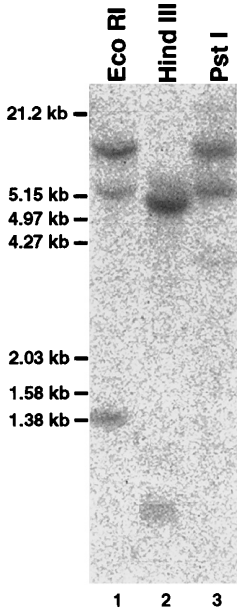

Fig. 1 Southern blot analysis of genomic DNA. A Southern blot was prepared from DNA (5 μg per lane) of ecotype Landsberg *erecta* of *Arabidopsis thaliana* digested with *Eco*RI (lane 1), *Hind*III (lane 2), or *Pst*I (lane 3) and fractionated on a 0.8% agarose gel. The blot was hybridized to radiolabeled 1.5 kb *At*UBP3 cDNA using hybridization and wash conditions (see Materials and methods) such that 30% mismatch was detectable

The *AtUBP3* and *AtUBP4* cDNAs encode highly similar UBPs

The *AtUBP3* and *AtUBP4* cDNAs were sequenced in their entirety and the sequences deposited in GenBank under accession numbers U76845 and U76846, respectively. The size difference between the cDNAs, 1542 bp vs 1332 bp, probably results from truncation of the 3′ untranslated region of *AtUBP4* at an *Eco*RI site. Owing to incomplete methylation at *Eco*RI sites, truncations are common in this cDNA library (M. Yanofsky, personal communication). *AtUBP3* and *AtUBP4* are predicted to encode proteins of 371 and 365 amino acids, respectively. Within the predicted coding region the cDNAs share 86% nucleotide identity, but only 56% and 39% in the 5′ and 3′ untranslated regions, respectively.

The *AtUBP4* cDNA contains an in-frame stop codon upstream of the first methionine, while the *AtUBP3* cDNA does not. This suggests that *AtUBP3* does not contain the entire coding region and that there is another methionine from which translation is initiated. This does not appear to be the case because Northern blot analysis with a probe specific for the 3′ untranslated region of *AtUBP3* detects an mRNA of approximately 1.6 kb (data not shown). This indicates that if another methionine is used, it is not much farther upstream. Two of the three *Arabidopsis* EST sequences identical to *AtUBP3* (119K8T7, 162M18T7, and 166K17T7) are only 7 nucleotides longer in the 5′ untranslated region, but they also do not have stop codons upstream of the putative initiator methionine. Thus, the Northern blot data along with the additional cDNAs suggest that *AtUBP3* contains the entire coding region.

The two proteins are 93% identical and 96% similar (Fig. 2). UBPs are characterized by the presence of four

conserved amino acid domains, one containing the putative active cysteine residue (Cys box), one containing two conserved histidines (His box), and two other regions of significant amino acid similarity (Papa and Hochstrasser 1993; Wilkinson et al. 1995). *At*UBP3 and *At*UBP4 contain all four of these domains (Fig. 2).

```
                                                        *
AtUBP3     1 MGAAGSKLEKALGDQFPEGERYFGFENFGNTCYCN 35
AtUBP4     1 ------------------------------------ 35

AtUBP3    36 SVLQALYFCVPFREQLLEYYTSNKSVADAEENLMT 70
AtUBP4    36 ---------A--------H-AN--..-------L- 68

AtUBP3    71 CLADLFSQISSQKKKTGVIAPKRFVQRLKKQNELF 105
AtUBP4    69 ----------_________________________- 103
                                  N1

AtUBP3   106 RSYMHQDAHEFLNYLLNEVVDILEKEAKATKTEHE 140
AtUBP4   104 ----------------L-E-----TQ---ADN- 138

AtUBP3   141 TSSSSSPEKIANGLKVPQANGVVHKEPIVTWVHNI 175
AtUBP4   139 -..---------V--A-L---.----------K- 170

AtUBP3   176 FQGILTNETRCLRCETVTARDETFLDLSLDIEQNS 210
AtUBP4   171 ------------------------------------ 205

AtUBP3   211 SITSCLKNFSSTETLHAEDKFFCDKCCSLQEAQKR 245
AtUBP4   206 ----------------------------__-- 240

AtUBP3   246 MKIKKPPHILVIHLKRFKYIEQLGRYKKLSYRVVF 280
AtUBP4   241 _____----------M---------- 275
             N2
                                        *       *
AtUBP3   281 PLELKLSNTVEPYADVEYSLFAVVVHVGSGPNHGH 315
AtUBP4   276 ----------DE-V-I-_________________ 310
                                      Block 1

AtUBP3   316 YVSLVKSHNHWLFFDDENVEMIEESAVQTFFGSSQ 350
AtUBP4   311 ___-------_____S--I-------------- 345
                    Block 2

AtUBP3   351 EYSSNTDHGYILFYESLGPTK 371
AtUBP4   346 --------____L-----.-R 365
                  Block 3
```

Fig. 2 Alignment of the deduced amino acid sequences of *At*UBP3 and *At*UBP4 cDNAs. Identical amino acids are indicated by *dashes* and gaps are indicated by *dots*. In *bold* are the four conserved domains found in ubiquitin-specific proteases (UBPs): the domain containing the conserved cysteine (Cys box, amino acids 22–39 of *At*UBP3 and *At*UBP4); the domain containing the conserved histidines (His box, amino acids 298–364 of *At*UBP3 and 292–359 of *At*UBP4); two other conserved regions (amino acids 110–126 of *At*UBP3 and 108–124 of *At*UBP4 and 251–262 of *At*UBP3 and 246–257 of *At*UBP4). The His box has been divided into three blocks, *underlined*, to facilitate comparison with other ubiquitin-specific proteases (see Fig. 4) *Asterisks* denote the conserved cysteine and histidine residues. The *underlined region N1* denotes a putative bipartite nuclear localization signal (amino acids 83–100 of *At*UBP3 and 81–98 of *At*UBP4). The *underlined region N2* denotes a putative SV40-like nuclear localization signal (amino acids 244–250 of *At*UBP3 and 239–245 of *At*UBP4)

Alignment of *At*UBP3 and *At*UBP4 with the deduced protein sequences of genes known to encode active UBPs showed that *At*UBP3 and *At*UBP4 possess all absolutely conserved residues within the Cys box (Fig. 3A). The His box can be broken down into three separate blocks of conserved residues. Within all three of the blocks *At*UBP3 and *At*UBP4 possess all absolutely conserved residues (Fig. 3B). *At*UBP3 and *At*UBP4 can clearly be considered to encode UBPs by these criteria.

Database searches revealed that outside of the conserved regions *At*UBP3 and *At*UBP4 share significant similarity with only one other UBP; this is a putative UBP encoded in the *Caenorhabditis elegans* genome, UBPX (SwissProt:p34547). UBPX shares an overall amino acid identity of 53% and 55% with *At*UBP3 and *At*UBP4, respectively, and 70% amino acid similarity with both. UBPX, at 408 amino acids, is slightly larger than *At*UBP3 and *At*UBP4 because it contains an additional block of amino acids in a region between the Cys and His boxes.

*At*UBP3 and *At*UBP4 contain a putative bipartite nuclear localization signal (amino acids 83–100 for

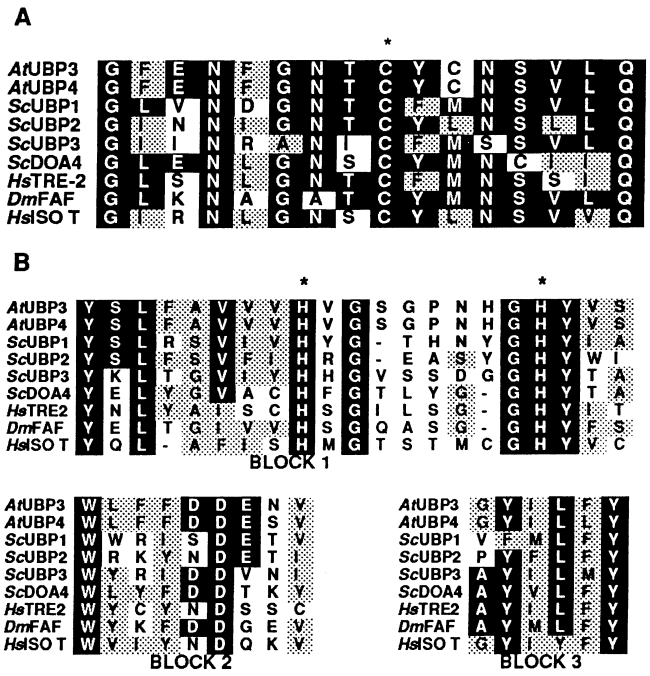

Fig. 3A, B Alignment of the active site cysteine and histidine boxes of *At*UBP3 and *At*UBP4 with those of other characterized UBPs. The Cys boxes (**A**) and the three conserved domains within the His boxes (**B**) of *At*UBP3, *At*UBP4, *Sc*UBP1 (M63484; Tobias and Varshavsky 1991), *Sc*UBP2 (M94916; Baker et al. 1992), *Sc*UBP3 (M94917; Baker et al. 1992), *Sc*DOA4 (U02518; Papa and Hochstrasser 1993), *Hs*TRE2 (EMBL accession numbers X63546 and X63547; Nakamura et al. 1992), *Dm*FAF (L04958, L04959, and L04960; Fischer-Vize et al. 1992), and *Hs*ISOT (U35116; Wilkinson et al. 1995) have been aligned using the Pileup program (Wisconsin Genetics Computer Group package). Unless otherwise indicated, numbers after genes are GenBank accession numbers. Amino acid identity is indicated with *reverse type*. Amino acid similarity is indicated by *shading*. Similarity is defined as follows: G = A = S, I = L = M = V, F = L = I = Y, and K = R

*At*UBP3 and 81–98 for *At*UBP4, N1 in Fig. 2) as well as a K/R-K/R-X-K/R-X-K/R SV40-like nuclear localization signal (amino acids 244–250 for *At*UBP3 and amino acids 239–245 for *At*UBP4, N2 in Fig. 2).

MBP*At*UBP3 and MBP*At*UBP4 can cleave the α-linked ubiquitin fusion protein Ub-X-β-gal but not the α-linked ubiquitin fusion protein Ub-CEP52

The cleavage of Ub-M-β-gal and Ub-R-β-gal by *At*UBP3 and *At*UBP4 was verified by Western blotting of protein extracts from *E. coli* coexpressing the substrates, and the UBPs as MBP fusion proteins. MBP-UBP proteins in which the conserved cysteine residue was replaced by a serine residue were also tested using the same assay. Immunoblot analysis of the cell lysates from these strains with anti-β-gal antibody (Fig. 4A) showed that both MBP*At*UBP3 and MBP*At*UBP4 were capable of cleaving this substrate. As expected, cleavage of Ub-M-β-gal left the stable M-β-gal and cleavage Ub-R-β-gal resulted in undetectable steady-state levels of β-gal owing to the instability of R-β-gal (Tobias et al. 1991). The mutant proteins, MBP*At*ubp3^{C32S} and MBP*At*ubp4^{C32S}, lacking the active site cysteine residue, were inactive against these substrates. No activity against the Ub-M-β-gal substrate was ever detected in in vitro reactions in which *E. coli* extracts expressing Ub-M-β-gal were mixed with *E. coli* extracts expressing either MBP*At*UBP3 or MBP*At*UBP4 (data not shown).

If *At*UBP3 and *At*UBP4 act in vivo to produce ubiquitin monomers from the ubiquitin ribosomal protein fusions and the polyubiquitin fusions, then they should be able to act upon these substrates in vivo in *E. coli*. MBP*At*UBP3 and MBP*At*UBP4 were coexpressed with Ub-CEP52 [*Arabidopsis UBQ1* gene product (Callis et al. 1990)] in BL21(DE3). Immunoblot analysis with anti-ubiquitin antibody of the cell lysates

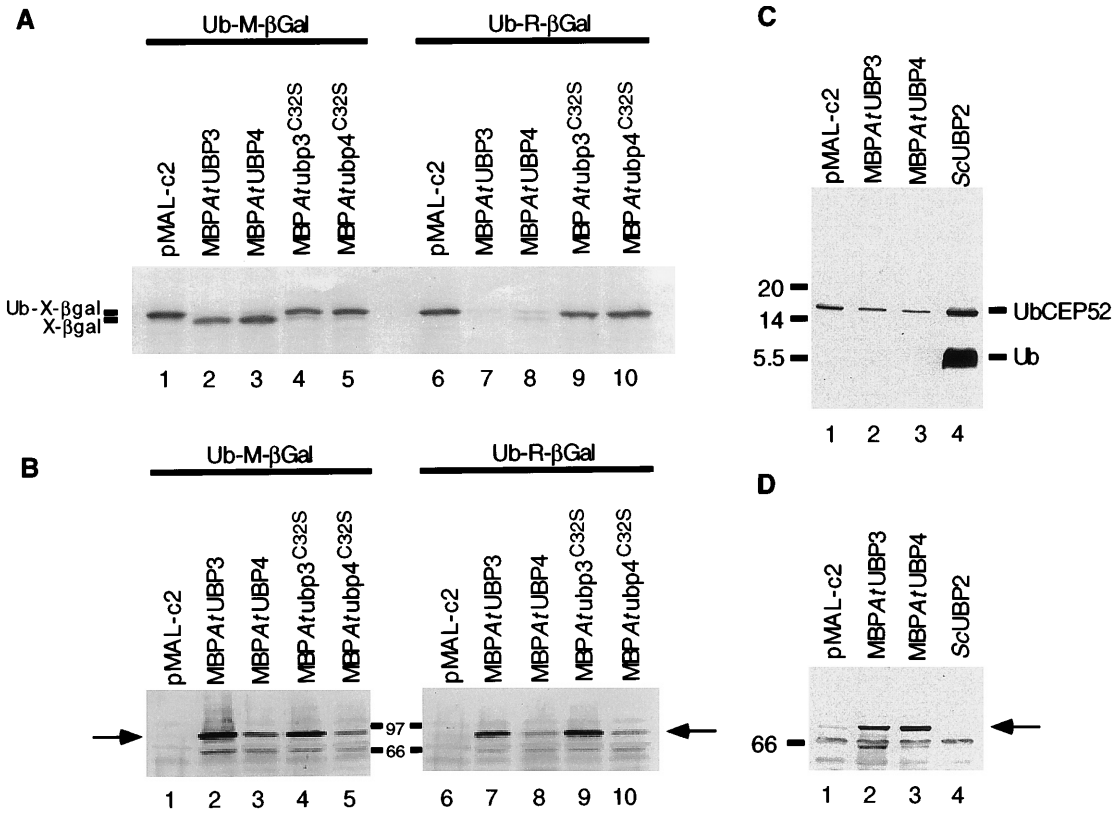

Fig. 4A–D *At*UBP3 and *At*UBP4 deubiquitinate the artificial α-linked substrate Ub-X-β-gal but not the natural α-linked substrate UbCEP52. *At*UBP3 and *At*UBP4 were assayed for their ability to catalyze the cleavage of ubiquitin α-linked to other proteins, β-gal (**A**, **B**) and CEP52 (**C**, **D**) using an in vivo *Escherichia coli* assay. Culture conditions for *E. coli* and preparation of protein extracts are described in Materials and methods. **A** β-Gal is visualized by anti-β-gal antibody immunoblot analysis of *E. coli* proteins fractionated by electrophoresis on a 7% SDS-polyacrylamide gel. Ub-M-β-gal (lanes 1–5) and Ub-R-β-gal (lanes 6–10) (Baker et al. 1992) were co-expressed in *E. coli* with the vector-only control (lanes 1, 6), MBP*At*UBP3 (lanes 2, 7), MBP*At*UBP4 (lanes 3, 8), MBP*At* ubp3^{C32S} (lanes 4, 9), and MBP*At* ubp4^{C32S} (lanes 5, 10). **B** *At*UBP3 and/or *At*UBP4 are visualized by anti-*At*UBP3 and *At*UBP4 immunoblot analysis of samples fractionated by electrophoresis on a 10% SDS-polyacrylamide gel. All samples were identical to those in **A**. **C** Ubiquitin is visualized by anti-ubiquitin antibody immunoblot analysis of *E. coli* proteins fractionated by electrophoresis on a 15% SDS-polyacrylamide gel. UbCEP52 was coexpressed in *E. coli* BL21 (DE3) with vector-only control (lane 1), MBP*At*UBP3 (lane 2), MBP*At*UBP4 (lane 3), and *Saccharomyces cerevisiae* UBP2 (Baker et al. 1992) (lane 4). **D** *At*UBP3 and/or *At*UBP4 are visualized by anti-*At*UBP3 and *At*UBP4 immunoblot analysis of the same samples fractionated by electrophoresis on a 10% SDS-polyacrylamide gel

from these strains showed that both MBP*At*UBP3 and MBP*At*UBP4 are inactive against this substrate (Fig. 4C). The ability of *Sc*UBP2 to cleave this substrate indicated that assay conditions were appropriate (Baker et al. 1992). Immunoblot analysis with anti-UBP3/4 indicates that the MBP*At*UBP enzymes were present in these strains (Fig. 4D) and at levels comparable to those capable of acting upon the Ub-X-β-gal substrates (Fig. 4B). In addition, MBP*At*UBP3 and MBP*At*UBP4 were coexpressed in vivo with polyubiquitin [produced from the *Arabidopsis UBQ10* gene (Callis et al. 1995)]. No activity was detected against this α-linked substrate (data not shown).

*At*UBP3 and/or *At*UBP4 are present in all organs of 6–8 week old *Arabidopsis*

To elucidate the function of *At*UBP3 and *At*UBP4 we determined their distribution of expression. Affinity-purified antibody primarily recognized a single protein at 42 kDa, the predicted molecular weight of *At*UBP3 and *At*UBP4 (Fig. 5). *At*UBP3 and/or *At*UBP4 are found in all organs examined, with lower levels in siliques (Fig. 5). In addition, hybridization of the *At*UBP3 cDNA sequence to a Northern blot made from poly(A)-enriched RNA isolated from 5 day old seedlings indicates that the transcript is present in this organ (data not shown).

*At*UBP3 and/or *At*UBP4 are present in the *Arabidopsis* nucleus

Because the deduced amino acid sequences of *At*UBP3 and *At*UBP4 contain putative nuclear localization signals, we determined whether the protein is found in nuclear extracts. *At*UBP3 and/or *At*UBP4 are present in the sample containing the nuclear extract (Fig. 6, lane N). *At*UBP3 and/or *At*UBP4 are also detected in the sample containing an amount of cytoplasmic marker equal to that found in the nuclear protein extract (Fig. 6, lane C). The lower signal in the cytosolic marker lane, relative to that in the lane containing nuclear extract (Beers et al. 1992), demonstrates that the signal from the nuclear extract cannot arise solely from cytosolic contamination. Thus, *At*UBP3 and/or *At*UBP4 are present in the nucleus.

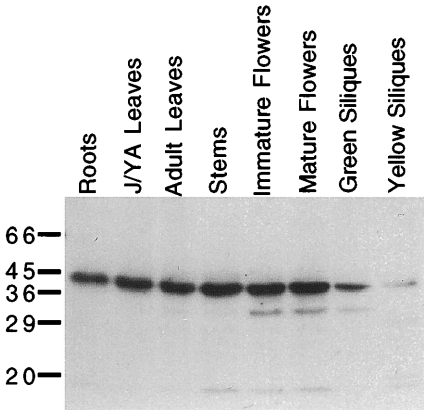

Fig. 5 *At*UBP3 and *At*UBP4 are found in all organs of 6–8 week old *Arabidopsis*. *At*UBP3 and/or *At*UBP4 are visualized by anti-*At*UBP3 and *At*UBP4 immunoblot analysis of samples fractionated by electrophoresis on a 10% SDS-polyacrylamide gel. Each lane contains 50 μg of protein isolated from the indicated organ. The J/YA Leaves sample is enriched in juvenile and young adult stage leaves and the Adult Leaves sample is enriched in adult leaves (Bowman 1994). Equivalent loading was verified by Coomassie blue staining of a duplicate gel (data not shown)

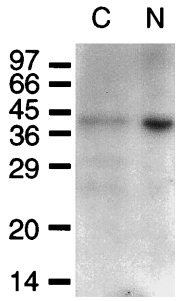

Fig. 6 *At*UBP3 and *At*UBP4 are enriched in *Arabidopsis* nuclear fractions. *At*UBP3 and/or *At*UBP4 are visualized by anti-*At*UBP3 and *At*UBP4 immunoblot analysis of samples fractionated by electrophoresis on a 13.5% SDS-polyacrylamide gel. Lane N contains 20 μg of protein from purified nuclei. Lane C contains 0.7 μg of protein from whole leaf extract. The two samples contain equal levels of the cytosol-localized protein phospho*enol*pyruvate carboxylase. Lane C represents a level of cytosolic contamination equivalent to that found in the nuclear extract (Beers et al. 1992)

Discussion

We have used a functional assay in *E. coli* (Baker et al. 1992) to identify an *Arabidopsis* cDNA encoding a UBP, *At*UBP3, and subsequently isolated a cDNA encoding a highly related second UBP, *At*UBP4. Analysis of the deduced amino acid sequences of *At*UBP3 and *At*UBP4 shows that these proteins have all of the conserved amino acid regions characteristic of UBPs, namely the domains containing the active site cysteine and the active site histidines and two other regions of similarity. *At*UBP3 and *At*UBP4 were predicted to be localized in the nucleus because they have both bipartite and SV40-like nuclear localization signals. Both bipartite and SV40-like nuclear localization signals have been observed in several plant nuclear proteins (Carrington et al. 1991; Abel and Theologis 1995). In addition, the presence of UBPs in the nucleus is consistent with the presence of ubiquitinated proteins in the plant nucleus (Beers et al. 1992). Immunoblotting with anti-*At*UBP3/4 shows enrichment of *At*UBP3 and/or *At*UBP4 in nuclear extracts when compared with the level of *At*UBP3 and/or *At*UBP4 in an amount of cytosolic extract equal to that contaminating the nuclear extract. Low levels of *At*UBP3 and/or *At*UBP4 are detected in the cytosol

control sample, owing either to the presence of *At*UBP3 and/or *At*UBP4 in the cytosol or to contamination of the cytosol samples with nuclear extract. Experiments are in progress to distinguish between these two possibilities by using transgenic plants expressing an *At*UBP-β-glucuronidase fusion. Nevertheless, *At*UBP3 and/or *At*UBP4 are indeed present in the *Arabidopsis* nucleus.

Most of the previously characterized UBPs are quite large (Wilkinson et al. 1995). All consist of a core, defined as the region flanked by the Cys box at the N-terminus and the His box at the C-terminus. Most UBPs, however, have large N-terminal extensions and occasionally have C-terminal extensions. Strikingly, *At*UBP3 and *At*UBP4 are the smallest proteins so far described in the UBP family (Wilkinson et al. 1995). The catalytic core has been defined as minimally 314 residues (Wilkinson et al. 1995). While the core region of *At*UBP3 and *At*UBP4 is a bit larger than the minimal core, spanning 348 and 342 residues, respectively, *At*UBP3 and *At*UBP4 contain no significant N- or C-terminal extensions. *At*UBP3 and *At*UBP4 share significant similarity outside of the conserved domains with only one other sequence, that of the putative UBP UBPX (SwissProt:p34547) encoded by a *C. elegans* gene. *At*UBP3 and *At*UBP4 are even slightly smaller than the predicted open reading frame of this putative UBP. *At*UBP3 (371 amino acids) and *At*UBP4 (365 amino acids) lack a block of amino acids found in the internal core region of UBPX (408 amino acids).

We have used the substrates from the in vivo *E. coli* screen, Ub-M-β-gal and Ub-R-β-gal, that identified *At*UBP3, to demonstrate the UBP activity of *At*UBP3 and *At*UBP4 directly (expressed as fusions with the MBP). We also investigated the activity of MBP-*At*UBP3 and MBP*At*UBP4 against natural α-amino linked substrates, namely polyubiquitin and ubiquitin-ribosomal protein fusions, upon which these enzymes might act in vivo. We have been unable to detect any activity against these substrates. It might be that the rather large N-terminal extension produced by the MBP is sterically interfering with the ability of the enzymes to act. It is equally possible, however, that *At*UBP3 and *At*UBP4 do not naturally act upon α-amino-linked substrates. This would not be unusual because another UBP, isopeptidase T, is very inefficient at processing α-amino-linked substrates (Wilkinson et al. 1995). *At*UBP3 and *At*UBP4 might be able to cleave ubiquitin from Ub-M-β-gal and Ub-R-β-gal only by virtue of some unique characteristic of these artificial substrates.

*At*UBP3 and *At*UBP4 might instead cleave ubiquitin at ε-amino linkages such as those found in polyubiquitin chains or ubiquitinated cellular proteins. Although we have no direct evidence to support this hypothesis, we do have indirect evidence. *At*UBP3 and *At*UBP4 are found in the nucleus, where the most likely substrates are polyubiquitin chains and ubiquitinated cellular proteins. The processing of the α-amino-linked substrates, polyubiquitin and ubiquitin-ribosomal fusion proteins, occurs in the cytosol, possibly co-translationally since intermediates cannot be detected (Finley et al. 1989). It therefore seems highly unlikely that these are the natural substrates, leaving only the ε-amino-linked substrates as candidate targets for these enzymes.

Additional information relevant to the in vivo role of *At*UBP3 and *At*UBP4 was obtained by examining the organ specificity of expression. *At*UBP3 and/or *At*UBP4 protein is not restricted to a particular organ. This suggests that these enzymes are not being produced in a organ-specific manner to act upon a substrate and effect a particular developmental process. This leaves several possibilities. *At*UBP3 and *At*UBP4 could be involved with (1) a general ubiquitin pathway function required in all organs, such as recycling ubiquitin monomers following proteolysis, (2) deubiquitinating a substrate protein that is present in many different organs, or (3) deubiquitinating a organ-specific substrate protein. Since ubiquitin-mediated proteolysis is the mechanism by which many cell cycle regulatory proteins are degraded (reviewed in King et al. 1996), the second possibility could include substrates involved in cell cycle regulation. This, however, seems unlikely because *At*UBP3 and/or *At*UBP4 are present in mature leaves and senescing siliques, organs that have few, if any, cells undergoing division. The third possibility might be true for at least one UBP. The *Drosophila* UBP Fat facets appears to be produced in many organs, yet the effects of *fat facets* mutations are only apparent during eye and embryo development (Fischer-Vize et al. 1992).

The next step toward understanding the in vivo function of *At*UBP3 and *At*UBP4 will be facilitated by identifying proteins with which these enzymes interact. This will provide additional information as to the type(s) of substrate upon which to focus future research efforts.

Acknowledgements We thank Dr. Eric Beers for kindly providing the subcellular fractionation blots. We thank Rohan Baker for providing the constructs encoding Ub-R-β-gal, Ub-M-β-gal and *Sc*UBP2. This work was supported by an NSF grant (DCB 93-06759) and an NSF Presidential Young Investigator Award (DCB 91-58453), both to Judy Callis.

References

Abel S, Theologis A (1995) A polymorphic bipartite motif signals nuclear targeting of early auxin-inducible proteins related to PS-IAA4 from pea (*Pisum sativum*). Plant J 8:87–96

Baker RT, Tobias JW, Varshavsky A (1992) Ubiquitin-specific proteases of *Saccharomyces cerevisiae*; cloning of *UBP2* and *UBP3*, and functional analysis of the *UBP* gene family. J Biol Chem 267:23364–23375

Beers EP, Moreno TN, Callis J (1992) Subcellular localization of ubiquitin and ubiquitinated proteins in *Arabidopsis thaliana*. J Biol Chem 267:15432–15439

Bowman J (1994) Vegetative development In: Bowman J (ed) Arabidopsis: an atlas of morphology and development. Springer-Verlag, New York, pp 1–89

Callis J, Raasch JA, Vierstra RD (1990) Ubiquitin extension proteins of *Arabidopsis thaliana*; structure, localization and ex-

pression of their promoters in transgenic tobacco. J Biol Chem 265:12486–12493
Callis J, Carpenter T, Sun C-W, Vierstra RD (1995) Structure and evolution of genes encoding polyubiquitin and ubiquitin-like proteins in *Arabidopsis thaliana* ecotype Columbia. Genetics 139:921–939
Carlson N, Rogers S, Rechsteiner M (1987) Microinjection of ubiquitin: changes in protein degradation in HeLa cells subjected to heat-shock. J Cell Biol 104:547–555
Carrington JC, Freed DD, Leinicke AJ (1991) Bipartite signal sequence mediates nuclear translocation of the plant potyviral Nla protein. Plant Cell 3:953–962
Chau V, Tobias JW, Bachmair A, Marriott D, Ecker DJ, Gonda DK, Varshavsky A (1989) A multiubiquitin chain is confined to specific lysine in a targeted short-lived protein. Science 243:1576–1583
Chen Z, Pickart CM (1990) A 25-kilodalton ubiquitin carrier protein (E2) catalyzes multi-ubiquitin chain synthesis via lysine 48 of ubiquitin. J Biol Chem 265:21835–21842
Dellaporta S, Wood J, Hicks JB (1983) A plant DNA minipreparation: version II. Plant Mol Biol Rep 1:19–21
Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to a high specific activity. Anal Biochem 132:6–13
Finley D, Bartel B, Varshavsky A (1989) The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. Nature 338:394–401
Fischer-Vize JA, Rubin GM, Lehmann R (1992) The *fat facets* gene is required for *Drosophila* eye and embryo development. Development 116:985–1000
Gallagher S, Winston SE, Fuller SA, Hurrell JGR (1994) Immunoblotting and immunodetection In: Ausubel FM, Brent R, Kingston RE et al (eds) Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York, pp 10.8.1–17
Haas AL, Bright PM (1985) The immunochemical detection and quantitation of intracellular ubiquitin-protein conjugates. J Biol Chem 260:12464–12473
Harlow E, Lane D (1988) Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Hershko A, Ciechanover A (1992) The ubiquitin system for protein degradation. Annu Rev Biochem 61:761–807
Hershko A, Eytan NE, Ciechanover A, Hass AL (1982) Immunochemical analysis of the turnover of ubquitin conjugates in intact cells. J Biol Chem 257:13964–13970
Huang Y, Baker RT, Fischer-Vize JA (1995) Control of cell fate by a deubiquitinating enzyme encoded by the *fat facets* gene. Science 270:1828–1831
Jentsch S (1992) The ubiquitin-conjugation system. Annu Rev Genet 26:179–207
King RW, Deshaies RJ, Peters J, Kirschner MW (1996) How proteolysis drives the cell cycle. Science 274:1652–1659
Kunkel TA (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc Natl Acad Sci USA 82:488–492
Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T_4. Nature 227:680–685
Moazed D, Johnson AD (1996) A deubiquitinating enzyme interacts with SIR4 and regulates silencing in *S. cerevisiae*. Cell 86:667–677
Nakamura T, Hillova J, Mariage-Samson R, Onno M, Huebner K, Cannizzaro LA, Boghosian-Sell L, Croce CM, Hill M (1992) A novel trascriptional unit of the *tre* oncogene widely expressed in human cancer cells. Oncogene 7:733–741
Papa FR, Hochstrasser M (1993) The yeast *DOA4* gene encodes a deubiquitinating enzyme related to a product of the human *tre-2* oncogene. Nature 366:313–319
Rose IA (1988) Ubiquitin carboxyl-terminal hydrolases. In: Rechsteiner M (ed) Ubiquitin. Plenum Press, New York, pp 135–155
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467
Sullivan ML, Callis J, Vierstra RD (1990) High performance liquid chromatography resolution of ubiquitin pathway enzymes from wheat germ. Plant Physiol 94:710–716
Thoma S, Hatfield P, Vierstra R (1994) Characterization of a ubiquitin carboxy-terminal hydrolase from Arabidopsis. Plant Physiol 105:S32
Tobias JW, Varshavsky A (1991) Cloning and functional analysis of the ubiquitin-specific protease gene UBP1 of *Saccharomyces cerevisiae*. J Biol Chem 266:12021–12028
Tobias JW, Shrader TE, Rocap G, Varshavsky A (1991) The N-end rule in bacteria. Science 254:1374–1377
Weigel D, Alvarez J, Smyth DR, Yanofsky MF, Meyerowitz EM (1992) *LEAFY* controls floral meristem identity in Arabidopsis. Cell 69:843–859
Wilkinson KD (1994) Cellular roles of ubiquitin In: Mayer RJ, Brown IR (eds) Heat shock proteins in the nervous system. Academic Press, London, pp 191–234
Wilkinson KD, Tashayev VL, O'Connor LB, Larsen CN, Kasperek E, Pickart CM (1995) Metabolism of the polyubiquitin degradation signal: structure, mechanism, and role of isopeptidase T. Biochemistry 34:14535–14546
Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *agamous* resembles transcription factors. Nature 346:35–39
Zhu Y, Carroll M, Papa FR, Hochstrasser M, D'Andrea AD (1996) *DUB*-1 a deubiquitinating enzyme with growth-suppressing activity. Proc Natl Acad Sci USA 93:3275–3279