Registry No. AChE, 9000-81-1; ethanolamine, 141-43-5; glucosamine, 3416-24-8.

REFERENCES

Baltz, T., Duvillier, G., Giroud, C., Richet, C., Baltz, D., & Degand, P. (1983) *FEBS Lett. 158*, 174–178.

Black, S. D., & Coon, M. J. (1982) *J. Biol. Chem. 257*, 5929–5938.

Borst, P., & Cross, G. A. M. (1982) *Cell (Cambridge, Mass.) 29*, 291–303.

Campbell, D. G., Gagnon, J., Reid, K. B. M., & Williams, A. F. (1981) *Biochem. J. 195*, 15–30.

Dutta-Choudhury, T. A., & Rosenberry, T. L. (1984) *J. Biol. Chem. 259*, 5653–5660.

Edman, P., & Henschen, A. (1975) in *Protein Sequence Determination* (Needleman, S. B., Ed.) 2nd ed., pp 232–279, Springer-Verlag, New York.

Ferguson, M. A. J., Haldar, K., & Cross, G. A. M. (1985a) *J. Biol. Chem. 260*, 4963–4968.

Ferguson, M. A. J., Low, M. G., & Cross, G. A. M. (1985b) *J. Biol. Chem. 260*, 14547–14555.

Futerman, A. H., Fiorini, R. M., Roth, E., Michaelson, D. M., Low, M. G., & Silman, I. (1984) in *Cholinesterases—Fundamental and Applied Aspects* (Brzin, M., Barnard, E. A., & Sket, D., Eds.) pp 99–113, de Gruyter, West Berlin.

Haas, R., & Rosenberry, T. L. (1985) *Anal. Biochem. 148*, 154–162.

Holder, A. A. (1983) *Biochem. J. 209*, 261–262.

Kim, B. H., & Rosenberry, T. L. (1985) *Biochemistry 24*, 3586–3592.

Low, M. G., & Kincade, P. W. (1985) *Nature (London) 318*, 62–64.

Massoulié, J., & Bon, S. (1982) *Annu. Rev. Neurosci. 5*, 57–106.

Roberts, W. L., & Rosenberry, T. L. (1985) *Biochem. Biophys. Res. Commun. 133*, 621–627.

Roberts, W. L., & Rosenberry, T. L. (1986a) *Biochemistry* (preceding paper in this issue).

Roberts, W. L., & Rosenberry, T. L. (1986b) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 45*, 1816.

Rosenberry, T. L. (1985) *Enzymes Biol. Membr. 4*, 403–429.

Rosenberry, T. L., & Scoggin, D. M. (1984) *J. Biol. Chem. 259*, 5643–5652.

Rosenberry, T. L., Scoggin, D. M., Dutta-Choudhury, T. A., & Haas, R. (1984) in *Cholinesterases—Fundamental and Applied Aspects* (Brzin, M., Barnard, E. A., & Sket, D., Eds.) pp 155–172, de Gruyter, West Berlin.

Rosenberry, T. L., Roberts, W. L., & Haas, R. (1986) *Fed. Proc., Fed. Am. Soc. Exp. Biol.* (in press).

Schmidt, M. F. G. (1983) *Curr. Top. Microbiol. Immunol. 102*, 101–129.

Seki, T., Chang, H.-C., Moriuchi, T., Denome, R., Ploegh, H., & Silver, J. (1985) *Science (Washington, D.C.) 227*, 649–651.

Sherman, G., Rosenberry, T. L., & Sternlicht, H. (1983) *J. Biol. Chem. 258*, 2148–2156.

Tse, A. G. D., Barclay, A. N., Watts, A., & Williams, A. F. (1985) *Science (Washington, D.C.) 230*, 1003–1008.

Complete Amino Acid Sequence of Ubiquitin from the Higher Plant *Avena sativa*†

Richard D. Vierstra,* Susan M. Langan, and G. Eric Schaller

Department of Horticulture, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received November 26, 1985; Revised Manuscript Received February 13, 1986

ABSTRACT: Ubiquitin is a 76 amino acid, eukaryotic polypeptide with several important regulatory functions that arise from its ability to become covalently ligated to cytoplasmic and nuclear proteins. The amino acid sequence of ubiquitin is remarkably conserved, being identical for all animal forms analyzed to date. Here, we present the complete amino acid sequence of ubiquitin isolated from a higher plant, oats (*Avena sativa L.*). This sequence was determined by repetitive Edman degradation of the intact molecule and of peptides derived by proteolytic digestion with trypsin or *Staphylococcus aureus* V8 protease. Comparison of this sequence with that obtained for animal ubiquitins indicates that the two forms are homologous but *not* identical. Like the animal form, oat ubiquitin contains 76 amino acid residues, no tryptophan or cysteine, and a carboxyl terminus of Leu-Arg-Gly-Gly. However, oat ubiquitin contains three amino acid substitutions at positions 19, 24, and 57.

Ubiquitin is a 76 amino acid polypeptide present in all eukaryotes. Its amino acid sequence is identical in organisms as diverse as mammals including humans (Schlesinger et al., 1975; Schlesinger & Goldstein, 1975), birds (Bond & Schlesinger, 1985), fish (Watson et al., 1978), amphibians (Dworkin-Rastl et al., 1984), and insects (Gavilanes et al., 1982), making it one of the most conserved proteins yet identified. The widespread occurrence and unparalleled sequence conservation of ubiquitin led Goldstein et al. (1975) to conclude long before its functions became evident that this protein plays an important role in eukaryotic cell physiology.

It is now known that ubiquitin has at least two functions in intracellular metabolism that are a result of its ability to become covalently ligated to other cytoplasmic and nuclear proteins [for reviews, see Hershko & Ciechanover (1982) and Finley & Varshavsky (1985)]. Ligation is via an unusual

† This work has been supported by grants from the National Science Foundation (DMB-8409210), the USDA-CRGO (85-CRCR-1-1547), and the Research Division of the College of Agriculture and Life Sciences (Hatch 2858) of the University of Wisconsin—Madison.

* Author to whom correspondence should be addressed.

0006-2960/86/0425-3105$01.50/0

peptide linkage between the α- and ϵ-amino groups on the various target proteins and the terminal glycine carboxyl group of ubiquitin. In the cytoplasm, ubiquitin conjugation appears to be an obligatory step for the degradation of many proteins, especially those proteins with short half-lives (Ciechanover et al., 1984; Hershko et al., 1984). This modification results in the rapid catabolism of the tagged protein with the concomitant release of free ubiquitin. In the nucleus, ubiquitin has been discovered conjugated to two specific proteins, histones H2A and H2B (Goldknopf & Busch, 1977). Here, conjugation is apparently reversible, functioning not as a method for degrading the affected histones (Wu et al., 1981) but possibly as a way to reversibly modify nucleosome structure. Variations in the level of these two ubiquitinated histones with respect to the cell cycle (Matsui et al., 1982; Mueller et al., 1985) and transcriptional activity (Goldknopf et al., 1980) and their selective occurrence within actively transcribed regions of chromatin (Levinger & Varshavsky, 1982) suggest that histone ubiquitination may play a role in altering gene expression.

We have recently begun to characterize ubiquitin and its related conjugation reactions in higher plants in an effort to understand its functions in these tissues. Initial characterization of the protein isolated from oats (*Avena sativa L.*) indicated that structural differences between animal and plant ubiquitins may exist (Vierstra et al., 1985). Here we report the complete amino acid sequence of oat ubiquitin. Comparison of this sequence with that obtained for animal ubiquitins provides definitive evidence that animal and plant ubiquitins do differ.

Materials and Methods

Chemicals. Ubiquitin was purified to homogeneity (based on $NaDodSO_4$–polyacrylamide gel electrophoresis)[1] from etiolated oat shoots (*Avena sativa L.*, cv. Garry) according to the procedure of Vierstra et al. (1985). L-1-(Tosylamido)-2-phenylethyl chloromethyl ketone treated bovine pancreatic trypsin was purchased from Sigma. *Staphylococcus aureus* V8 protease, high-performance liquid chromatography (HPLC-) grade trifluoroacetic acid, triethylamine, and all the reagents and supports used for Edman degradation of intact ubiquitin were obtained from Pierce. HPLC-grade acetonitrile was a product of Burdick and Jackson.

Enzymatic Digestions and Purification of Peptides. Tryptic digests of purified oat ubiquitin were prepared according to the method of Cox et al. (1986). Ubiquitin (4 mg/mL) was dissolved in 0.3 M potassium phosphate and 6 M urea, pH 7.8, and incubated with 5% w/w trypsin for 2 h at 37 °C. After 2 h, an additional 5% w/w trypsin was added and the reaction allowed to proceed for another 2 h. *S. aureus* V8 protease digestions were performed under the conditions in which the cleavages are specific for glutamyl residues (Houmard & Drapeau, 1972). Ubiquitin (2.5 mg/mL) was dissolved in 100 mM NH_4HCO_3 and 2 mM tetrasodium ethylenediaminetetraacetic acid, pH 7.8, and incubated with 10% w/w V8 protease for 72 h at 37 °C. Preliminary experiments indicated that complete digestion of ubiquitin with V8 protease required such an extended incubation period. The digestions with either protease were terminated by rapid freezing of the sample to 77K.

Peptides generated by digestion with either protease were purified by reverse-phase chromatography with a Vydac 25-cm stainless steel C18 column in conjunction with a Varian Model 5000 HPLC. Absorbance of the HPLC eluate was monitored at 210 and 280 nm on a Hewlett-Packard Model 1040 diode array detector. Absorbance profiles at 280 nm were used to tentatively identify tyrosine- and phenylalanine-containing peptides. Solvent A was 0.1% trifluoroacetic acid–0.036% triethylamine dissolved in water, and solvent B was 0.083% trifluoroacetic acid–0.03% triethylamine dissolved in 65% acetonitrile. Separation of peptides was accomplished with a 34-min linear gradient (flow rate = 1 mL/min) from 2 to 65% solvent B (Cox et al., 1986). Peptide-containing peaks from the HPLC were collected, lyophilized, and used directly for sequence analysis. Approximately 0.4 μg (50 nmol) of tryptic digest and 0.2 μg of (25 nmol) of V8 digest were injected per HPLC run. Yields of the various peptides from the HPLC column were ~10%.

Automated Sequence Analyses. Automated amino acid sequence analyses of the intact protein were accomplished with the Sequemat Model 12 solid-phase protein sequencer after first coupling the protein via lysine residues to *p*-phenylene diisothiocyanate activated aminopropyl glass (L'Italien & Strickler, 1982). Approximately 50 nmol was used per run with repetitive yields of 90–92%. Sequence analyses of the HPLC-purified fragments were performed on an Applied Biosystems 470A gas-phase protein sequencer with no vacuum programs (Hunkapillar et al., 1983). Between 2 and 5 nmol of peptide was used per analysis with repetitive yields between 88 and 91%. Phenyl isothiocyanate derivatized amino acids were converted to the phenylhydantoin forms with methanolic HCl and identified and quantitated by HPLC with a C18 column.

Results

The complete amino acid sequence of oat ubiquitin was determined from automated sequence analysis of the intact molecule and of peptides derived by proteolytic digestion with either trypsin or *S. aureus* V8 protease. From the intact molecule, the sequence of residues 2–31 was obtained. However, because of the linkage method used for the solid-phase sequencing, we were unable to identify the amino-terminal residue or positively determine the positions of lysines. In order to obtain the sequence for the remaining residues and eliminate ambiguities, further sequence analyses were performed in the gas phase with HPLC-purified proteolytic fragments derived from the intact molecule (Figure 1). As described earlier for bovine ubiquitin (Schlesinger et al., 1975), we find that oat ubiquitin is highly resistant to digestion with either trypsin or V8 protease. In the case of trypsin, complete digestion was accomplished only in the presence of 6 M urea (Cox et al., 1986). This treatment made ubiquitin sufficiently susceptible to proteolysis without totally eliminating the activity of trypsin. In the absence of urea, the only fragment cleaved even after prolonged incubations was the carboxyl-terminal Gly-Gly (S. M. Langan and R. D. Vierstra, unpublished). Under the digestion conditions used here, two trypsin-resistant bonds were identified at Lys-33 and Arg-74. As suggested by Cox et al. (1986), this protection may result from the presence of neighboring acidic residues in the former case and neighboring hydrophobic residues in the later. Digestion with V8 protease was accomplished only with high enzyme to substrate ratios and long incubation times (up to 72 h).

From analysis of the sequence data obtained here for oat ubiquitin and comparison with that obtained previously for the animal form, the complete amino acid sequence of oat ubiquitin was deduced (Figure 2). The exact position in the complete sequence of all the tryptic peptides except T1 was

[1] Abbreviations: HPLC, high-performance liquid chromatography; $NaDodSO_4$, sodium dodecyl sulfate.

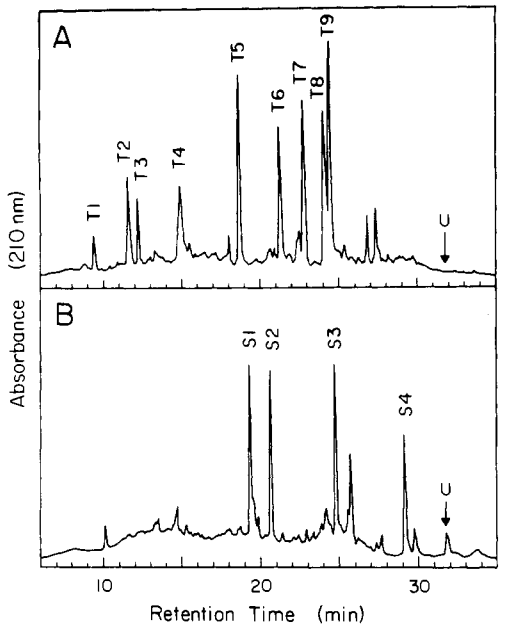

FIGURE 1: HPLC elution profile of *Avena* ubiquitin digested with either trypsin (A) or *S. aureus* V8 protease (B). Ubiquitin was digested to completion with either protease (4 h for trypsin and 72 h for V8 protease), and the resulting peptides were chromatographed by reverse-phase HPLC on a C18 column, monitored at 210 nm as described under Materials and Methods. Peptides, designated T1-T9 and S1-S4 for the trypsin and V8 protease digests, respectively, were subjected to repetitive Edman degradation and the derived amino acid sequences used to construct the complete amino acid sequence presented in Figure 2. U is the elution position of intact ubiquitin.

corroborated with sequence information from the intact protein or from overlapping V8 protease peptides. The position of T1 at the carboxy terminus was concluded from a comparison with the animal sequence and from previous immunological and biochemical evidence demonstrating that the two forms have homologous carboxy termini (Vierstra et al., 1985). The lack of overlapping sequence information between amino acids 72 and 73 does not preclude the possibility that additional residues are inserted in this region. However, we consider this possibility unlikely on the basis of the high degree of immunological cross-reactivity (Vierstra et al., 1985) and sequence homology (see above) in this region for all the ubiquitins analyzed to date.

Similar to the animal form, oat ubiquitin contains 76 amino acids, no tryptophan or cysteine, one tyrosine, a sole methionine as the amino terminus, and a carboxyl-terminal sequence of Leu-Arg-Gly-Gly. Even though previous amino acid compositional analyses indicated the presence of 1 mol of cysteine/mol of oat ubiquitin (Vierstra et al., 1985), we could not identify any cysteine residues in the protein by sequence determination. Since subsequent compositional analyses of more highly purified preparations than used previously yielded much lower molar ratios of cysteic acid, we conclude that the earlier observation that oat ubiquitin contains a cysteine residue was artifactual.

Despite strong homologies between the oat and animal ubiquitin, differences in amino acid sequence have been detected at positions 19 (Ser for Pro), 24 (Asp for Glu), and 57 (Ala for Ser). The oat sequence is also divergent from that obtained for yeast ubiquitin, the only other ubiquitin variant identified thus far (Ozkaynak et al., 1984). Amino acid substitutions between these two forms occur at positions 28 (Ala for Ser) and 57 (Ala for Ser). None of the changes between the animal and oat forms would affect the charge of the molecule as expected from previous observations that the p*I* values of human and oat ubiquitin are identical (Vierstra et al., 1985).

Discussion

Despite the sequence divergence between oat and human ubiquitin, three substitutions in 76 residues, ubiquitin still represents one of the most conserved proteins yet identified, rivaled only by histone 4, which has two substitutions in 102 residues between cattle and pea (Delange et al., 1969). All the substitutions identified can be accounted for by single nucleotide changes in the ubiquitin gene(s). This striking stability indicates that the structural constraints necessary for catalytic activity and/or recognition of ubiquitin by other proteins involved in its function(s) are very strict. On the basis of the three-dimensional structure of human ubiquitin, determined by X-ray diffraction (Vijay et al., 1985), we note that all the amino acid substitutions identified to date are clustered in a region opposite the carboxyl-terminal, conjugation site. Because sequence divergence can be tolerated in this region, it may indicate that this region is not directly involved in the protein's function(s). Whether these amino acid differences actually alter the three-dimensional structure of oat ubiquitin from that determined for the human form is currently under investigation. It is also interesting that the only two ubiquitin variants identified to date are from the plant kingdom, yeast and oats. Whether this sequence divergence is common to all plant species and hence may provide information on the evolution of the animal and plant kingdoms awaits further analyses.

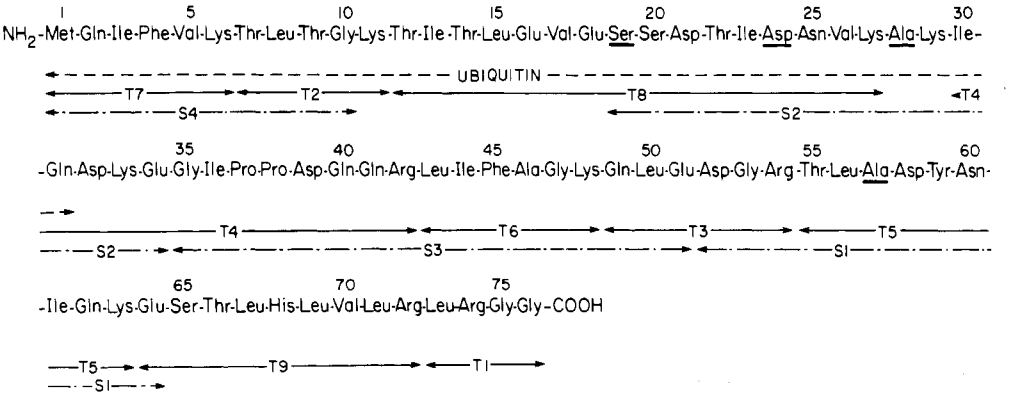

FIGURE 2: Complete amino acid sequence of *Avena* ubiquitin. Arrows indicate the sequence of amino acid residues identified by repetitive Edman degradation of the intact molecule (ubiquitin) or of peptides derived from ubiquitin by proteolytic digestion with either trypsin (T1-T9) or *S. aureus* V8 protease (S1-S4) (see Figure 1). The complete sequence was identified for all the proteolytically generated fragments except S4. The underlined residues are those that differ in the amino acid sequences determined for human and yeast ubiquitins (Schlesinger & Goldstein, 1975; Ozkaynak et al., 1984).

ACKNOWLEDGMENTS

We thank Dr. K. Wilkinson for making available his methods for trypsin digestion of ubiquitin and subsequent HPLC separation of the peptides. We are grateful to Drs. R. Niece and M. Sussman for assistance with the peptide purification and sequence analyses. Sequence determinations were performed in part by the Protein Sequence-DNA Synthesis Facility at the University of Wisconsin, supported by funds from the U.S. Public Health Service, National Institutes of Health (Shared Equipment Grant S10-RR01684, National Cancer Institute Continuing Support Grant CA-07175, and the general research support grant to the University of Wisconsin Medical School), and from the University of Wisconsin Graduate School.

Registry No. Ubiquitin, 60267-61-0; ubiquitin (oat), 101953-02-0.

REFERENCES

Bond, U., & Schlesinger, M. L. (1985) *Mol. Cell. Biol. 5*, 949–956.

Ciechanover, A., Finley, D., & Varshavsky, A. (1984) *Cell (Cambridge, Mass.) 37*, 57–66.

Cox, M. J., Shapira, R., & Wilkinson, K. D. (1986) *Anal. Biochem.* (in press).

Delange, R. J., Fambrough, D. M., Smith, E. K., & Bonner, J. (1969) *J. Biol. Chem. 244*, 5669–5679.

Dworkin-Rastl, E., Shrutkowski, A., & Dworkin, M. B. (1984) *Cell (Cambridge, Mass.) 39*, 321–325.

Finley, D., & Varshavsky, A. (1985) *Trends Biochem. Sci. (Pers. Ed.) 117*, 344–347.

Gavilanes, J. G., Gonzalez de Buitrago, G., Perez-Castells, R., & Rodriquez, R. (1982) *J. Biol. Chem. 257*, 10267–10270.

Goldknopf, I. L., & Busch, H. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 864–868.

Goldknopf, I. L., Wilson, G., Ballal, N. R., & Busch, H. (1980) *J. Biol. Chem. 255*, 10555–10558.

Goldstein, G., Scheid, M., Hammerling, U., Boyse, E. A., Schlesinger, D. A., & Niall, H. D. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 11–15.

Hershko, A., & Ciechanover, A. (1982) *Annu. Rev. Biochem. 51*, 335–364.

Hershko, A., Leshinsky, E., Ganoth, D., & Heller, H. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1619–1623.

Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 3506–3509.

Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., & Hood, L. E. (1983) *Methods Enzymol. 91*, 399–413.

Levinger, L., & Varshavsky, A. (1982) *Cell (Cambridge, Mass.) 28*, 375–385.

L'Italien, J. J., & Strickler, J. E. (1982) *Anal. Biochem. 127*, 198–212.

Matsui, S., Seon, B. K., & Sandberg, A. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6386–6390.

Mueller, R. D., Yasuda, H., Hatch, C. L., Bonner, W. M., & Bradbury, E. M. (1985) *J. Biol. Chem. 260*, 5147–5153.

Ozkaynak, E. Finley, D., & Varshavsky, A. (1984) *Nature (London) 312*, 663–666.

Schlesinger, D. H., & Goldstein, G. (1975) *Nature (London) 255*, 423–424.

Schlesinger, D. H., Goldstein, G., & Niall, H. D. (1975) *Biochemistry 14*, 2214–2218.

Vierstra, R. D., Langan, S. M., & Haas, A. L. (1985) *J. Biol. Chem. 260*, 12015–12021.

Vijay-Kumar, S., Bugg, C. E., Wilkinson, K. D., & Cook, W. J. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 3582–3585.

Watson, D. C., Levy, B., & Dixon, D. G. (1978) *Nature (London) 276*, 196–198.

Wu, R. S., Kohn, K. W., & Bonner, W. M. (1981) *J. Biol. Chem. 256*, 5916–5920.