Characterization of a ubiquitin-fusion gene from the tobacco hawkmoth, *Manduca sexta*

Stephen T. Bishoff* and Lawrence M. Schwartz

Department of Zoology, Morrill Science Center, University of Massachusetts, Amherst, MA 01003, USA

Received June 26, 1990; Revised and Accepted September 18, 1990

EMBL accession no X53524

ABSTRACT

A gene encoding a ubiquitin-fusion protein was isolated from a cDNA library made from the intersegmental muscles (ISM) of the moth, *Manduca sexta*. The predicted amino acid sequence of this fusion protein is highly conserved when compared to the sequence of homologous proteins from diverse species. The *Manduca* clone encoding this ubiquitin fusion gene hybridized with a single, abundantly expressed transcript in all tissues examined. In the ISM, the transcript was present at high levels, independent of the developmental stage or hormonal treatment of these muscles. Data from other species indicate that ubiquitin-fusion genes participate in ribosome biogenesis.

INTRODUCTION

Ubiquitin is a highly conserved, 76-amino acid protein found in all eukaryotic cells examined to date (1). Ubiquitin is covalently linked to many proteins within the cytoplasm as a post-translation modification. When multiple copies of ubiquitin are added to a single protein that targets the acceptor protein for extra-lysosomal proteolysis (2). However, many proteins in the cell have only a single ubiquitin attached to them; examples include histones H2a and H2b (3), and one of the actin proteins found in *Drosophila* flight muscle (4). While the role of mono-ubiquitin conjugation in these systems remains unclear, it does not appear to target those proteins for degradation.

Within the genome, ubiquitin is encoded by two distinct classes of genes. The first contains tandem head-to-tail repeats of the sequence encoding ubiquitin. These polyubiquitin genes provide a large pool of free ubiquitin monomers which are required for the targeting of proteins for degradation during the cellular stress response (5). The second class of ubiquitin genes encodes a single copy of the ubiquitin protein, fused in-frame with a second protein coding sequence on the 3'-end. These fusion proteins are either 52 (Ubf52) (6,7,8,9,10) or 76–80 (Ubf80) (7,8,11,12) amino acids long. The products of these ubiquitin-fusion genes are required for ribosome biogenesis, since deletions in the Ubf52 or Ubf80 sequence result in dramatic reductions in the number of large or small ribosomal subunits, respectively (13).

We have been characterizing the ubiquitin gene family in the

tobacco hawkmoth, *Manduca sexta*, since ubiquitin appears to play a role in the programmed degeneration of the intersegmental muscles (ISM) (14,15). The ISM function to extricate the developing adult moth from the pupal cuticle at the end of metamorphosis. Following eclosion (emergence), the ISM are no longer needed by the animal and undergo rapid degeneration during a 36 hour period (16,17). The trigger for ISM death in *Manduca* is a decline in the insect molting hormone, 20-hydroxyecdysone (20-HE) (17). Late on day 17 of pupal development, the steroid titer falls below a specific threshold value, thus committing the ISM to degenerate following eclosion, late on day 18. Coincident with this commitment is the selective repression and expression of ecdysteroid-responsive genes (14,18). One gene whose expression increases dramatically is polyubiquitin (19). In characterizing the ubiquitin gene family in *Manduca*, we isolated a cDNA clone for a ubiquitin-fusion protein, clone 18–2. Here we present the nucleotide and predicted amino acid sequences for the ubiquitin-fusion protein encoded within clone 18–2, and characterize its pattern of expression in *Manduca*.

MATERIALS AND METHODS

Animals

Eggs of *Manduca sexta* were the gift of Dr. Jim Buckner of the USDA \ ARS Biosciences Research Lab, Fargo, ND. Animals were reared on an artificial diet as previously described (20). Comparably sized animals were staged by distinct cuticular markers (17). When hormonal effects were examined, pupae were injected in the dorsal thorax with 25 μ g of 20-HE (Sigma).

RNA Isolation and Northern Blots

At the specified stages of development, insects were rapidly dissected under ice-cold saline (21), and the ISM placed in liquid nitrogen. Muscles were homogenized with a TissueMizer (Tekmar Co.) in 4M guanidinium isothiocyanate extraction buffer (22). Total RNA was isolated by centrifugation through a 5.7 M CsCl pad in an SW41 rotor (Beckman) for 23 hours at 30k RPM at 18°C. The pellet was resuspended in 5% beta-mercaptoethanol, 0.5% Sarkosyl and 5 mM EDTA, then extracted with phenol/chloroform, and precipitated from 0.3 M sodium acetate by the addition of cold ethanol.

* To whom correspondence should be addressed

For Northern blots, total RNA (15 μ g, determined spectrophotometrically) from each developmental stage was denatured in formaldehyde, separated electrophoretically by size in an agarose (1.5%)/formaldehyde (2.2 M) gel, and stained with ethidium bromide to insure equal loading. The RNA was transferred to Zeta-Probe membrane (Bio-Rad). Blots were hybridized overnight in buffer (1 mM EDTA, 0.5 M NaH_2PO_4 , pH 7.2, 7% SDS) with ^{32}P -labeled probe and washed (1 mM EDTA, 40 mM NaH_2PO_4 , pH 7.2, 5% SDS then 1% SDS) under stringent conditions (65°C) according to the manufacturer's recommendations. Hybridization was detected by autoradiography on Kodak XAR film with intensifying screens at -80°C.

DNA Isolation and Southern Blots

Genomic DNA was prepared from whole larvae after removal of the gut to prevent bacterial contamination. The tissue was frozen in liquid nitrogen, pulverized, and digested overnight at 50°C with 0.1 mg/ml of protease K in digestion buffer (100 mM NaCl, 10 mM Tris-HCl, pH 8, 25 mM EDTA, pH 8, and 0.5% SDS) (23). The sample was digested with RNase A, then extracted with phenol/chloroform, and precipitated with ethanol as described above. For Southern blots (24), 15 μ g aliquots of the genomic DNA were digested with restriction endonucleases (Promega), fractionated in agarose (0.7%), and blotted onto nylon membrane (MSI), then hybridized in buffer (50% formamide, 5 $\times$ Denhardt's solution, 5 $\times$ SSPE, 0.1% SDS, 200 μ g/ml tRNA) and ^{32}P -labeled probe overnight at 50°C. The membranes were washed with buffer (0.1% SDS, 1 $\times$ SSPE then 0.1 $\times$ SSPE) under stringent conditions (65°C). Hybridization was detected by autoradiography on Kodak XAR film with intensifying screens at -80°C.

cDNA Clones

Clone 18-2 was obtained as previously described (14) by differential hybridizations with radiolabeled cDNAs from pre-commitment to degenerate (day 15) and post-commitment (day 18) ISM mRNA. Clone 18-2 was sub-cloned into Bluescript KS+ (Stratagene). In addition, a portion of clone 18-2, 18-2a, which was comprised of the 127 bp of the 3' end of the open reading frame, the untranslated region, and the poly A⁺ tail, were excised from clone 18-2 with *Sau* 3A1 and *Eco* R1, and sub-cloned into pGEMEX 1 (Promega).

Nucleotide Sequence Analysis

The nucleotide sequence of clone 18-2 was determined by the dideoxy method (25) using T7 DNA polymerase (Pharmacia or US Biochemical). Autoradiographs were read using an IBI gel reader; the resulting nucleotide sequence was analyzed on a microVax computer, using the University of Wisconsin Genetics Computer Group Sequence Analysis Software, version 6.1 (26). The nucleotide sequence data reported in this paper will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under the accession number X53524.

RESULTS

Sequence Analysis of clone 18-2

Clone 18-2 was isolated during the screening of a cDNA library made from *Manduca* ISM poly A⁺ RNA (14). Clone 18-2 was sequenced to identify its putative protein product. The insert contained within 18-2, which was not full length (see Northern

Blots below) consisted of 550 nucleotides plus a polyadenylated tail. At the 5' end of the clone there is a small region of untranslated sequence upstream of the predicted AUG initiation codon (Figure 1A). Translation of the nucleotide sequence predicts a single open reading frame (ORF) encoding a protein with 155 amino acids. This ORF is followed by a stop codon (TAG), a 79 base untranslated region, and a polyadenylated tail.

The protein predicted from the nucleotide sequence has a calculated molecular weight of 17,847 daltons and an isoelectric point of 10.53. The N-terminus contains the complete amino acid sequence for a single copy of the ubiquitin protein. Ubiquitin

Figure 1A. Nucleotide sequence of clone 18-2 with the predicted amino acid sequence for the ORF. The start of translation (bold type) is labeled base 1. Ubiquitin is encoded from nucleotides 1-228; Ubf80 from 229-465 (start at filled arrow); clone 18-2a from 339 (open arrow 1) to the *Eco* R1 site following the polyadenylation tail (open arrow 2). The cleavage-polyadenylation signal (underlined) is found in the 3'-untranslated region. The untranslated sequence is in lower case. **Figure 1B.** Diagram of the 18-2 clone containing ubiquitin (stippled) and the fusion protein (open). The fusion protein contains a highly basic transitional region and a zinc finger motif with the cysteines shown (broken lines).

has 76 amino acids, a molecular weight of 8,564 daltons, and an isoelectric point of 7.53. In frame with the ubiquitin, is a C-terminal extension comprising an additional 79 amino acids (9,301 daltons). This fusion element, Ubf80, has several distinguishing features (Figure 1B). 35% of the amino acid residues of Ubf80 are basic; at the N-terminus, five of the first seven amino acids are lysines. This preponderance of basic residues results in a pI of 10.76 for Ubf80. Towards the C-terminal end of the protein, there is a zinc-finger motif of the form, Cys-X₂₋₄-Cys-X₂₋₁₅-Cys-X₂₋₄-Cys (27).

Ubiquitin encoded by *Manduca* clone 18-2 is identical to the amino acid sequence of ubiquitin from *Drosophila* and humans. When the fusion element alone is compared to the homologous region from ubiquitin-fusion genes from four other species, a 63.3% homology is found across taxa (Figure 2). All reported fusion elements contain both the highly basic transitional domain, as well as the identical zinc-finger motif form: Cys-X₄-Cys-X₁₄-Cys-X₂-Cys. As shown by the shading in Figure 2, several regions of the sequence are invariant.

Southern Analysis

In all eukaryotic species examined to date, ubiquitin is represented in the genome as a multi-gene family (5). To determine if 18-2 is present as a single copy gene, genomic Southern blots were probed with clone 18-2 under high stringency conditions (Figure 3). Digestion with *Eco* R1, *Hind* III, and *Pst* I generate single hybridizing bands, while *Sal* I generated two. Since none of these enzymes have restriction sites within the 18-2 clone, these data suggest that 18-2 exists as a single copy gene in *Manduca*. The second band in the *Sal* I lane may reflect hybridization to another ubiquitin sequence despite high stringency conditions, a cut within an intron, or incomplete digestion.

Pattern of ubiquitin-fusion expression

To determine the pattern of expression for 18-2, RNA was isolated from the ISM at various developmental stages, beginning before the onset of muscle atrophy (day 14) and continuing into the active phase of the degeneration program (5hr post-eclosion). One group of animals was injected with 20-HE on day 17 to block degeneration (17), and ISM RNA isolated 5 hours after the normal time of eclosion. For each stage and treatment, 15 µg of total RNA were fractionated by size in agarose, blotted, and hybridized to ³²P-labeled 18-2 DNA. The autoradiograph revealed a single transcript of 0.9 kb, which is expressed at a high level relative to total RNA (Figure 4A). This level is attained independent of developmental stage or hormonal treatment.

When the stringency of hybridization was reduced by decreasing the wash temperature to 60°C, the pattern on Northern blots became much more complex (Figure 4B). In addition to the constitutively expressed 0.9 kb transcript, several larger transcripts appeared whose abundance was strongly influenced by developmental stage or hormonal manipulation. These transcripts are expressed at low levels on days 15 and 16, and at high levels at all time points examined on or after day 18. Pre-injection with 20-HE dramatically reduced the accumulation of these larger transcripts. Several lines of data, including DNA sequence analysis, suggest that these additional transcripts encode polyubiquitin (14,15).

Since the ISM contains multiple ubiquitin encoding transcripts, we wanted to verify that the 0.9 kb transcript was, in fact, the product of the 18-2 gene, rather than another ubiquitin encoding transcript. Membranes identical to the one in Figure 4B were

hybridized to clone 18-2a which contains only part of the 3' end of 18-2 (Figure 1A). This probe labeled only the constitutively expressed transcript seen in Figure 4A (data not shown) indicating that the 18-2 probe, at high stringency, specifically recognized the fusion gene product.

Since this ubiquitin-fusion gene was constitutively expressed at high level in the ISM, we wanted to determine its distribution in other tissues from *Manduca*. Day 18 animals were dissected, and the following tissues taken for RNA isolation: Malpighian

Figure 2. Alignment of the amino acid sequence of the Ubf80 proteins from *Manduca sexta*, *Drosophila melanogaster* (12), *Homo sapiens* (11), *Saccharomyces cerevisiae* (7), and *Dictyostelium discoideum* (8). Regions of sequence identical to *Manduca* are stippled.

Figure 3. Autoradiograph of a Southern blot of genomic DNA from *Manduca sexta*. The blot was hybridized to labeled clone 18-2 DNA and washed under high stringency conditions (65°C). Lanes: 1 *Eco* R1; 2 *Hind* III; 3 *Pst* I; 4 and *Sal* I. Size is indicated in kilobases (kb).

tubules, fat body, flight muscle, ISM, male accessory gland and ovary. For each sample, 15 μ g of total RNA were fractionated, blotted, and hybridized with the 18-2 probe. As can be seen in Figure 5, the probe labeled a single, abundantly expressed transcript of 0.9 Kb in every sample. The level of expression varied relative to the total RNA with each tissue; the greatest amount was in the ovary, followed by ISM, fat body, and male accessory gland. Malpighian tubules and flight muscle had the lowest levels. After stripping the 18-2 probe, the blots were re-probed with 18-2a, which gave identical results (data not shown).

Figure 4A. Autoradiograph of a Northern blot of *Manduca* ISM. Total RNA (15 μ g/lane) was isolated from muscles on pupal days 14-18, 3 hr and 5 hr post-eclosion, and 5 hr post-eclosion from animals treated with 25 μ g of 20-HE on day 17. The membrane was hybridized with labeled 18-2 DNA and washed under high stringency conditions (65°C). **Figure 4B.** Autoradiograph of a Northern blot identical to that in Figure 4A. This membrane was hybridized with labeled 18-2 DNA but washed at lower stringency (60°C).

Figure 5. Autoradiograph of a Northern blot of *Manduca* tissues taken on pupal day 18. Lanes: 1 Malpighian tubules; 2 fat body; 3 flight muscle; 4 ISM; 5 male accessory gland; and 6 ovary. The membrane was probed with labeled 18-2 DNA under high stringency conditions (65°C).

DISCUSSION

The data presented here demonstrate that clone 18-2 encodes a ubiquitin-fusion protein, which is abundantly expressed in all tissues examined. In the ISM, expression of the ubiquitin-fusion gene which contains Ubf80 is independent of developmental stage or 20-HE treatment. This is in sharp contrast to the polyubiquitin gene, which undergoes dramatic changes in its expression during development and in response to steroid treatment (15).

Both ubiquitin-fusion genes of yeast, containing Ubf52 or Ubf80, participate in ribosome biogenesis (13). By size, form of the Zn²⁺ finger motif, and sequence homology, we conclude that clone 18-2 encodes the larger member of the ubiquitin-fusions, which contains Ubf80. We have not yet detected a ubiquitin-fusion sequence containing Ubf52 on our Northern blots, presumably due to the lack of sequence homology among the regions encoding ubiquitin, which differ significantly in *Manduca* (28). The Southern blot data also suggest that an 18-2 probe will not detect other ubiquitin genes under high stringency conditions.

Ubf80 encoded by 18-2 shares 63.3% identity with the Ubf80 amino acid sequences from diverse organisms. Retention of the stretch of basic residues at the N-terminus of the fusion moiety and the zinc finger motif near the C-terminus, suggest that these regions are required for homologous function among species. Both motifs are thought to function in binding to nucleic acids such as rRNA, which is consistent with a role in ribosome biogenesis (13).

The ubiquitin-fusion genes containing Ubf80 are required for production of the small ribosomal subunit. Deletion of the Ubf80 sequence in the ubiquitin-fusion gene in yeast drastically reduces production of the small ribosomal subunit (13) while re-introduction of the wild-type sequence fully complements the deletion (13). In addition, an antibody directed against the N-terminus of the fusion element stains the rat ribosomal protein S27a on 2-dimensional western blots (29,30). While we have not determined the role of 18-2 in *Manduca*, the high levels of 18-2 expression suggest that it serves house-keeping function. Furthermore, the ovaries, whose trophocytes are enriched for ribosome sequences (31), have the greatest level of the 18-2 transcript.

Why these putative ribosomal proteins are produced as fusions to ubiquitin is still unclear. Ubiquitin may serve either to extend the half-life of the fusion element or to function as a chaperonin (13). Independent of its role, it appears that the ubiquitin moiety is cleaved from the fusion element at some point following translation (32). The fusion element then participates in ribosome biogenesis, and the ubiquitin enters the pool of free ubiquitin monomers in the cytoplasm. Ubiquitin derived from the fusion proteins may be the major source of free ubiquitin in the cell during non-stress conditions (13).

We are continuing our examination of the ubiquitin-fusion genes in *Manduca*. In particular, we are interested in isolating the promoter for the ubiquitin-fusion gene containing Ubf80 since it may provide a strong, non-tissue specific promoter for transfection experiments.

ACKNOWLEDGEMENTS

We wish to thank J. Buckner for the provision of *Manduca* eggs, L. Kosz for technical assistance and S. Robinson, A. Myer, and

B. Osborne for a critical reading of the manuscript. Supported by N.I.H. grant GM40458 to LMS.

REFERENCES

- Schlesinger, M.J., and Bond, U. (1987) Oxford Surveys on Eukaryotic Genes 4, 77–91.
- Chau, V., Tobias, J.W., Bachmair, A., Marriott, D., Ecker, D.J., Gonda, D.K., and Varshavsky, A. (1989) Science 243, 1576–1583.
- Bonner, W.M., Hatch, C.L., and Wu, R.S. (1988) In Rechsteiner, M. (ed.), Ubiquitin. Plenum Press, New York, 157–172.
- Ball, E., Karlik, C.C., Beall, C.J., Saville, D.L., Sparrow, J.C., Bullard, B., Fyrberg, E.A. (1987) Cell 51, 221–228.
- Finley, D., Ozkaynak, E., and Varshavsky, A. (1987) Cell 48, 1035–1046.
- Salvensen, G., Lloyd, C., and Farley, D. (1987) Nucleic Acids Res. 15, 5485.
- Ozkaynak, E., Finley, D., Solomon, M.J., and Varshavsky, A. (1987) EMBO J. 6, 1429–1439.
- Ohmachi, T., Giorda, R., Show, D.R., and Ennis, H.L. (1989) Biochem. 28, 5226–5231.
- Swindle, J., Ajioka, J., Eisen, H., Sanwal, B., Jacquemot, C., Browder, Z., and Buck, G. (1988) EMBO J. 7, 1121–1127.
- Mezquita, J., Pau, M., and Mezquita, C. (1988) Nucleic Acids Res. 16, 11838.
- Lund, P.K., Moats-Staats, B.M., Simmons, J.G., Hoyt, E., D'Ercole, A.J., Martin, F., and Van Wyk, J.J. (1985) J. Biol. Chem. 260, 7609–7613.
- Lee, H., Simon, J.A., and Lis, J.T. (1988) Mol. Cell. Biol. 8, 4727–4735.
- Finley, D., Bartel, B., and Varshavsky, A. (1989) Nature 338, 394–401.
- Schwartz, L.M., Kosz, L., and Kay, B.K. (1990) Proc. Natl. Acad. Sci. USA in press.
- Myer, A., and Schwartz, L.M., in preparation.
- Schwartz, L.M. and Truman, J.W. (1982) Science 215:1420–1421.
- Schwartz, L.M. and Truman, J.W. (1983) Dev. Biol. 99, 103–114.
- Schwartz, L.M., and Kosz, L. In preparation.
- Schwartz, L.M., Myer, A., Kosz, L., Engelstein, M., and Maier, C. submitted.
- Bell, R.A., and Joachim, R.A. (1976) Ann. Ent. Soc. Amer. 69, 365–373.
- Schwartz, L.M. (1981) J. Insect Physiol. 27, 727–734.
- MacDonald, R.J., Swift, G.H., Przybyla, A.E., and Chirgwin, J.M. (1987) Meth. Enzymol. 152, 219–227.
- Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1987) Current Protocols in Molecular Biology. John Wiley & Sons, New York.
- Southern, E. (1975) J. Mol. Biol. 98, 503–517.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- Deveraux, J., Haeblerli, P., and Smithies, O. (1984) Nuc. Acids Res. 12, 387–395.
- Berg, J.M. (1986) Science 232, 485–487.
- Myer, A., Bishoff, S.T., and Schwartz, L.M. unpublished.
- Redman, K.L. and Rechsteiner, M. (1988) J. Biol. Chem. 263, 4926–4931.
- Redman, K.L. and Rechsteiner, M. (1988) Nature 338, 438–440.
- Berry, S.J. (1985) In Blum, M.S. (ed.), Fundamentals of Insect Physiology. John Wiley & Sons, New York. p. 437–466.
- Monia, B.P., Ecker, D.J., Jonnalagadda, S., Marsh, J., Gotlib, L., Butt, T.R., and Crook, S.T. (1989) J. Biol. Chem. 264, 4093–4103.