Volume 17 Number 8 1989 Nucleic Acids Research

Polyadenylated H3 histone transcripts and H3 histone variants in alfalfa

Sheng-Cheng Wu, János Györgyey and Dénes Dudits*

Institute of Genetics, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, H-6701 Szeged, Hungary

Received December 22, 1988; Revised and Accepted March 9, 1989 EMBL accession nos X13673–X13677

ABSTRACT

Histone H3 mRNAs were found in polyA(+) fractions of total RNA prepared from alfalfa plants, calli and somatic embryos. The sequence analysis of cDNAs revealed the presence of a polyA tail on independent alfalfa H3 mRNAs. A highly conserved sequence motif AAUGAAA identified about 20bp upstream from the 3′ ends of the alfalfa H3 cDNAs was suggested to be one of the possible regulatory elements in the 3′ end formation and polyadenylation. Three out of the four analysed H3 cDNAs have more than 97% homology with a genomic clone and encode the same protein. While the fourth represents a minor species with only 78.8% homology to the coding region of the genomic clone and encodes a H3 histone with four amino acid replacements. On the basis of compilation analysis we suggest a consensus sequence for plant H3 histones which differs from that of animal's by four amino acid changes.

INTRODUCTION

The common features of most animal histone genes are the absence of introns and polyadenylation signal sequences and the presence of a T-hyphenated palindromic structure and its 'downstream element' in the 3′ untranslated region (UTR, reviews in 1, 2). The latter has been demonstrated to be essential for the regulation of 3′ processing of histone pre-mRNAs (3,4,5). Recently, a number of histone variants and their corresponding genes have been discovered (review in 6). Unlike the classical histone genes initially studied, these variant histone genes frequently contain introns and encode polyadenylated mRNA (7,8). It has been suggested that the polyadenylation of histone mRNAs is an archaic trace (9) and this feature may have particular importance in evolution. In contrast to these minor cases in animals, histone mRNAs are ubiquitously polyadenylated in yeast (10) and *Tetrahymena* (11). Recent studies on plant histone genes have also shown that histone H3 and H4 mRNAs are polyadenylated in maize (12), *Arabidopsis* and probably tobacco and sunflower (13). Consistently, a pea histone H1 (14) and a barley histone H3 (15) mRNAs were also shown to be polyadenylated. A comparison revealed that the 3′ UTRs of plant histone genes are much longer than that of animal's and do not contain the typical T-hyphenated hairpin structure (review in 16). These structural characteristics may be responsible for a different mechanism(s) of the 3′ processing in plant histone pre-mRNAs. At present, however, the data in the literature are insufficient to give a general picture about plant histone genes, particularly in relation to the status of polyadenylation of their mRNAs. Here we report data which confirm that histone H3 mRNAs are polyadenylated in alfalfa plants, calli and somatic embryos. We also outline a consensus amino acid sequence for plant H3 histones through the analysis of different alfalfa H3 cDNAs and other plant H3 proteins or H3 genes.

MATERIAL AND METHODS

Plant Source

Plants of *Medicago sativa* L cv. Regen S, line RA3, obtained from D.A.Stuart were propagated by nodal cutting. Initiation of callus cultures and induction of somatic embryos was carried out according to Stuart et al (17).

RNA Preparation and Northern Hybridization

Total cellular RNAs were prepared from in vitro grown alfalfa plants, calli and somatic embryos as described by Cathala, et al (18). PolyA(+) RNAs and polyA(−) RNA fractions were obtained by oligo-dT cellulose column chromatography (19). Northern hybridizations were performed as we described previously (20).

Construction of Alfalfa cDNA Library

RA3 calli grown on SH medium (21) with low concentration of plant hormone (17) were exposed to high concentration of 2,4-dichlorophenoxyacetic acid (2,4-D, 10mg/l) for 60 hours. This hormone treatment is essential for the induction of somatic embryos (17). By use of the polyA(+) RNA from the 2,4-D treated calli, a cDNA library was constructed (J.G. et al, unpubl.) according to Gubler and Hoffman in an *in vitro* transcription vector, pGEM-2 (22).

DNA Sequencing

DNA fragments were subcloned into M13 mp18/19 vectors and sequenced by the dideoxy chain termination method (23).

RESULTS AND DISCUSSION

Alfalfa H3 Histone Transcripts Are Polyadenylated

Polyadenylation of histone H3 and H4 mRNAs has been shown in germinating embryos and young plantlets of maize (12) and cell suspension and plantlets of *Arabidopsis* (13). Our previous observation of the unusually high molecular weight of alfalfa histone H3 mRNAs (20) also suggested the presence of long 3′ UTR in alfalfa H3 mRNAs possibly with polyadenylate tails. To test this prediction, polyA(+) and polyA(−) mRNAs were isolated by oligo-dT cellulose chromatography from alfalfa plants, calli and somatic embryos. The amount of H3 transcripts in polyA(+) and polyA(−) RNA fractions was analysed by RNA blot hybridizaton (Fig.1). In comparison to the non-separated total RNA fractions, the signal detected in the polyA(+) RNA fractions was stronger, whereas no detectable amount of H3 mRNA was observed in the polyA(−) RNA fractions. These results indicated that histone H3 mRNAs are polyadenylated in the studied alfalfa tissues. The hybridization signals shown in Figure 1 are also consistent with our previous findings indicating that the histone H3 genes are highly expressed in somatic embryos (20). According to the present molecular weight determination, the main H3 transcripts are shorter (780bp) than we suggested previously (20). In attempt to prove unequivocally that alfalfa H3 transcripts carry polyA tail cDNA clones were analysed. About 40,000 cDNA clones from an alfalfa cDNA library were screened with an alfalfa H3 gene, ALH3-1.1 (20), as a probe. Out of 25 positive clones four were further characterized by DNA sequencing. Figure 2 shows the nucleotide sequence alignment of the genomic H3 clone, ALH3-1.1, and the four H3 cDNAs. Among the isolated cDNAs, only clone pH3c-1 carried a full length H3 transcript, while the others were incomplete copies of the H3 mRNAs. All the analysed cDNAs possess long 3′ UTR (146bp or 217bp, Fig.2). The detection of polyA tract in the studied cDNA clones (pH3c-1, pH3c-11 and pH3c-12) is in agreement with

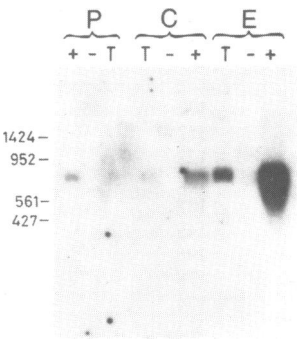

Fig.1. Northern blot hybridization analysis of total RNA (T), polyA(−) RNA (−) and polyA(+) RNA (+). RNAs were prepared from 16 days old alfalfa plants propagated by nodal cutting (P), three week old calli (C) and somatic embryos with different stages (E). 25μg of total and polyA(−) RNA and 5μg of polyA(+) RNA were immobilized on nitrocellulose filter and hybridized to the EcoRV-StyI fragment of an alfalfa genomic clone, ALH3-1.1 (20). The polyA(+) RNA samples were loaded about four times in excess, assuming an appropriate 5% of polyA(+) RNA in the total RNA.

```
          10        20        30        40        50        60        70        80        90       100       110       120
1,  CTATATAACTCACCCCTTTCATCTCTTCTTCCTCATCACTCTCAATTCTTCAAAGCACAAAAACAACCTTCAAGTTTCTCTGTTTGATACTGTTCTTTTCAGTTATTTTTCATAATCAAT
2,                                                   ...C..............AC..T....C.A.....C...T.................GC.....**
3,                                                   ------------------------------------------------------------------
4,                                                   ------------------------------------------------------------------
5,                                                   ------------------------------------------------------------------

          130       140       150       160       170       180       190       200       210       220       230       240
1,  GGCACGTACCAAGCAAACCGCTCGCAAATCCACCGGTGGCAAAGCTCCAAGGAAGCAACTCGCCACAAAAGCCGCTCGCAAATCTGCTCCGGCCACCGGCGGAGTGAAGAAACCTCACCG
2,  *.....
3,  ------------------------------------------------------------------------------------------------------------------------
4,  ----------------------------------------.....
5,  ----------------------------------------.....G........C..G..T...A.G.....TA.T..T..A.....C.....

          250       260       270       280       290       300       310       320       330       340       350       360
1,  TTTCAGGCCAGGAACCGTCGCTCTCCGTGAGATCCGCAAGTACCAGAAGAGCACTGAGCTCCTCATCCGCAAACTCCCCTTCCAGCGTCTTGTCCGTGAGATCGCTCAGGATTTCAAGAC
2,  ........................................................................................................................
3,  ------------------------------------------------------------------------------------------------------------------------
4,  .........T..............................................................................................................
5,  A.A.C.C..T.....T.........T.............T.....T..C.....TT.G.........G..T..A..T.....................A..T.....A.....

          370       380       390       400       410       420       430       440       450       460       470       480
1,  TGATCTCCGTTTCCAGAGCTCCGCCGTGTCGGCTTTGCAAGAAGCGGCCGAGGCTTATCTCGTCGGTCTCTTTGAGGATACTAACCTCTGCGCCATTCATGCTAAGCGTGTCACTATCAT
2,  ..............................AC..........................................T.....................................
3,  ..............................AC..........................................C.....................................
4,  ..............................AC..........................................C.....................................
5,  G.....GA.A.........CAT..A..TCTT..AC.T..G.....A..T.........C..G..T..AT.G.........C..C..T..G..T..A.........C...A.G.TG..A..T..

          490       500       510       520       530       540       550       560       570       580       590       600
1,  GCCTAAGGATATCCAGCTCGCTAGGCGTATCCGTGGCGAGCGTGCTTGATCTTGTTGATTCGCTTTGTTAGGGTTTGTGTAGATAGGTTCATGATGTAGTTAAATCACAAACCGTTGCTA
2,  ...C.....................................****.....................................A.....................................
3,  ........................................................................................T.....................
4,  ........................................................................................T.....................
5,  .........C..T.....T...C.T..C..T..C..T..A.......AGGG.G..GA.GG......TAGC.TTA.G...G.TTAGTA..TTG..A.G.T...GGGTTTT.TGAA...AAT

          610       620       630       640       650       660       670
1,  TAAGTTTCTCTATGGATTTTGTTATATTGTAATGTGCTTAACGCTTAATCAATGAAATCGATCATCTTTTGTTA
2,  ..............C.....................................................AAAAAAAAAAAAAA
3,  ..............C.....................................TA............TAAAAAAAAAAAAAA
4,  ..............C.....................................TA............----
5,  .TTC...TATGCGTTGTATAGTTCTGAACCTATAATGTTCAATCTTTAACAACAGACATATTTTGGATTATGATTAGTTTTTTGCGGACAAATTTGTGATGTAATTGGTCAATTACAATT

5,  GAAGTCTCTGCAACTATTTTACTTATAAAAAAAAAAAAAA
```

Fig.2. Comparison of the nucleotide sequence of alfalfa histone H3 cDNAs with that of genomic clone, ALH3-1.1. The TATA box is framed, the cap site is indicated by dashed line, the initiation and stop codons are marked with asterisks and the putative polyadenylation signals are underlined. Dots represent base pairs identical to the genomic ones and dashes represent deletions. The polyA tracts in three cDNAs were represented by 15 A sequences. 1, ALH3-1.1; 2, pH3c-1; 3, pH3c-12; 4, pH3c-3 and 5, pH3c-11

Table 1. Putative polyadenylation signal sequences of plant histone H3 and H4 genes

Species	Clone	bp	Signal*	Reference
Alfalfa	ALH3-1.1	122	AAUG-AAA	This paper
	pH3c-1	122	AAUG-AAA	
	pH3c-3	122	AAUG-AAA	
	pH3c-12	122	AAUG-AAA	
	pH3c-11	188	AAUUG-AA	
Arabidopsis	H4748	149	G-UUGAAA	13
	H4777	196	GAUG-AAA	
Corn	H3C2	178	AAUGGAAA	12
	H3C3	134	AAUGGAAA	
	H3C4	215	AAUGGAAA	
	H4C7	194	GAUG-AAA	
	H4C13	194	GAUG-AAA	
Barley	–	237	GAUG--AA	15
Consensus sequence			A/GAUG(G)AAA	

* Base pairs from stop codon.

the hybridization data shown in Figure 1. We might suggest that the H3 cDNA in plasmid pH3c-3 originally also possessed a polyA tail which could be lost during cloning.

Polyadenylation Signals in Alfalfa H3 Histone Genes

About 20bp upstream from the 3′ ends of the alfalfa H3 transcripts a highly conserved sequence motif, AAUGAAA, was found (Fig.2). This motif contains one G insertion as compared to the conventional polyadenylation signal, AAUAAA. The latter is usually located 10−30bp upstream of the 3′ end of RNA polymerase II-transcribed messengers and, together with a stretch of GU or U-rich sequence, is essential for the cleavage and polyadenylation of the 3′ end (see reviews in 3,24). Most of the histone polyA(+) messenger RNAs including those from higher plants lack a typical polyadenylation signal.

This structural feature may not necessarily result in a different mechanism for the cleavage-polyadenylation reaction, since many divergent types of AAUAAA have been reported (24) and the pre-messenger RNAs are occassionally found to be cleaved and polyadenylated at multiple sites *in vivo* or *in vitro* (25,26). Similar forms of this putative polyadenylation signal detected in alfalfa H3 genes exist also in histone H3 and H4 genes of maize and *Arabidopsis* (Table 1), which have recently been shown to be polyadenylated (12,13). Coincidently, AAUGAA or AAUGGAAA sequences were also suggested to be the polyadenylation signal in a Cab gene of *Petunia* (25) and in a rbcS gene of pea (27). In addition to the AATGAAA heptamer the alfalfa H3 gene, ALH3-1.1, possesses a stretch of T-rich sequence (20) downstream from the 3′ end of its transcript. Thus, it is reasonable to propose that the cleavage and adenylate addition of the alfalfa histone H3 mRNAs occurs by a mechanism similar to that described for other eukaryotic mRNAs (24).

The Diversity of Alfalfa H3 cDNAs and H3 Histones

According to the compilation of nucleotide sequences, the four H3 cDNAs and the genomic sequence, ALH3-1.1 (20), can be divided into two classes: a major class and a minor class, based on the homology between them. The major class is represented by clone pH3c-1, pH3c-3 and pH3c-12 and the genomic clone, ALH3-1.1. These sequences vary only in a few nucleotides in both translated and untranslated region (Fig.2) and encode the same protein. Three additional alfalfa H3 cDNA clones characterised by restriction enzymes were also shown to belong to this major class (data not shown). In contrast, clone

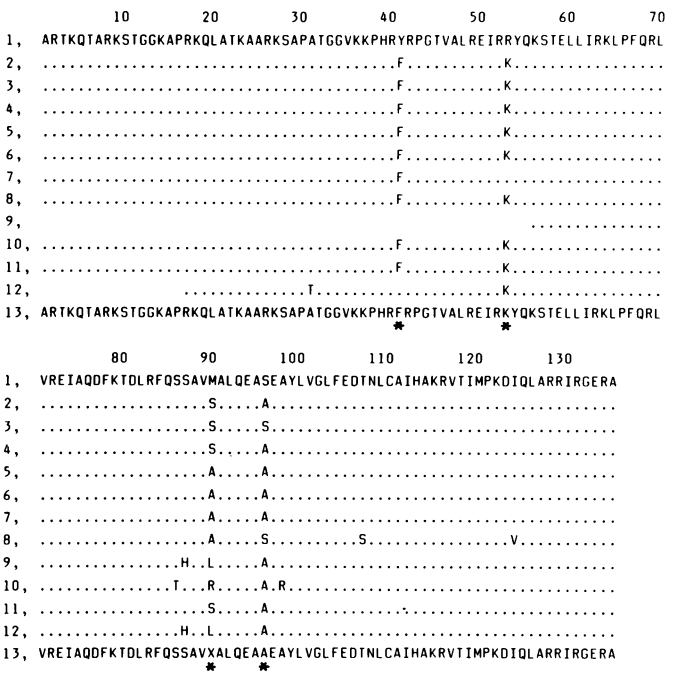

Fig.3. Comparison of amino acid sequences of plant H3 histones with consensus amino acid sequence of animal H3 histones. Dots represent amino acids identical to animal ones and asterisks indicate the main amino acid replacements in plant H3 histones in comparison with animal H3 consensus. Data are from 1, animal consensus (31); 2 and 3, pea embryo (32); 4, wheat H3 gene (33); 5, *Arabidopsis* H3 genes (34); 6, Corn H3 genes (35); 7 and 8, cycad pollen (36); 9, barley H3 cDNA (15); 10, rice H3 gene (37); 11, major alfalfa H3 genes (20); 12, minor alfalfa H3 gene (this paper) and 13, the suggested consensus amino acid sequence of plant H3 histones. The letter X represents variable amino acids at position 90.

pH3c-11 is considered to represent a minor gene which shows only 78.8% homology with the major class in the translated region and encodes a longer 3′ UTR with little homology to that of major class (Fig.2). The low homology between these two classes of alfalfa H3 genes results in four unique amino acid replacements (Fig.3). These include a change of a hydrophobic Ala to a polar Thr at position 31, a hydrophobic Phe to a polar Tyr at position 41, a polar Ser to a basic His at position 87 and a polar Ser to a hydrophobic Leu at position 90, respectively. Surprisingly, the amino acid sequence of this minor alfalfa H3 protein was found to be identical to that of a barley H3 histone deduced from a partial H3 cDNA sequence (15), although the codon usage of the barley and the minor alfalfa H3 genes belong to two completely different categories (20). Furthermore, the expression of this minor gene might be very low, since by using the 3′UTR of pH3c-11 as a specific probe the presumed longer mRNA band (Fig.2) possibly encoded by the minor gene was not detected in the same RNA samples under the same conditions as those in figure 1 (data not shown). We have recently reported that there were about 160 copies of histone H3 sequences in the tetraploid alfalfa genome (20). This high copy number may reflect a composition of major and minor gene families, and probably pseudogenes. A recent study on alfalfa histone proteins has also revealed the existence of different subtypes of

alfalfa H3 histones (28). The diversity of alfalfa cDNA sequences is consistent with these observations.

As in animals, many plant species were found to contain more than one histone variant for each type of histone. The amino acid sequences of some of these histone variants have been either directly determined or deduced from their genes (29). However, despite the diversity, plant H3 histones are highly conserved, particularly at four positions of their amino acid sequences which are different from the animals'(Fig.3). The most intriguing difference between the animal and plant H3 histones is located at position 90. At this position, the amino acids can be a hydrophobic Ala or Leu, a polar Ser and a basic Arg (Fig.3) in plant H3 histone. While in animal, at the same position, the amino acid is exclusively either a Met in H3.1 and H3.2 histones or a Gly in H3.3 histone which was considered to be an ancient protein (9). It should be of particular interest to study whether these amino acid replacements in plant H3 histones have any structural effect on plant chromatin. Surprisingly, two H3 histones (30) from a ciliate protozoan, *Tetrahymena*, where the H3 mRNAs are polyadenylated (11), possess the same amino acids at all the four positions as the plant H3 histone consensus. Whereas H3 histones from fungi (31), where the H3 transcripts are also polyadenylated(10), are closer to the animal basal variants H3.3 (9) than to the plant H3 histone. The conservation and variation of different classes of H3 histone among eukaryotes may be of evolutionary importance, this remains to be studied.

ACKNOWLEDGEMENT

We thank Drs. Cs. Koncz (Max-Planck Institut für Züchtungsforschung, Köln), F. Deák and F. Nagy for critical reading of this manuscript. The alfalfa cDNA library was constructed by J.Györgyey in the Laboratory of Genetics, State University of Gent, Belgium, with the generous help of Dr. D. Inzé and T. Alliotte.

*To whom correspondence should be addressed

REFERENCES

1. Hentschel, C.C. and Birnstiel, M.L. (1981) Cell 25, 301–313.
2. Maxson, R., Cohn, R. and Kedes, L. (1983) Ann. Rev. Genet.17, 239–277.
3. Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
4. Levine, B.J., Chodchoy, N, Marzluff, W.F. and Skoultchi, A.L. (1987) Proc. Natl. Aca. Sci. USA 84, 6189–6193.
5. Cotten, M., Gick, O., Vasserot, A., Schaffner, G. and Birnstiel, M. (1988) EMBO J. 7, 801–808.
6. Old, R.W. and Woodland, H.R. (1984) Cell 38, 624–626.
7. Wells, D., Hoffman, D. and Kedes, L. (1987) Nucl. Acids Res. 15, 2871- 2886.
8. Lieber, T., Angerer, L.M., Angerer, R.C. and Childs, G. (1988) Proc. Natl. Aca. Sci. USA 85, 4123–4127.
9. Wells, D., Bains, W. and Kedes, L. (1986) J. Mol. Evol. 23, 224–241.
10. Fahrner, K., Yarger, J. and Hereford, L. (1980) Nucl. Acids Res. 8, 5725- 5737.
11. Baum, G.A., Calzone, F., Bowen, J.K., Allis, C.D. and Gorovsky, M.A. (1983) Nucl. Acids Res. 11, 3903–3917.
12. Chaubet, N., Chabout, M.-E., Clement, B., Ehling, M., Philipps, G. and Gigot, C. (1988) Nucl. Acids Res. 16, 1295–1304.
13. Chaboute, M.-E., Chaubet, N., Clement, B., Gigot, C. and Philipps, G. (1988) Gene 71, 217–223.
14. Gantt, J.S. and Key, J.L. (1987) Eur. J. Biochem. 66, 119–125.
15. Chojecki, J. (1986) Carlsberg Res. Commun. 51, 211–217.
16. Gigot, C. (1988) In Kahl, G.(ed), Architecture of Eukaryotic Genes, Verlags-gesellschaft, Frankfurt, pp.229–240.
17. Stuart, D.A., Nelson, J., Stricland, S.G. and Nichol, J.W. (1985) In Henke, R.R., Hughes, W.K., Constantin, M.P. and Hollaeder, A.(eds), Tissue Culture in Forestry and Agriculture, Plenum Press, New York, pp.59–73.

18. Cathala, G., Savouret, J.-F., Mendez, B., West, B.L., Karin, M., Martial, J.A. and Baxter, J.D. (1983) DNA 2, 329–335.
19. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. USA 69, 1408–1412.
20. Wu, S.-C., Bögre, L., Vincze, E., Kiss, G.B. and Dudits, D. (1988) Plant Mol. Biol., 11, 641–649.
21. Schenck, R.U. and Hildebrandt, A.C. (1972) Can. J. Bot. 50, 199–204.
22. Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
24. Manley, J.L. (1988) Biochem. Biophys. Acta 950, 1–12.
25. Dean, C., Tamaki, S., Dunsmuir, P., Favreau, M., Katayama, C., Dooner, H. and Bedbrook, J. (1986) Nucl. Acids Res. 14, 2229–2240.
26. Hunt, A.G. (1988) DNA 7, 329–336.
27. Coruzzi, G., Broglie, R., Edwards, C. and Chua, N.-H. (1984) EMBL J. 3, 1671–1679.
28. Waterborg, J.H., Winicov, I. and Harington, R.E. (1987) Arch. Bioch. Biophys. 256, 167–178.
29. Spiker, S. (1988) In Kahl, G.(ed), Architecture of Eukaryotic Genes, Verlags-gesellschaft, Frankfurt, pp.143–162.
30. Hayashi, T., Hayashi, H., Fusauchi, Y. and Iwai, K. (1984) J. Biochem. 95, 1741–1749.
31. Wells, D.E. (1986) Nucl. Acids Res., 14, r119–149.
32. Patthy, L., Smith, E.L. and Johnson, J. (1973) J. Biol. Chem. 248, 6834–6840.
33. Tabata, T., Fukasava, M. and Iwabuchi, M. (1984) Mol. Gen. Genet. 196, 397–400.
34. Chaboute, M.-E., Chaubet, N., Ehling, M. and Gigot, C. (1987) Plant Mol. Biol. 8, 179–191.
35. Chaubet, N., Philipps, G., Chabout, M.-E., Ehling, M. and Gigot, C. (1986) Plant Mol. Biol. 6, 253–263.
36. Brandt, W.F. and Von Holt, C. (1986) FEBS Lett. 194, 278–281.
37. Peng, Z.-G. and Wu, R. (1986) Gene 45, 247–252.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.