Volume 1 · Number 3 · 1991
GENE EXPRESSION

Multiple polyadenylation signals and 3′ untranslated sequences are conserved between chicken and human cellular myosin II transcripts

Joanne M. Volosky and Thomas C. S. Keller III

Department of Biological Science, Florida State University, Tallahassee, Florida

We have isolated a chicken cellular myosin II heavy chain isoform cDNA clone that overlaps the published sequence for MHC-A (Shohet et al., 1989, Proc Natl Acad Sci 86, 7726–7730) and contains three canonical AAUAAA-polyadenylation signals in an additional 374 nucleotides at its 3′ end. S1 nuclease protection analysis and PCR-amplification of MHC-A cDNA 3′ ends have confirmed that all three of the signals are used in vivo. Differential usage of these signals without differential splicing in this region yields three messages that differ at their 3′ ends but appear to encode the same protein. Comparison of the new chicken sequence with the homologous human MHC-A cDNA sequence (Saez et al., 1990, Proc Natl Acad Sci 87, 1164–1168) has revealed a number of similarities at this end of their long 3′ untranslated regions (3′-UTRs). The three chicken polyadenylation signals reported here are positioned similarly to three signals evident in the human sequence. This region also contains distinct stretches of identity that are interspersed with regions of little homology. Within these regions of identity are a number of conserved sequence motifs, some of which have been demonstrated to be involved in mRNA metabolism in other systems. The pattern of mRNA sequence conservation demonstrated here suggests that the mechanisms for regulating MHC-A mRNA metabolism have been conserved between chickens and humans.

Myosin II, which functions as a mechanoenzyme in both muscle and nonmuscle cells, is composed of two identical heavy chains and two pairs of nonidentical light chains. Myosin heavy chain isoforms found in skeletal and cardiac muscle are encoded by a multigene family (Leinwand et al., 1983; Nguyen et al., 1982; Robbins et al., 1986; Emerson and Bernstein, 1987). Expression of heavy chain isoforms from this gene family is regulated by a variety of factors and correlates with the physiological and developmental state of the muscle (Yanagisawa et al., 1987). Despite differences in their physiological properties, however, all of the muscle myosin heavy chain isoforms function similarly in the context of a stable and highly organized contractile sarcomere.

By contrast, nonmuscle myosin has been implicated in a number of functions which are dependent on a variety of structural organizations (Warrick and Spudich, 1987). These cellular structures range from the transiently organized cleavage furrow common to all dividing cells to more stably organized, cell-specific structures such as the intestinal epithelial cell circumferential ring (Hirokawa et al., 1983). Two

Received July 10, 1991; revision accepted August 16, 1991.

Correspondence: Thomas C. S. Keller III, Department of Biological Science, Florida State University, Tallahassee, FL 32306-3050
Tel (904) 644-5572 Fax (904) 644-0481

 1052-2166/91/0103/223-09$2.00

isoforms of vertebrate nonmuscle or cellular myosin heavy chain (MHC-A and MHC-B) have been verified by cloning (Shohet et al., 1989; Katsuragawa et al., 1989; Saez et al., 1990). MHC-A and MHC-B mRNAs are coexpressed differentially in a wide variety of chicken cells at ratios that can be physiologically regulated (Kawamoto and Adelstein, 1991). The possible involvement of each isoform in multiple structures and functions may require complex regulation at several levels.

Analysis of similarities and differences in homologous cellular myosin isoforms from different species should yield clues to conserved mechanisms for regulation at the level of the mRNA. MHC-A isoform cDNAs have been cloned from chickens and humans (Shohet et al., 1989; Katsuragawa et al., 1989; Saez et al., 1990). Comparison of the available nucleotide sequences reveals significant similarities: the deduced amino acid sequences for chicken and human MHC-A are 90% identical (Saez et al., 1990), and both the chicken and human MHC-A cDNA sequences contain long (>1 Kb) 3′ untranslated regions (3′-UTRs) (Saez et al., 1990; Shohet et al., 1989).

There are significant differences, however, in the characteristics of the reported 3′ ends of the cDNA sequences. The human MHC-A sequence contains multiple canonical AAUAAA-polyadenylation signals near its 3′ end. In several human cell types, differential usage of these polyadenylation signals without differential splicing yields at least two, and possibly three, MHC-A mRNAs that appear to encode the same protein but differ in their site of polyadenylation (Saez et al., 1990). The existence of these different messages suggests that this region of the 3′-UTR might be important in regulating some aspect of human MHC-A expression. In contrast to the human sequence, however, the published homologous chicken MHC-A sequence, which terminates in 10 A residues, lacks any canonical polyadenylation signal (Shohet et al., 1989).

We report finding additional chicken MHC-A 3′-UTR sequence. This sequence is present in a MHC-A cDNA clone isolated from a chicken colonic epithelial cell λgt11 expression library with an affinity-purified polyclonal antibody to brush border myosin. This additional sequence contains three canonical AAUAAA polyadenylation signals in positions similar to those in the human MHC-A transcript. Furthermore, alignment of the human and chicken 3′ end sequences reveals discrete regions of identity that are interspersed with regions that have diverged significantly. Use of all three polyadenylation signals in chicken intestinal epithelial cells, where cellular myosin A is a major component of the brush border cytoskeleton, results in expression of three mRNAs that apparently encode the same protein but differ at their 3′ ends.

Materials and methods

Tissue isolation

Small intestines (duodenum) and colons were removed from freshly sacrificed adult leghorn chickens (obtained from Gold Kist, Live Oak, FL) and flushed thoroughly with cold saline. Epithelial cells from small intestines were prepared with sterile solutions essentially as described in Mooseker and Howe (1981). Colons were slit lengthwise, and laid flat in a petri dish. Cold, sterile cell dissociation medium (Mooseker and Howe, 1981) was pipetted onto the epithelial surface of the colon. After soaking for 10–15 minutes, the epithelial cells were gently scraped from the colon with a rubber policeman. Epithelial cells were collected by brief centrifugation at 150 × g., and the supernatant containing contaminating erythrocytes was carefully decanted. The epithelial cells were washed by resuspension in fresh cell dissociation medium and collected by centrifugation at 1000 × g. for 5 minutes. Small intestines were dissected from embryonic day 17 chickens and washed with sterile saline solution. Brains were removed intact from adult chickens and minced before use.

Preparation of RNA and colon epithelial cell cDNA library

RNA was isolated from whole embryonic small intestine and adult epithelial cells and brain by the method of Chomczynski and Sacchi (1987). Poly-A$^+$ RNA was purified by poly-U Sephadex chromatography according to the manufacturer's instructions (Gibco-BRL).

cDNA was synthesized using 5 μg Poly-A$^+$ RNA by the procedure of Gubler and Hoffman (1983), with modifications. Briefly, the first strand cDNA was synthesized using oligo-dT primers and M-MLV reverse transcriptase (Gibco-BRL). Second strand cDNA was synthe-

sized using RNase H and DNA pol I (Gibco-BRL). The double-stranded cDNA then was blunt-ended, and ligated first to EcoR I linkers and then to λgt11 phage arms (Promega). Phage DNA was packaged using Gigapak Gold Packaging Extract (Stratagene) according to the manufacturer's instructions. Screening of the library was performed by standard procedures with an affinity-purified antibody to intestinal epithelial cell brush border myosin (Keller et al., 1985), followed by a goat anti-rabbit IgG-alkaline phosphatase conjugated secondary antibody (Boehringer Mannheim Biochemicals).

RNA analysis

The RNA blot was prepared essentially as described in Forney et al. (1983), except that 1 μg of poly A^+ RNA from chicken small intestine, colon, and brain was used. SP6 RNA polymerase (Promega) was used to prepare an in vitro transcript complementary to the final 81 nucleotides of c4–11. Hybridization and washes were done at 65°C and the stringency of the final wash was 0.2× SSPE.

S1 nuclease analysis was performed essentially by standard procedures (Berk and Sharp, 1977) using 30 μg of chicken embryonic and adult small intestine total RNA. The probe was end-labeled following digestion of c4–11 with BglII (Promega) at a position corresponding to +6578 of the published chicken MHC-A sequence (Shohet et al., 1989).

PCR

First strand cDNA was synthesized from 1 μg poly-A^+ RNA, essentially as described above. The PCR reactions included 0.01% of the first strand cDNA, 0.2 μg each of a MHC-A-specific primer (5′-TTAATCTGCACAGATATATG-3′) and an oligo-dT adapter primer (5′-AGGATCCCCGGGT_{12}-3′), 200 μM dNTPs, 2.5 U Taq polymerase, and Taq polymerase buffer (Promega). DNA was initially denatured at 94°C for 7 minutes, followed by 5 cycles of 2 minutes denaturation at 94°C, 2 minutes annealing at 37°C, and 3 minutes extension at 72°C. The next 30 cycles consisted of 1 minute denaturation, 2 minutes annealing, and 3 minutes extension at the temperatures indicated, with the extension time lengthened by 5 seconds in each cycle. Products were electrophoresed on 2% agarose/TAE gels, and stained with ethidium bromide.

Cloning and sequencing of PCR products

The PCR products were digested with Hae III, which cuts the MHC-A sequence between the specific primer and the first polyadenylation signal, and Sma I, which cuts in the oligo-dT primer anchor sequence. The digested products were then electrophoresed on a 2% low-melting temperature agarose gel (International Biotechnologies Incorporated). Each band to be cloned was excised from the gel, melted at 65°C, ligated into Hinc II digested pGEM3zf(+) (Promega), and transformed into E. coli JM101. Single-stranded DNA made from the resulting clones was sequenced with Sequenase (version 2.0, United States Biochemical) according to the manufacturer's instructions.

Computer analysis

Computer analyses of human and chicken cellular MHC-A sequences were performed using programs in the GCG Sequence Analysis Software Package (version 7.0; (Devereux et al., 1984)).

Results

Several clones for MHC-A were isolated by screening a chicken colon epithelial cell cDNA λgt11 expression library with an affinity-purified brush border myosin antibody. One of the resulting clones, designated c4–11, was 2.5 Kb in length. c4–11 overlaps the 3′ end of the published sequence for MHC-A (Shohet et al., 1989), but also contains 374 bases of additional 3′-end sequence (Fig. 1). The existence of this additional sequence indicates that the 10 A residues terminating the published chicken MHC-A sequence are components of the 3′-UTR and not part of the poly-A tail. RNA blot analysis using a RNA probe complementary to the 81 nucleotides at the 3′ end of this additional sequence verified that it is part of a >7 Kb message in chicken intestinal epithelial cell, colonic epithelial cell, and brain poly-A^+ RNA (Fig. 2).

Three canonical AAUAAA polyadenylation signals are present in this new sequence. S1 nuclease protection analysis (Fig. 3) confirmed the existence in vivo of messages corresponding to usage of all three polyadenylation signals. The ends of two additional products map to the A/T-rich region that includes the 15 adenosines preceding the first polyadenylation signal and the A/T-rich region preceding the second sig-

```
aaaaaaaaaaAAAAAT AATAAAGGTCTTTATCACTGCCTTTCTATTGGGACCATGGTTATATATAGATAGTCTTTACTTT   80
TCTACACCGTACACTGGTTGCTGGATTTACCTGTATTCTTAACCATATTGTATATGCTGCATTTAGACCTACTTATGAGC   160
AAAGTAAAAATAATTGAGTATGAAGCTCCATACTGTATGCCTGAAGCGCTCTAAATGCAGGAGCCTCGTGTCCCAATTGC   240
TGTCACAGGAAAGTTTTAATTTTTTTTTTATATATATAATAAAAGTGCCTTAGCATGTGCCTCAGCTGTGTGTCACCACT   320
ACAGTCAGTAATGGATTACTGGTGCTCCACTGATGTTACCAATAAAGATTATCCATGTGCTGCA                    384
```

Figure 1. 3′-end sequence of chicken MHC-A cDNA clone. Shown are 384 nucleotides of chicken MHC-A 3′-untranslated sequence from clone c4–11 (2.5kb). This sequence includes ten adenosine residues (lower case) that overlap the published chicken MHC-A sequence (Shohet et al., 1989). Three canonical polyadenylation signals (AAUAAA) are indicated in bold face. Sites of polyadenylation following each signal, determined by sequencing of PCR-amplified cDNA, are designated (large arrows). Variations at the first site also are indicated (small arrows).

nal. Such A/T-rich regions have been shown to be S1-sensitive areas that can result in the production of spurious products (Wellman et al., 1987).

To determine the actual sites of MHC-A polyadenylation, we used the polymerase chain reaction (PCR) to amplify the 3′ ends of intestinal epithelial cell MHC-A transcripts. The sequence of one PCR primer, which was common to all of the MHC-A messages, was identical to a region near the end of the published chicken MHC-A sequence (Shohet et al., 1989). The other PCR primer consisted of a stretch of 12 T residues capable of priming from poly-A tails. It also contained an additional 12 nucleotides on the 3′ end which provided a restriction site in the amplified products to facilitate cloning.

Several distinct products resulted from PCR amplification of intestinal epithelial cell cDNA using these primers (Fig. 4). Southern blot analysis of these products using a 1-Kb 3′ restriction fragment of c4–11 as a probe revealed that four major bands, approximately 83, 119, 374, and 458 base pairs, contained MHC-A sequence (data

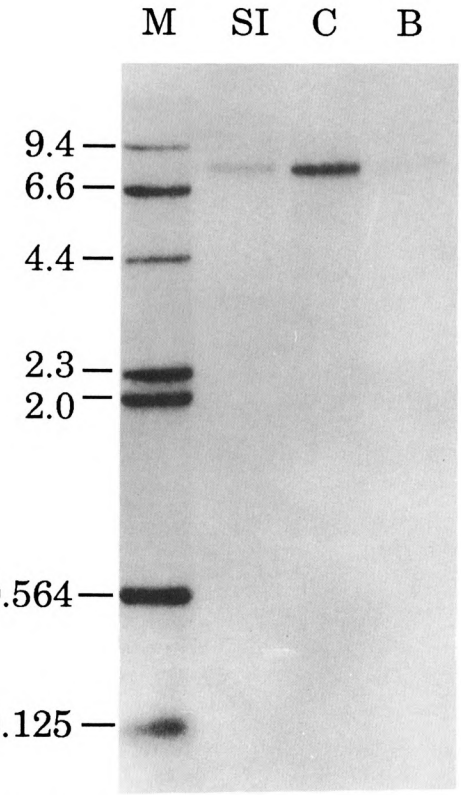

Figure 2. RNA blot analysis. RNA blot of small intestine (SI), colon (C), and brain (B) poly A⁺ RNAs from chicken. Blot was probed with an in vitro transcript complementary to the last 81 nucleotides of c4–11. All lanes contain 1 μg of poly A⁺RNA. M, end-labeled λ/HindIII DNA size markers (kilobases).

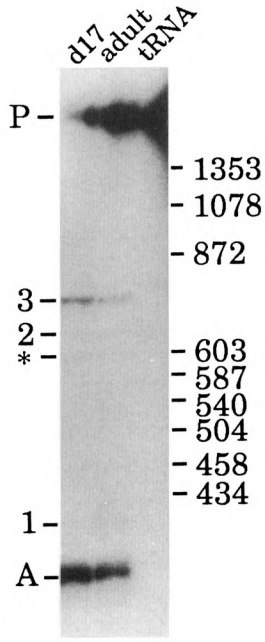

Figure 3. S1 nuclease protection analysis. An end-labeled probe was hybridized to 30 μg of total RNA from chicken embryonic day 17 small intestine (d17) and adult small intestinal epithelial cells and 30 μg tRNA and subjected to S1 nuclease treatment prior to electrophoresis on a 7.5% polyacrylamide/7M urea gel. Protected fragments with ends that map to sites following the first (1), second (2), and third (3) polyadenylation signals are indicated. Products resulting from cleavage of the probe at A/T-rich regions preceding the first (A) and the second (*) polyadenylation signals are indicated. Migration of full-length probe (P) and DNA size markers (base pairs) are indicated.

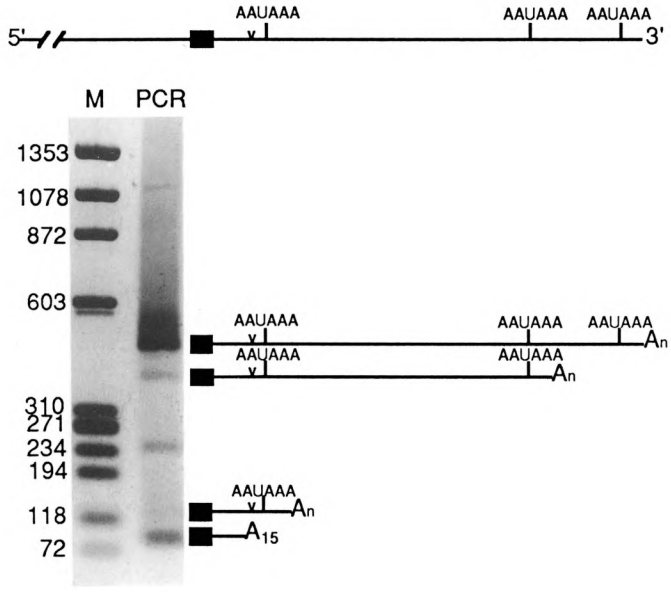

Figure 4. Products from PCR amplification of chicken MHC-A cDNA. Chicken MHC-A cDNA was amplified as described in Materials and Methods and analyzed on 2% agarose/TAE gels. Shown above the gel is a schematic of the longest chicken MHC-A mRNA, with the location of the polyadenylation signals indicated. Also indicated is an internal stretch of 15 adenosine residues, including those that end the published sequence (v), and the location of the MHC-A-specific primer used in amplification (black box). The 3′ ends of the messages giving rise to three of the major PCR products are shown next to the corresponding bands on the gel (119, 374, and 458 base pairs). Also shown is a product (83 base pairs) that results from annealing of the oligo-dT adaptor primer to the stretch of 15 adenosine residues located internally (v, A_{15}). M, DNA size markers, sizes (base pairs) indicated to left of gel; PCR, MHC-A PCR products, A_n, poly-A tail.

not shown). Two distinct bands, approximately 240 and 1100 base pairs, showed no hybridization.

DNA from each of the four MHC-A-positive bands and one negative band (240 base pairs) was cloned and sequenced (see Materials and Methods). The three largest positive bands contained the MHC-A-specific sequence, polyadenylation signals, and poly-A tail that are diagramed adjacent to the bands on the gel in Figure 4. The sites of poly-A addition that were verified by sequencing these clones are indicated in Figure 1 (black arrows). Several variations were found in the site of polyadenylation following the first signal. Of the clones sequenced in which the first signal was used, the majority (six) were polyadenylated 16 nucleotides downstream (Fig. 1, first large arrow) from the polyadenylation signal. Two other clones were polyadenylated at positions located 20 and 28 nucleotides downstream from the signal (Fig. 1, small arrows). The fourth positive band (approximately 83 base pairs) yielded clones of MHC-A sequence that termintae with the internal stretch of 15 adenosine residues; these clones resulted from priming from this region by the oligo-dT primer. Clones of DNA from the negative band did not contain MHC-A sequence, a result consistent with its lack of hybridization in the Southern analysis.

Since this property of multiple polyadenylation signal usage is shared by chicken and human MHC-A mRNAs, we compared their 3′ end sequences to ascertain similarities at the level of the primary sequence. Dotplot comparison of the chicken and human MHC-A nucleotide sequences reveals that they are highly homologous in the protein coding region (Fig. 5). In contrast, the sequences are highly diverged over most of their long 3′-UTRs, except for discrete regions near their 3′ ends corresponding to sequences that surround the polyadenylation signals.

Direct comparison of the human and chicken nucleotide sequences in this region reveals that each of the three chicken AAUAAA-polyadenylation signals has a counterpart in the human sequence (Fig. 6). This includes one previously unidentified human polyadenylation signal (Saez et al., 1990) that corresponds to the second polyadenylation signal in the chicken sequence. The sequence alignment also reveals that there are surprisingly long stretches of identity that contain homologies to previously described sequence motifs. These identical regions are interspersed with gaps in both messages and stretches of highly divergent sequence.

Discussion

Numerous investigations have confirmed the importance of the poly-A tail and other 3′-UTR sequences in governing mRNA stability, translational efficiency, and possibly localization in

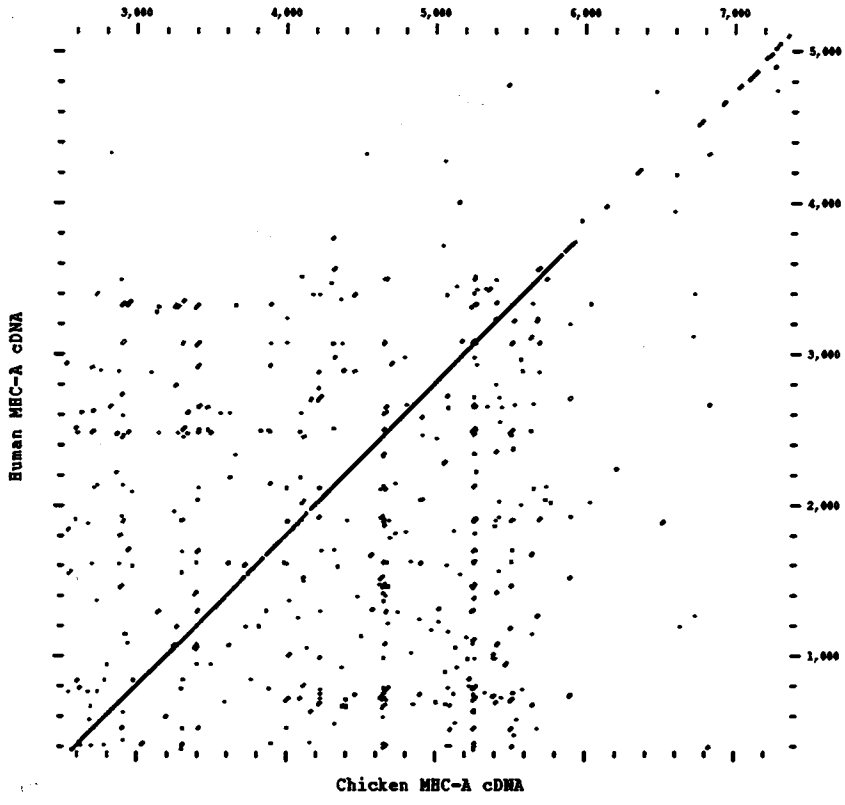

Figure 5. Dot matrix comparison of cDNA sequence for protein tail and 3′-UTR of chicken and human MHC-A. Chicken and human MHC-A cDNAs were compared beginning at nucleotide 2507 of the published chicken sequence (Shohet et al., 1989) and nucleotide 365 of the human sequence (Saez et al., 1990). The 3′-UTRs start at nucleotide 5929 of the chicken sequence and nucleotide 3743 of the human sequence. The chicken sequence used in this comparison contained the additional 374 nucleotides reported here. Comparison was performed using the Compare and Dotplot programs of the Wisconsin GCG Software Package (Devereux et al., 1984).

cells (for recent reviews, see Jackson and Standart, 1990; Brawerman, 1989; Wickens, 1990b; Hunt, 1988; Bernstein and Ross, 1989; Munroe and Jacobson, 1990). The conserved sequence and structural characteristics of the chicken and human MHC-A mRNA 3′-UTRs demonstrated here suggest that there are conserved mechanisms for regulating the expression of this cellular myosin II isoform.

With regard to overall structure of the

```
chk  aaaaaaaaaaAAAAAUAAUAAA.............G.GUCUUUAUCACUGCCUUUCUAUUGGGACCAUGGUUAUAUAUAGA
hum  UUUUCCCCUUUCUUGUAAUAAAUGAUAAAAUUCCGAGUCUUUCUCACUGCCUUU.............GUUUAGA.AGAGA

chk  .UA...GUCUUUACUUUUCUACACCGUACACUGGUUGCUGGAUUUACCUGUAUUCUUAACCAUAUUGUAUAUGCUGCAUU
hum  GUACUCGUCCUCACUGGUC.......UACACUGGUUGCCGAAUUUACUUGUAUUCCUAACUGUUUUGUAUAUGCUGCAUU

chk  UAGACCUAC..UUAUGA...GC.............AAAGUAA..AAA...U..AAU..UGA...G.UAU.GAAGCUCCAU
hum  GAGACUUACGGGCAAGAAGGGCAUUUUUUUUUUUUUAAAGGAAACAAACUCUCAAAUCAUGAAGUGAUAUAAAAGCUGCA.

chk  ACUGUAUGCCUGAAGCGCUCUAAAUGCAGGAGCCUCGUGUCCCAAUUGCUGUCACA..GGA..AAGUUUUAA........
hum  ....UAUGCCUACAAAGCUCUGAAUUCAG.........GUCCCAGUUGCUGUCACAAAGGAGUGAGUGAAAACACCCACC

chk  .U...UUUUUUUUUAUAUAUAUAAUAAAAGUGCCUUAGCAUGUGCCUCAGCUGUGUGUCACCACUACAGUCAGUAAUGGA
hum  CUACCCCCUUUUUUAU....AUAAUAAAAGUGCCUUAGCAUGUGUUGCAGC....UGUCACCACUACAGUAAG..CUGGU

chk  UUACUG..GUGCUCCACUGAUG..UUACCAAUAAAGAUUAUCCAUGUGCUGCA
hum  UUACAGAUGUUUUCCACUGA.GCAUCA.CAAUAAAGA.GAACCAUGUGCU
```

Figure 6. Alignment of chicken and human MHC-A 3′ ends. Shown is a direct alignment of the 3′-end sequences from the chicken and human MHC-A cDNAs. Polyadenylation signals (AAUAAA) are shown in bold face. Identities between the two sequences are indicated by vertical lines, with the following exceptions: (+) denotes conserved UUUUUAU motif and (I) denotes conserved (G/C)CAUGUG motif. GU-rich elements downstream from the first and second polyadenylation signals in each sequence are evident. Gaps have been inserted (.) to allow for optimal alignment.

mRNAs, our analysis has revealed that there is less than a 10% difference between the chicken and human sequences in both 3′-UTR length and in the distances between multiple polyadenylation signals in both transcripts. This degree of 3′-UTR length conservation has been found in messages for a number of isotypic proteins from different species (Yaffe et al., 1985), including sarcomeric myosin (Saez and Leinwand, 1986), further reflecting the possible importance of 3′-UTR spatial considerations in message regulation.

Also conserved are the existence and positions of three polyadenylation signals in the two transcripts. Multiple poly-A signal sequences have been found in primary transcripts encoding several different types of proteins. In many transcripts, selection of a polyadenylation signal is made in association with alternative splicing, resulting in a mature transcript with one polyadenylation signal (reviewed in Breitbart et al., 1987). In contrast, other mature mRNAs, like those of the chicken and human cellular myosins analyzed here, can contain multiple polyadenylation signals, suggesting that different mechanisms govern expression of these mRNAs.

Comparison of the primary structure of the two mRNAs has revealed extensive sequence conservation between the chicken and human messages in the region surrounding the MHC-A polyadenylation signals. The extent of this conservation (approximately 78% identity, GCG Gap program, Devereux et al., 1984) is much greater than would be predicted (<30% identity; Yaffe et al., 1985) if the sequences had been neutral over the 300 million years of evolution since chickens and mammals diverged. It also compares favorably with the 79% identity between the nucleotide sequences in the region of the messages that encode the highly conserved tail portion of the homologous myosin proteins (J. Volosky and T. Keller, unpublished observations). The fact that these 3′-UTR termini are as conserved as the protein coding sequences indicates their potential importance in the regulation of mRNA metabolism.

One possible function of specific sequences in this conserved 3′-UTR region is that they regulate polyadenylation during formation of the three ends. It has been well established that U/GU-rich elements that are located downstream from polyadenylation signals are important in the selection of polyadenylation sites (Wickens, 1990a). Although not identical between the two sequences, such a GU-rich element is present following the first polyadenylation signal of each sequence in a position where it could function in 3′-end formation. The GU-rich element downstream from the second polyadenylation signal in each sequence is homologous to the YGUGUUYY motif found in a wide variety of genes (McLauchlan et al., 1985); the exact sequence of this motif and its position relative to the polyadenylation signal, however, differ between the two sequences. Whether similar GU-rich elements exist following the third polyadenylation signal has yet to be determined for either sequence because this downstream region is removed during 3′ end formation.

Within a transcript, the differences in these downstream elements might affect the relative efficiency of the three polyadenylation signals; this could dictate a specific ratio of what are likely to be functionally distinct mRNAs. Our S1 nuclease and PCR analysis is consistent with the longest message being most prevalent in chicken intestinal epithelial cells. In human macrophages, products of the first and third signals appear to be equally prevalent (Fig. 5, in Saez et al., 1990). HeLa cells, on the other hand, appear to express all three products in approximately equivalent amounts (Fig. 5, in Saez et al., 1990). This ratio, therefore, might be regulated by favoring formation of particular ends in a tissue- or developmental stage-specific manner.

A sequence that might impart a distinct functional characteristic to one of the mature MHC-A messages is the conserved UUUUUAU that precedes the second MHC-A polyadenylation signal. Identical sequence motifs preceding AAUAAA polyadenylation signals dictate cytoplasmic polyadenylation of messages during Xenopus oocyte maturation; this activity has been correlated with increases in translational efficiency of these messages (Wickens, 1990b; Fox et al., 1989). It has been speculated that similar mechanisms exist in somatic cells, accounting for the increases in poly-A tail length that correlate with enhanced translation of insulin, vasopressin, and growth hormone (reviewed in Wickens, 1990b). If this motif governs cytoplasmic elongation of the poly-A tail in MHC-A messages, it may function only in messages polyadenylated at the second site: it is cleaved from

messages that are polyadenylated at the first site, and its position with respect to the poly-A tail may render it nonfunctional in messages that are polyadenylated at the third site. Consequently, use of the second polyadenylation signal may yield more stable messages.

A conserved motif of unknown function is the sequence (G/C)CAUGUG which follows the second and third MHC-A polyadenylation signals. This motif is homologous to the GCAUGU motif that follows the last AAUAAA signal in a number collagen gene transcripts (Maatta et al., 1991; Myers et al., 1986; Myers et al., 1983). In human and mouse collagen COL1A1 transcripts, this conserved sequence contains an additional G residue, making it GCAUGUG (Maatta et al., 1991), which is identical to the MHC-A sequence that follows the second polyadenylation signal. Whether these motifs are important in MHC-A messages or in other transcripts has yet to be verified.

Clearly, both the length and the specific sequences that remain in the mature MHC-A messages depend on the polyadenylation sites that are used. The 3′ end resulting from specific site selection might be important in governing message stability, translational efficiency, and localization through both the sequence motifs that are retained and possibly the secondary structures that are formed. Elucidation of the functional significance of the different MHC-A 3′ ends awaits characterization of their specific usage in various cells at different stages of development and under different physiological conditions.

Acknowledgments

We gratefully acknowledge the assistance of Scott Rabe, Christiana Eichholz, Mitchell Martin, and Ken Eilertsen in the cloning and sequencing. We also thank Drs. L. Epstein, L. Keller, and W. Marzluff for helpful discussions and comments on the manuscript, and Dr. Leslie Leinwand for sharing her data prior to publication. This work was supported by a grant from the NSF (DCB-8945577) and a Sigurd M. Johnson Research Grant from the American Cancer Society, Florida Division.

The nucleotide sequence data reported in this paper have been submitted to GenBank and have been assigned accession number M73982.

The costs of publishing this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC Section 1734 solely to indicate this fact.

References

A. J. Berk and P. A. Sharp (1977), Cell 12, 721–732.
P. Bernstein and J. Ross (1989), Trends Biochem Sci 14, 373–377.
G. Brawerman (1989), Cell 57, 9–10.
R. E. Breitbart, A. Andreadis, and B. Nadal-Ginard (1987), Annu Rev Biochem 56, 467–495.
P. Chomczynski and N. Sacchi (1987), Anal Biochem 162, 156–159.
J. Devereux, P. Haeberli, and O. Smithies (1984), Nucl Acids Res 12, 387–395.
C. P. Emerson, Jr. and S. I. Bernstein (1987), Annu Rev Biochem 56, 695–726.
J. D. Forney, L. M. Epstein, L. B. Preer, B. M. Rudman, D. J. Widmayer, W. H. Klein, and J. R. Preer (1983), Mol Cell Biol 3, 466–474.
C. A. Fox, M. D. Sheets, and M. P. Wickens (1989), Genes Dev 3, 2151–2162.
U. Gubler and B. J. Hoffman (1983), Gene 25, 263–269.
N. Hirokawa, T. C. S. Keller, R. Chasan, and M. S. Mooseker (1983), J Cell Biol 96, 1325–1336.
T. Hunt (1988), Nature 334, 567–568.
R. J. Jackson and N. Standart (1990), Cell 62, 15–24.
Y. Katsuragawa, M. Yanagisawa, A. Inoue, and T. Masaki (1989), Eur J Biochem 184, 611–616.
S. Kawamoto and R. S. Adelstein (1991), J Cell Biol 112, 915–924.
T. C. S. Keller, K. A. Conzelman, R. Chasan, and M. S. Mooseker (1985), J Cell Biol 100, 1647–1655.
L. A. Leinwand, R. E. K. Fournier, B. Nadal-Ginard, and T. B. Shows (1983), Science 221, 766–769.
A. Maatta, P. Bornstein, and R. P. K. Penttinen (1991), FEBS Lett 279, 9–13.
J. McLauchlan, D. Gaffney, J. L. Whitton, and J. B. Clements (1985), Nucl Acids Res 13, 1347–1368.
M. S. Mooseker and C. L. Howe (1981), in Methods in Cell Biology (L. Wilson, ed.), Academic Press, Inc., New York, pp. 143–174.
D. Munroe and A. Jacobson (1990), Gene 91, 151–158.
J. C. Myers, L. A. Dickson, W. J. de Wet, M. P. Bernard, M.-L. Chu, M. Di Liberto, G. Pepe, F. O. Sangiorgi, and F. Ramirez (1983), J Biol Chem 258, 10128–10135.
J. C. Myers, J. M. Brinker, N. A. Kefalides, J. Rosenbloom, S.-Y. Wang, and L. Gudas (1986), Nucl Acids Res 14, 4499–4517.
H. T. Nguyen, R. M. Gubits, R. M. Wydro, and B. Nadal-Ginard (1982), Proc Natl Acad Sci USA 79, 5230–5234.
J. Robbins, T. Horan, J. Gulick, and K. Kropp (1986), J Biol Chem 261, 6606–6612.
C. G. Saez, J. C. Myers, T. B. Shows, and L. A. Leinwand (1990), Proc Natl Acad Sci USA 87, 1164–1168.
L. Saez and L. A. Leinwand (1986), Nucl Acids Res 14, 2951–2969.
R. V. Shohet, M. A. Conti, S. Kawamoto, Y. A. Preston,

D. A. Brill, and R. S. Adelstein (1989), Proc Natl Acad Sci USA 86, 7726–7730.
H. M. Warrick and J. A. Spudich (1987), Annu Rev Cell Biol 3, 379–421.
S. E. Wellman, P. J. Casano, D. R. Pilch, W. F. Marzluff, and D. B. Sittman (1987), Gene 59, 29–39.
M. Wickens (1990a) Trends Biochem Sci 15, 277–281.
M. Wickens (1990b), Trends Biochem Sci 15, 320–324.
D. Yaffe, U. Nudel, Y. Mayer, and S. Neuman (1985), Nucl Acids Res 13, 3723–3737.
M. Yanagisawa, Y. Hamada, Y. Katsuragawa, M. Imamura, T. Mikawa, and T. Masaki (1987), J Mol Biol 198, 143–157.