Volume 11 Number 4 1983 Nucleic Acids Research

The nucleotide sequence of myosin light chain (L-2A) mRNA from embryonic chicken cardiac muscle tissue

H.H.Arnold[1]*, M.Krauskopf[2] and M.A.Q.Siddiqui[2]

[1]University of Hamburg, Department of Toxicology, Medical School Hamburg, Hamburg, FRG, and [2]Roche Institute of Molecular Biology, Nutley, NJ 07110, USA

Received 15 October 1982; Revised 15 December 1982; Accepted 10 January 1983

ABSTRACT

The nucleotide sequence of a cDNA clone (pML10) for chicken cardiac myosin light chain is described. The cDNA insert contains 613 nucleotides representing the entire coding sequence, with the exception of nine NH_2-terminal amino acids, and the full 3'-non-coding region of 146 nucleotides. The missing 5' terminus of the mRNA, not represented in the clone pML10, was obtained by extension of the cDNA using a 43 nucleotide long internal EcoR1 fragment as a primer. The non-coding region contains several direct and inverted repeated sequences and the polyadenylation signal sequence AATAAA. The coding portion exhibits non-random usage of synonymous codons with a strong bias for codons ending in G and C.

INTRODUCTION

Myosin from various vertebrate muscles consists of 2 heavy chains, polypeptides of 200,000 M.W. and 2 to 4 light chains (MLC) of 17,000 - 25,000 M.W. depending on the muscle type (1-4). Chicken cardiac muscle myosin contains two kinds of MLC designated L-1 and L-2 whose amino acid sequence has been recently described (5-6). Myosin light chain L-2 is known to consist of two components, L-2A and L-2B, which are very highly related but differ at specific positions of their primary structure (7). At least 8 different isozymic forms of avian MLC have been identified by amino acid sequence analysis (7). There are embryonic forms of MLC which are not found in the adult (8,9). Each of these MLC polypeptides represents presumably the product of a different gene. It is generally believed that proliferating myoblasts do not synthesize any of the myosin isozymes but synthesis is initiated at myoblast fusion in skeletal muscle cells. The molecular mechanism(s) that controls the expression of genes specific for the multiple myosin light chains is not clear. The evidence available so far suggests that several possible levels of control, i.e.,transcriptional, post-transcriptional or post-translational control, exist during muscle development (10-12). Similar information on cardiac myoblast differentiation, where no fusion occurs, is not available.

We have previously reported the construction of a recombinant plasmid (pML10) containing cDNA for embryonic chicken cardiac MLC (13). The specificity of the cloned sequence was demonstrated by hybrid-arrested translation assay. In the following, we present the nucleotide sequence for the entire cloned DNA representing the coding and 3'-non-coding regions of MLC mRNA. The amino acid sequence derived from the DNA sequence establishes the identity of clone, pML10, as specific for MLC-2A.

MATERIALS AND METHODS

Preparation of cloned DNA

The contruction of the recombinant plasmid pML10 was previously described (13). Plasmid DNA was isolated (14) and restriction fragments were fractionated on polyacrylamide gel electrophoresis and eluted either by electroelution or the crush and freeze technique (15). DNA fragments were also isolated by electrophoresis on low-melting agarose (Sea Plaque) gel and DNA was recovered by repeated phenol extraction.

Restriction mapping of recombinant DNA

Restriction enzymes were used as recommended by the supplier (Bethesda Research Laboratories). Enzyme digestion at multiple sites was mapped by electrophoresis on polyacrylamide gel using essentially the partial digestion technique of Smith and Birnstiel (16).

Sequencing of plasmid DNA

The sequencing method was essentially as described by Maxam and Gilbert (15). Fragments were labeled at the 5' end by polynucleotide kinase (Boehringer) and ^{32}P-γ-ATP (Amersham). Sequence samples were resolved on 20%, 10%, and 6 and 8% polyacrylamide gels, 0.3 mm thick and 40 cm long.

Primer extension of cloned cDNA

The internal EcoRI fragment of 43 nucleotides in length was labeled with ^{32}P-γ-ATP and polynucleotide kinase on both 5' ends. The strands were separated as described (15), and the complementary strand was used to hybridize to poly(A$^+$)RNA from embryonic chick heart in 50% formamide, 0.4 NaCl, 100 mM Pipes-buffer pH 4.5 and 0.1% SDS at 50°C for 5 hours. The hybrid was precipitated with ethanol and incubated in the presence of 10 mM $MgCl_2$, 50 mM Tris pH 8.3, 0.5 mM DTT, 1 mM each of dCTP, dATP and dTTP and 200 U/ml of reverse transcriptase for 1 hour at 42°C. The reaction was stopped with chloroform/phenol 1:1 and the denatured reaction products subjected to polyacrylamide gel electrophoresis on a 6% sequencing gel.

Secondary structure analysis of 3'noncoding sequence

Determination of potentially base pared regions in the 3'noncoding sequence of MLC-2AmRNA was performed by computer analysis on a VAX 11/780 using the program of Zucker and Steigler (24). The energy parameters for optimal binding energies were modified as described by Henco (25).

RESULTS AND DISCUSSION

The construction of the recombinant plasmid pML10 was described earlier (13). The DNA upon hybridization to total mRNA from 16-day old chicken cardiac muscle arrested specifically the translation of mRNA for myosin light chain 2, although the arrest was not complete (13). The identity of the plasmid pML10 has now been fully characterized by direct sequence analysis of the total DNA insert (17). Double-stranded DNA fragments of defined lengths were generated by digestion with restriction endonucleases (Fig. 1).
The fragments were radio-labeled on both 5'ends and subjected to secondary cleavages with appropriate restriction enzymes to obtain singly labeled subfragments, which were then sequenced separately. The complete cDNA sequence thus derived is shown in Fig. 2. The final sequence was obtained by repeated sequencing of the same fragment and by sequencing the overlapping fragments of the same and opposite polarity as shown in Fig. 1. The orientation and the reading frame were established by comparison of the derived amino acid sequence to the published primary structure of the protein (7). The cloned cDNA contains 613 nucleotides of coding sequence plus 30 and 70 adenosin re-

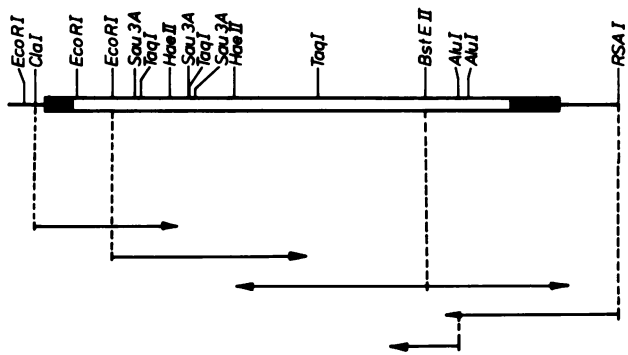

Figure 1: Restriction Map of the MLC-2 Insertion of PML-10 and Diagram of the Restriction Fragments used for Sequencing. The horizontal arrows are indicative for the orientation and the approximate length of sequences obtained. The hatched areas represent the poly(A-T)homopolymers used for inserting the cDNA into the vector.

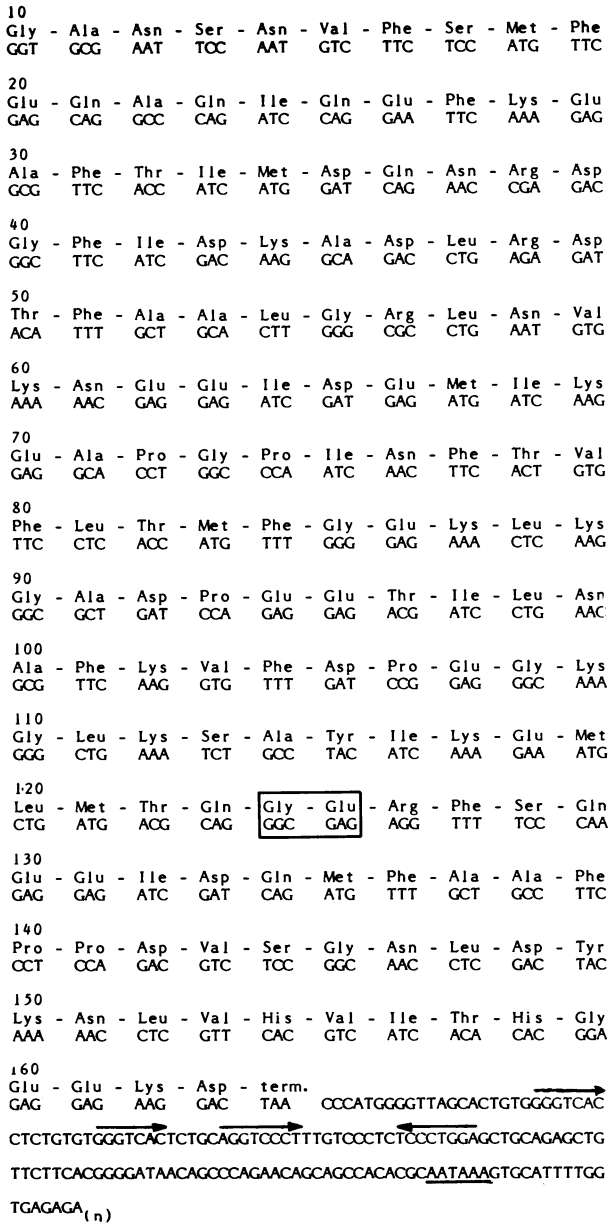

Figure 2: The complete nucleotide sequence of recombinant clone PML-10, and the derived amino acid sequence of MLC-2A. The boxed-in amino acids 124 and 125 appear in reversed order when compared to the published protein sequence (7). The underlined nucleotides represent the poly-adenylation signal of mRNAs. Direct and inverted repeated sequences are indicated by arrows above the sequence.

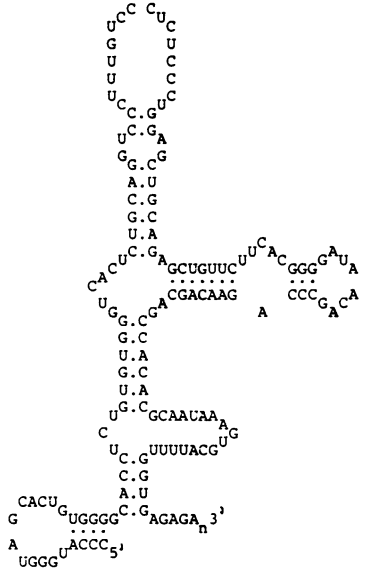

Figure 3: Possible secondary structure of the 3'-non-coding sequence of chicken MLC-2A mRNA. Maximal base pairing was determined by computer analysis as described under Materials and Methods.

sidues at the left and right ends, respectively. The inserted DNA appears to be smaller than we had suggested previously (13).

The sequence agrees at 152 positions with the published amino acid sequence of cardiac MLC-2A. However, the codons for amino acids glutamic acid and glycine located at positions 124 and 125, respectively (7) appear in opposite order in pML10 DNA. The DNA coding for these two amino acids has been sequenced on both strands and was read independently by two investigators. The hexanucleotide sequence GGCGAG which we have proposed would have to differ in at least 3 positions in order to agree with the published protein sequence. Since 3 sequencing errors within 6 nucleotides seem to be very unlikely, we feel that the reverse order of amino acids 124 and 125 is correct. An explanation for the disagreement of sequences could lie in genetic polymorphism, since different chicken strains were used in the studies (7, 13). MLC-2A differs from MLC-2B by 16 amino acid substitutions, one insertion and one deletion of an amino acid (7). In addition, differences may be found in the 3'-non-coding segments of the two mRNAs. These differences might explain the partial arrest of translation of MLC-2 mRNA when hybridized to pML10 under highly stringent conditions.

The pML10 DNA contains coding information for amino acid 9-163 of cardiac MLC2. Following the termination codon TAA, the entire 3'non-coding sequence of 148 nucleotides plus the poly A residues are present in the plasmid. In the 17 nucleotide segment before the poly(A), there is the hexanucleotide sequence AATAAA, the putative signal for polyadenylation of eucaryotic mRNA (18). The non-coding part of the cloned sequence contains multiple direct repeats of 6 to 7 nucleotides in lenth and an inverted repeated sequence of seven nucleotides separated by 9 nucleotides around the middle of the 3'-non-coding sequence. The functional significance of these sequences remains unclear. On the basis of simplified rules for estimating thermodynamic stability (24, 25), the most plausible secondery structure of the 3'trailing sequence was determined by computer analysis as shown in Fig. 3. About 50% of the nucleotides form helical structures, bringing the stop codon and the 3'end very close together, thus leading to a rather compact configuration of this sequence. It is noteworthy that a stable loop structure is located very close to the UAA stop codon of the mRNA. This arragment could suggest a role in translation, since it might support termination.

Alternatively, the loop could function in termination of transcription, as has been shown for the attenuation of bacterial operons (26) or the premature termination of late SV 40 transcription (27) in which potential stem-and-loop configurations of the transcript seem to be involved. A similar stem-and-loop structure has been observed in the 3'-non-coding sequence of the pGAP 30 DNA, coding for GAPDH mRNA of chicken (28).

To obtain information on the precise length of the MLC-2A mRNA a primer extension experiment was performed as described under Materials and Methods. As shown in Fig.4 , a maximally extended cDNA of 145 ($\pm$3) nucleotides was synthesized, indicating that the complete MLC-2A mRNA has a size of 727 ($\pm$3) nucleotides plus poly(A), consequently 96 ($\pm$3) nucleotides from the 5'terminus are not represented in the cDNA clone PML10. Since only coding information for 9 NH_2-terminal amino acids is missing in the clone, it can be calculated that the 5'-non-coding sequence comprises 69 ($\pm$3) nucleotides. The amount of radioactivity present in the extended cDNA was unfortunately not sufficient for sequence analysis.

The G + C content (51,5 %) of pML10 is significantly higher than that of total chicken DNA (42,3 %). The 3'-non-coding sequence is particularly rich in G + C (56,8 %), unlike the 3'-non-coding sequences of rabbit globin mRNA (19) and rat actin mRNA which are A + T rich (20). As shown for other mRNAs

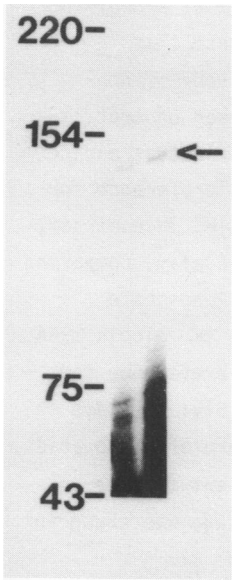

Figure 4: Autoradiograph of primer extended cDNA of the 5'end portion of MLC-2A mRNA. The priming DNA fragment and the reaction conditions are described under Materials and Methods. The maximally extended primer is indicated by the arrow. The shorter bands are presumably premature termination products.

Table 1

Phe	UUU 5	Ser	UCU 1	Tyr	UAU 0	Cys	UGU -
	UUC 9		UCC 4		UAC 2		UGC -
	UUA 0		UCA 0	Term	UAA -	Term	UGA -
	UUG 0		UCG 0		UAG -	Trp	UGG -
Leu	CUU 1	Pro	CCU 1	His	CAU 0	Arg	CGU 0
	CUC 4		CCC 0		CAC 2		CGC 1
	CUA 0		CCA 3	Gln	CAA 1		CGA 1
	CUG 5		CCG 2		CAG 6		CGG 0
Ile	AUU 0	Thr	ACU 1	Asn	AAU 3	Ser	AGU 0
	AUC 10		ACC 2		AAC 6		AGC 0
	AUA 0		ACA 2	Lys	AAA 7	Arg	AGA 1
Met	AUG 7		ACG 2		AAG 5		AGG 1
Val	GUU 0	Ala	GCU 3	Asp	GAU 6	Gly	GGU 1
	GUC 3		GCC 4		GAC 6		GGC 6
	GUA 0		GCA 3	Glu	GAA 2		GGA 1
	GUG 3		GCG 3		GAG 14		GGG 3

The frequency of use of each codon, as revealed by the LMC-2A mRNA from chicken heart, is indicated. Amino acids showing a highly significant preference are underlined with solid lines, marginal preference is indicated by dotted lines.

(19, 21, 22, 18), it appears that there is a non-random usage of synonymous codons for MLC as well. The data, summarized in Table 1 suggest that synonymous codons are not neutral, contradicting ideas in favour of non-Darwinian evolution (23). Particularly noteworthy is the marked underutilization of codons terminating in A or T and the preference for codons ending in G or C. There is a high statistical significance of codon preference for Ile, Glu, and Ser and to a lesser extent for Leu, Val, and GlN. In contrast, codons for Ala, Thr, Asp, Lys, Gly and Arg are used almost equally. Comparing the codon usage in MLC-2A mRNA with that in glyceraldehyde-3-phosphate dehydrogenase mRNA from the same chicken tissue (28) or with α-globin mRNA from different chicken tissue (29), we have found a general preference for codons ending in G or C. For some amino acids, like Ile and Glu, the same codons are preferentially used in all three mRNAs. The pattern of amino acids, however, which exhibit nonrandom codon utilization is different for the 3 mRNAs; for Thr and Gly, for instance, prodominant codon usage was shown in α-globin mRNA (29) but not in the other two mRNA. If there is nonrandom codon usage for a given amino acid, the same codons seem to be preferentially used in all three chicken mRNAs. Thus the bias in codon utilization seems to correlate to the organism, whereas the pattern of amino acids using codons preferentially appears to be specific for the mRNA per se, rather than for the organism or tissue. The functional and/or evolutionary meaning of non-random codon selection remains unresolved.

*To whom requests for reprints should be addressed

REFERENCES

1. Perrie, W.T. and Perry, S.V. (1970) Biochem. J. 119, 31-38
2. Weeds, A.G. and Lowey, S. (1971) J.Mol.Biol. 61, 701-715
3. Lowey, S. and Risby, D. (1971) Nature 234, 81-86
4. Kendrick-Jones, J. (1973) Phil. Trans. Roy. Soc. London B 265, 183-189
5. Matsuda, G., Maita, T., Umgegane, T. and Kato, Y. (1980) in Frontiers in Protein Chemistry, Liu, T.Y. et al.,Eds., pp. 323-334, Elsevier/North Holland, Amsterdam, New York.
6. Maita, T., Umegane, T., Kato, Y. and Matsuda, G. (1980) Eur. J. Biochem. 107, 565-575
7. Matsuda, G., Maita, T., Kato, Y., Chen, J. and Umegane, T. (1981) FEBS Lett. 135, 232-236
8. Rushbrook, J.I. and Sliacher, A. (1979) Proc. Natl. Acad. Sci. USA 74, 4331-4334
9. Whalen, R.E., Butler-Brown, E.S. and Gros, F. (1978) J. Mol. Biol. 126, 415-431
10. Reeburgh, Keller, L. and Emerson, Jr. Ch.P. (1980) Proc. Natl. Acad. Sci. USA 77, 1020-1024

11. Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A.M. and Yaffe, D. (1980) Proc. Natl. Acad. Sci. USA 77, 960-962
12. Stockdale, F.E., Raman, N. and Baden, H. (1981) Proc. Natl. Acad. Sci. USA 78, 931-935
13. Arnold, H.H. and Siddiqui, M.A.Q., (1979) Biochemistry 18,5641-5647
14. Bazaral, M. and Helsinki, D.R. (1968) J. Mol. Biol. 36, 185-194
15. Maxam, A.M. and Gilbert, W. (1980) in Methods in Enzymology, Colowick, S.P. and Kaplan, N.O., Eds., Vol. 65, pp. 499-559
16. Smith, H.O. and Birnstiel, M.L. (1976) Nucleic Acids Res. 3., 2387-2398
17. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
18. Proudfoot, N.Y. and Brownlee, G.G. (1976) Nature 263, 211-214
19. Efstradiadis, A., Kafatos, F.C. and Maniatis, T. (1977) Cell 10, 571-585
20. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D. Carmon, Y., Reiter, J., Frischauf, A.M. and Yaffee, D. (1981) Nucleic Acids Res. 9, 579-589
21. Suzuki, Y. and Brown, D.D. (1972) J. Mol. Biol. 63, 409-429
22. Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Min-Jou, W., Molemans, F., Raeymakers, A., van den Berghe, A., Volckaert, G. and Ysebaert, M. (1976) Nature 260, 500-507
23. King, J.L. and Jukes, T.M. (1969) Science 164, 788-798
24. Zucker and Steigler (1982) Nucleic Acid Res. 9, 133-148
25. Henco, H., (1979) Ph.D. Thesis, T.H. Darmstadt FRG
26. Stroynowski, J. and Yanofsky, C. (1982) Nature 298, 34-38
27. Skolnik-David, H., Hay, N. and Aloni, Y. (1982) Proc. Natl. Acad.Sci. USA 79, 2743-2747
28. Domdey, H. Wiebauer, K., Klapthor, H. and Arnold, H.H. (1982) Europ. J. Biochem. in press
29. Deacon, N.J., Shine, J. and Naora, H. (1980) Nucleic Acids Res. 8, 1187-1199