Myogenin Contains Two Domains Conserved among Myogenic Factors*

Atsuko Fujisawa-Sehara, Yoko Nabeshima, Yoko Hosoda, Takashi Obinata*, and Yo-ichi Nabeshima

From the Division of Molecular Genetics, National Institute of Neuroscience, National Center of Neurology and Psychiatry, 4-1-1 Ogawahigashi, Kodaira, Tokyo 187 and the *Department of Biology, Chiba University, Yayoi-cho, Chiba 260, Japan

Previously, three structurally related proteins, MyoD, myogenin, and Myf5, have been identified, and each of them was found to convert C3H10T1/2 fibroblasts to myoblasts when their respective cDNAs were expressed under the control of a viral promoter. Here, we describe the cloning and DNA sequencing of myogenin cDNAs from chicken and mouse. They encode polypeptides highly homologous to each other, but the polypeptide sequences we have obtained are not homologous in the carboxyl-terminal 70 amino acids with those previously reported for mouse, rat, and human because of a single base deletion in the previously reported cDNAs. Determination of genomic sequence coding for myogenin revealed that the mouse myogenin cDNA presented here corresponds to a correct transcript of the gene, and the nucleotide sequence of the previously reported cDNA is incorrect. A comparison of chicken and mouse myogenins with other myogenic regulatory factors, MyoD and Myf-5, identified a domain with an interesting feature located in the carboxyl terminus of these proteins in addition to the myc homology domain previously reported.

Myogenesis is generally considered to begin with the determination of a subset of mesodermal cells to become myoblasts, which are myogenic precursors. The mouse embryonic cell line, C3H10T1/2 (10T1/2),1 has provided an excellent system for investigating the molecular mechanism of myogenic determination. The initial observation was that treatment with 5-azacytidine converts 10T1/2 cells into morphologically and biochemically differentiated myotubes at high frequency (1, 4). Davis et al. (3) isolated cDNA clones that are expressed in proliferating myoblasts but not in 10T1/2 cells. When one of the clones, MyoD, was expressed in 10T1/2 and other fibroblast cell lines under the control of Moloney sarcoma virus long terminal repeat, MyoD could convert them to myoblasts, indicating that the expression of MyoD is sufficient for myogenic conversion (3). More recently, it was reported that cDNA clones encoding myogenin and Myf-5 which were isolated from L6 rat myoblasts and human fetal skeletal muscles, respectively, are also capable of converting 10T1/2 cells to myoblasts (4, 5). MyoD, myogenin, and Myf-5 share the myc homology domain and a preceding cluster of basic amino acids, suggesting similar functions for these proteins.

This paper presents the nucleotide sequences of mouse and chicken myogenin cDNAs that encode highly homologous polypeptides of 224 and 227 amino acid residues, respectively. The myogenin cDNAs from mouse (6), rat (4), and human (7) have been previously reported to encode polypeptides of 246, 287, and 246 amino acids, respectively. Their polypeptide sequences show no homology from amino acid 155 to the carboxyl terminus with those of chicken myogenin and the mouse myogenin we have cloned because of a single base deletion in the middle of the previously reported cDNAs. Determination of the genomic sequence of myogenin confirmed that the mouse myogenin sequence presented here corresponds to a correct transcript of the gene. Furthermore, the insertion of a 39-base pair fragment containing the genomic DNA sequence in question, but not the insertion of a 40-base pair fragment, in pUC vector resulted in the production of the fusion protein with β-galactosidase activity, which clearly indicates that the DNA sequence we have obtained represents the real DNA sequence. From the sequence comparison among chicken and mouse myogenins and other myogenic regulatory factors, MyoD and Myf-5, we found the existence of a domain located at the carboxyl terminus of these proteins other than the myc homology domain.

MATERIALS AND METHODS

Cells—C3H10T1/2 cells were obtained from the American Type Culture Collection. C2C12 myoblasts (14), 10T1/2 cells, and the transformants were grown in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal calf serum. Myogenic differentiation was induced by incubation in DMEM with 2% fetal calf serum (differentiation medium).

Isolation of Mouse Myogenin cDNAs—The cDNA library from mouse myoblast cell line, C2C12, was prepared by standard methods in λgt10 (15). Approximately 3 × 10^6 plaques were screened on nitrocellulose filters with a synthetic MyoD probe kinased with [γ-32P]ATP. The MyoD probe binds to a mouse homologous region was synthesized with a DNA synthesizer (Pharmacia LKB Biotechnology Inc.) and was purified by using polyacrylamide gel electrophoresis. The sequence of the probe is as follows: 5'-AGGCCGAGACCCCTGACATGGGACATTCTCACCTT-3'. The hybridization was carried out for 16 h at 42°C in 6 × SSC, 0.05% sodium pyrophosphate (pH 6.8), 5 × Denhardt’s solution, 100 μg/ml salmon sperm DNA, 20% formamide, 10% dextran sulfate, and 5 × 10^6 cpm of probe/ml. Washings were performed twice in 0.2 × SSC and 0.1% SDS for 1 h at 55°C. The cDNA inserts of positive clones were sequenced using the EcoRI linker on both ends.
were subcloned into pBluescriptSK(+) (Stratagene). The nucleotide sequence of these clones was determined essentially according to Sanger et al. (16) by using a 7-deaza sequence kit (Takara).

Isolation of Chick Myogenin cDNAs—The cDNA library from 11-day chick breast muscle was prepared in Xgt10. Approximately, 10^9 plaques were screened with the nick-translated cDNA of mouse myogenin (cut out with EcoRI from pBluescript). The hybridization was performed at 55 °C in 6 X SSC, 0.5% Tris-EDTA (pH 8.0), 1 M NaCl, 1 X Denhardt's, 20 mg/mL EDTA, and 0.2% SDS. Washing was performed in 5 X SSC and 0.1% SDS at 55 °C. Subcloning of the cDNAs inserts into pUC18 and DNA sequencing were performed as described above.

Isolation of Genomic DNA of Mouse Myogenin—A mouse (Balb/c) genomic library constructed in the AEMBL vector was purchased from Clontech. The library was screened with the full-length myogenin cDNA labeled with nick translation. Genomic DNA fragments from positive clones were subcloned into pBluescriptSK(+). The nucleotide sequence was determined as described above. In vitro amplification of a part of myogenin gene with the polymerase chain reaction (9) using genomic DNA from mouse liver (ICR) as a substrate was performed by using a GeneAmp DNA amplification kit (Perkin-Elmer Cetus) and automated with the DNA thermal cyclers (Cetus). The DNA sequence determination was performed by the chemical modification method (10).

Disruption of lac Z Gene by Insertion of DNA Fragments of Myogenin Gene—The 5′-ends of Sau3A 1-NcoI (476-514 in the cDNA) and FnuDII-NcoI (475-514 in the cDNA) fragments isolated from mouse myogenin cDNA were filled up with Klenow fragment (DNA polymerase I large fragment), and then, the resulting fragments (39 and 40 bp, respectively) were ligated with pUC119 at its SmaI site in vitro. This construct was transfected using a GeneAmp DNA amplification kit (Perkin-Elmer Cetus) and automated with the DNA thermal cyclers (Cetus). The DNA sequence determination was performed by the chemical modification method (10).

Expression of cDNAs—pHjYAPr-1, an expression vector which expresses the lac Z gene by insertion of DNA fragment, was kindly provided by Dr. Peter Gunning (8). Subcloned myogenin cDNAs from mouse and chicken were digested with HindIII and Sall and ligated in their Smal site in vitro. Transfectants of Escherichia coli host cells (DH5α) bearing the constructed plasmids were grown on plates supplemented with 100 mg/liter X-Gal, 100 mg/liter IPTG and 40 μg/ml ampicillin, allowing a color indication of in-frame and out-of-frame insertion of DNA fragments.

RESULTS AND DISCUSSION

Isolation of Mouse Myogenin cDNAs—Using a synthetic oligonucleotide corresponding to the myc homology region of MyoD as a probe, we screened a cDNA library from a mouse myoblast cell line, C2C12. Five MyoD cDNA clones were isolated; two had a 5′ untranslated region that was 41 bases longer than those reported previously (3). Three positive clones contained cDNAs which showed an identical restriction map among each other except for various lengths of the 5′-end. As is shown in Fig. 1, nucleotide sequence analysis revealed that these clones essentially correspond to the myogenin cDNA reported previously (6). However, there is a significant difference between the polypeptides encoded by the cDNAs we have isolated and the one reported previously. That is, they diverge completely between amino acid 155 to 197. As is shown in Fig. 1, nucleotide sequence analysis revealed that these clones essentially correspond to the myogenin cDNA described previously. When one of the cDNAs we have isolated was expressed in 10T1/2 cells under the control of the human β-actin promoter (8), colonies forming multinucleated myotubes could be ob-

![Fig. 1. Nucleotide and deduced amino acid sequence of mouse myogenin cDNA. Domain I and Domain II are underlined with solid lines. The nucleotide marked with a triangle is missing in the mouse myogenin cDNA.](image-url)
served (Table I), indicating that the cDNA encodes an active myogenic regulatory protein.

Isolation of Chicken Myogenin cDNAs—Subsequently, we screened a cDNA library prepared from 11-day chick breast muscle with mouse myogenin cDNA as a probe under low stringency hybridization conditions. Four positive clones had identical restriction maps except for various lengths of their 5’-ends. The DNA sequence of the longest clone and its open reading frame of 227 amino acids are shown in Fig. 2. Comparison of the nucleotide sequence of the coding region with that of mouse myogenin cDNA revealed 71% similarity. This was found to extend over the entire coding sequence. Furthermore, as is shown in Fig. 3, the polypeptide sequence encoded by the chick cDNA has 72% similarity with that encoded by our mouse cDNA, and this was found to extend over the entire protein. When the chicken cDNA was expressed in 10T1/2 cells under the control of the human β-actin promoter, myogenic colonies could be obtained with an efficiency similar to the mouse myogenin (Fig. 4 and Table I). The structural similarity together with the myogenic activity suggests that the chick cDNA represents the chicken homologue to mouse myogenin. However, the chicken myogenin loses similarity at amino acid 155, not only with the mouse myogenin reported previously but also with the rat and human myogenin due to a single base deletion in their cDNAs.

**TABLE I**

<table>
<thead>
<tr>
<th>Expression vectors</th>
<th>Number of myogenic colonies* per neo-resistant colonies</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHi/Pr(-)</td>
<td>0/54</td>
</tr>
<tr>
<td>pHi/Pr(+; mouse MyoD)</td>
<td>21/40</td>
</tr>
<tr>
<td>pHi/Pr(+; mouse myogenin)</td>
<td>11/54</td>
</tr>
<tr>
<td>pHi/Pr(+; chicken myogenin)</td>
<td>9/43</td>
</tr>
</tbody>
</table>

*Myogenic colonies were scored by the morphological appearance of multinucleated myotubes 7–10 days after medium shift.

![FIG. 2. Nucleotide and deduced amino acid sequence of chicken myogenin cDNA. Domain I and Domain II are underlined with solid lines.](image)

![FIG. 3. Amino acid sequence comparison of the chicken and mouse myogenin polypeptides. Bars indicate identity, and dots indicate conservative amino acid substitutions.](image)

![FIG. 4. Anti-troponin T staining of myotubes formed by a transformant expressing chicken myogenin. Immunopositive cells visualized with horseradish peroxidase-coupled second antibody appear dark in the photograph. Cells were photographed at × 200 power.](image)

![FIG. 5. Nucleotide sequence of mouse myogenin gene around the position at amino acid 155. The coding sequence is split into two different exons between amino acid 157 and 158 by an intron of approximately 600 bp.](image)
three different ways. First, the mouse myogenin gene was isolated to determine its DNA sequence. Second, the genomic sequence was determined by direct sequencing of a region of the myogenin gene which was amplified with the polymerase chain reaction using genomic DNA from mouse liver as a substrate (9). The direct sequencing excluded the possibility that we sequenced cloning artifacts. From these experiments, we concluded that the genomic DNA sequence of mouse myogenin around amino acid 155 is identical to the sequence of the cDNA which we have isolated (Fig. 5). The myogenin gene was found to contain an intron between amino acids 157 and 158. Finally, we expressed fusion genes of lacZ containing small pieces of genomic DNA fragments which cover the position in question in the SmaI site of pUC plasmid. When the 39-bp fragment which should keep the gene in-frame was inserted, the resulting plasmid could complement the gal-gene of an E. coli host as expected. In contrast, the insertion of a 40-bp fragment which should disrupt the reading frame of lac Z could not complement the gal-gene of host cells. Fig. 6 shows the sequence ladder of the genomic 39-bp fragment and the expression of the fusion genes in the cells grown on an agar plate supplemented with IPTG and X-Gal. The last experiment clearly indicates that the appearance of 4 C residues (or 4 G residues in the complementary strand) in the sequence ladder is not due to artificial factors which result in the appearance of a “ghost” band but represents the real DNA sequence of the gene. Since the direct sequencing of the gene amplified in vitro suggests that mouse has a single myogenin gene, we conclude that the cDNA reported by Edmondson and Olson (6) is a cloning artifact, or sequence determination of the cDNA is incorrect. Since all the myogenin cDNAs reported previously have a single base deletion at the same position, it is conceivable that the compression of sequence ladders might lead the authors to misreading of the sequences because there is a cluster of guanines and cytidines around the deleted position in each myogenin gene. It remains to be clarified whether rat and human myogenins actually diverge in the carboxyl terminus from those of mouse and chicken.

Existence of Two Conserved Regions among Myogenic Regulatory Factors—When the predicted protein sequence of mouse or chicken myogenin is compared with MyoD and Myf-5, significant homology is not restricted to the myc homology region (Domain I) as was previously reported but is also found in a region close to the carboxyl terminus (Domain II, Fig. 7). Domain I is highly conserved among three myogenic regulatory factors, and related domains are found in three myc genes, various regulatory genes including two genes of the Drosophila achaete-scute complex (T4 and T6), and the im-
munoglobulin enhancer binding myogenic activity of MyoD (12). Domain II is localized close to the carboxyl terminus of each protein. An interesting feature of Domain II is the presence of several hydrophobic residues interrupted by a few hydrophilic residues. A similar structure is also found in the carboxyl termini of T4 and T5 of the achaete scute complex (13), although their stretches of dispersed hydrophobic residues are shorter than those of myogenic proteins. A careful comparison between Domain II and the carboxyl half of Domain I reveals a striking similarity, implying a redundant structure or function within a single polypeptide. Murre et al. (11) showed that Domain I has the potential to form two amphipathic helices separated by an intervening loop. The similarity of Domain II to the carboxyl helix of Domain I suggests that Domain II is involved in protein-protein or intramolecule interactions by forming a helix similar to the carboxyl helix of Domain I. Our preliminary experiments on the formation of myogenic colonies suggest that a truncation of myogenins which removes domain II results in the decreased myogenic activity. Further studies will elucidate how different myogenic factors with similar structures take part in myogenesis.

Acknowledgments—We thank Peter Gunning for providing the plasmid pHSAPr-1. We also gratefully acknowledge the synthesis of the oligonucleotides by Fumihiro Taguchi and Taichi Uetsuki, and critical comments on this manuscript by Fumio Matsuzaki, Robert O. Kwon, Vincent Mauro, and Elaine Parrish.

REFERENCES