Title: MYOCYTE-SPECIFIC TRANSCRIPTION ENHANCING FACTOR 2

Abstract

The invention generally features members of the Myocyte-specific Enhancer Factor 2 (MEF2) protein family which has myocyte transcription enhancing activity, the MEF2 nucleic acids or proteins being used to increase muscle cell mass or activity in transgenic animals, or in victims of muscle cell atrophy.
FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AT</td>
<td>Austria</td>
<td>FR</td>
<td>France</td>
<td>MR</td>
<td>Mauritania</td>
</tr>
<tr>
<td>AU</td>
<td>Australia</td>
<td>GA</td>
<td>Gabon</td>
<td>MW</td>
<td>Malawi</td>
</tr>
<tr>
<td>BB</td>
<td>Barbados</td>
<td>GB</td>
<td>United Kingdom</td>
<td>NE</td>
<td>Niger</td>
</tr>
<tr>
<td>BE</td>
<td>Belgium</td>
<td>GN</td>
<td>Guinea</td>
<td>NL</td>
<td>Netherlands</td>
</tr>
<tr>
<td>BF</td>
<td>Burkina Faso</td>
<td>GR</td>
<td>Greece</td>
<td>NO</td>
<td>Norway</td>
</tr>
<tr>
<td>BG</td>
<td>Bulgaria</td>
<td>HU</td>
<td>Hungary</td>
<td>NZ</td>
<td>New Zealand</td>
</tr>
<tr>
<td>BJ</td>
<td>Benin</td>
<td>IE</td>
<td>Ireland</td>
<td>PL</td>
<td>Poland</td>
</tr>
<tr>
<td>BR</td>
<td>Brazil</td>
<td>IT</td>
<td>Italy</td>
<td>PT</td>
<td>Portugal</td>
</tr>
<tr>
<td>BY</td>
<td>Belarus</td>
<td>JP</td>
<td>Japan</td>
<td>RO</td>
<td>Romania</td>
</tr>
<tr>
<td>CA</td>
<td>Canada</td>
<td>KP</td>
<td>Democratic People's Republic of Korea</td>
<td>RU</td>
<td>Russian Federation</td>
</tr>
<tr>
<td>CF</td>
<td>Central African Republic</td>
<td>KR</td>
<td>Republic of Korea</td>
<td>SD</td>
<td>Sudan</td>
</tr>
<tr>
<td>CG</td>
<td>Congo</td>
<td>LZ</td>
<td>Liechtenstein</td>
<td>SE</td>
<td>Sweden</td>
</tr>
<tr>
<td>CH</td>
<td>Switzerland</td>
<td>LX</td>
<td>Sri Lanka</td>
<td>SI</td>
<td>Slovenia</td>
</tr>
<tr>
<td>CI</td>
<td>Côte d'Ivoire</td>
<td>LU</td>
<td>Luxembourg</td>
<td>SK</td>
<td>Slovak Republic</td>
</tr>
<tr>
<td>CM</td>
<td>Cameroon</td>
<td>LV</td>
<td>Latvia</td>
<td>SN</td>
<td>Senegal</td>
</tr>
<tr>
<td>CN</td>
<td>China</td>
<td>MC</td>
<td>Monaco</td>
<td>TD</td>
<td>Chad</td>
</tr>
<tr>
<td>CS</td>
<td>Czechoslovakia</td>
<td>MG</td>
<td>Madagascar</td>
<td>TG</td>
<td>Togo</td>
</tr>
<tr>
<td>CZ</td>
<td>Czech Republic</td>
<td>ML</td>
<td>Mali</td>
<td>TC</td>
<td>Togo</td>
</tr>
<tr>
<td>DE</td>
<td>Germany</td>
<td>MN</td>
<td>Mongolia</td>
<td>UA</td>
<td>Ukraine</td>
</tr>
<tr>
<td>DK</td>
<td>Denmark</td>
<td>MS</td>
<td>Mauritania</td>
<td>US</td>
<td>United States of America</td>
</tr>
<tr>
<td>ES</td>
<td>Spain</td>
<td>MU</td>
<td>Mozambique</td>
<td>UZ</td>
<td>Uzbekistan</td>
</tr>
<tr>
<td>FI</td>
<td>Finland</td>
<td>NZ</td>
<td>New Zealand</td>
<td>VN</td>
<td>Viet Nam</td>
</tr>
</tbody>
</table>
MYOCYTE-SPECIFIC TRANSCRIPTION ENHANCING FACTOR 2

Background of the Invention

Funding for the work described herein was provided by the federal government through the National Institutes of Health, which has certain rights in the invention. The invention relates to the use of muscle cell transcriptional regulators.


Summary of the Invention

Applicants have identified and isolated a family of muscle-specific transcription factors, the Myocyte-specific Enhancer Factor 2 (MEF2) protein family, and cDNA's encoding them. These transcription factors are characterized by their ability to enhance the transcription of genes in muscle cells that play important roles in muscle cell proliferation and differentiation. The transcription factors of the invention are useful for increasing muscle mass in agricultural or domestic animals, or in humans that suffer from muscle cell atrophy.

Accordingly, the invention generally features a transgenic non-human mammal, hereinafter referred to as a transgenic mammal of the invention, that includes a first transgene encoding a member of the Myocyte-specific Enhancer Factor 2 (MEF2) protein family having myocyte transcription enhancing activity. The term transgene is used to cover mammals comprising a transgene introduced at an embryonic stage into the mammal or into an ancestor of the mammal.
"A member of the Myocyte-specific Enhancer Factor 2 protein family", referred to herein as a MEF2 polypeptide, refers to a polypeptide that enhances the transcriptional activity of a set of structural genes that include a MEF2 consensus recognition site, 5'-CTAAAAATAAA-3' (SEQ ID NO: 18) or 5'-CTA(AT)₄ TAG-3' (SEQ ID NO: 19), as part of their 5' regulatory sequences. A MEF2 polypeptide will include a sequence substantially homologous to the MADS enhancer sequence (Fig. 1B) (SEQ ID NO: 2), and a sequence substantially homologous to the MEF2 region (Fig. 1A SEQ ID NO: 1) and 1C). The MEF2 family can include any active form of MEF2, including forms whose activity is potentiated by other substances. Myocyte transcription activity means activity in the assay described below or an equivalent assay.

In preferred embodiments, the nucleotide sequence of the first transgene can include at least one of the following elements: a) a nucleotide sequence encoding at least eleven consecutive glutamine residues, or b) a nucleotide sequence encoding the amino acid sequence SEEEEEL (SEQ ID NO: 20). The mammal can be any agricultural or domestic mammal, or any mammal used for laboratory, research, or diagnostic purposes. The MEF2 protein encoded by the transgene can include at least a 54 amino acid portion of the amino acid sequence of Fig. 1A (SEQ ID NO: 1). Where the wild-type protein includes an inactivation domain, the MEF2 protein can be a mutant of the wild-type protein, such that the first transgene is deleted for sequences encoding the inactivation domain.

Specific MEF2 family members according to the inventions are isoforms of the MEF2 sequence shown in Fig. 1, e.g., aMEF2 SEQ ID NO: 7, or xMEF2 SEQ ID NO: 3.
The transgenic mammal of the invention can further include a second transgene introduced into the mammal, or an ancestor of the mammal, at an embryonic stage, the second transgene including a promoter positioned to effect expression of a structural gene, the promoter being characterized in that the expression is enhanced by the MEF2 protein family member. Alternatively, the transgenic mammal of the invention can further include a second transgene introduced into the mammal, or an ancestor of the mammal, at an embryonic stage, the second transgene enhancing the activity of the MEF2 protein family member. The enhancing activity can be any enhancing activity that increases MEF2 activity, i.e., by increasing the amount of MEF2 transcribed, e.g., by increasing the expression of MEF2; or by increasing the activity of an at least partially inactive form of the MEF2 protein family member. The activity of a partially inactive MEF2 protein can be increased, for example, by including a transgene that phosphorylates MEF2, or by including a transgene that codes for a protease that deletes inactivating sequences from the primary sequence of the MEF2 polypeptide, or by including a transgene that codes for an activator molecule, e.g., a hormone. Examples of proteins that can enhance MEF2 activity include, but are not limited to, a MyoD polypeptide, a myogenin polypeptide, or a homeobox protein. A "myoD polypeptide", as used herein, can include any member of the myogenic basic-helix-loop-helix (bHLH) polypeptide family.

A transgene of the invention, e.g., a first transgene, or a second transgene, can be expressed by a tissue-specific promoter, e.g., a muscle cell specific promoter.

In another aspect, the invention includes an essentially pure nucleic acid encoding a member of the
Myocyte-specific Enhancer Factor (MEF2) protein family which has myocyte transcription enhancing activity.

In preferred embodiments, a MEF2 nucleic acid can include at least one of the following elements: a) a nucleotide sequence encoding at least eleven consecutive glutamine residues, or b) a nucleotide sequence encoding the amino acid sequence SEEEEEEL. The MEF2 nucleic acid can also encode a 54 amino acid portion of the amino acid sequence of Fig. 1A (SEQ ID NO: 1), e.g., a sequence including the conserved MADS domain, or a sequence including the MEF2 DNA binding domain. The MEF2 nucleic acid can be an isoform of the MEF2 sequence shown in Fig. 1, e.g., aMEF2 (SEQ ID NO: 7), or xMEF2 (SEQ ID NO: 3). The MEF2 nucleic acid can be part of a nucleic acid vector, wherein the vector can also, but does not of necessity, include a transcriptional regulatory sequence positioned and oriented to regulate expression of the nucleic acid encoding the MEF2 family member. A cell that contains such a vector is also included in the invention.

An additional preferred embodiment is a substantially pure MEF2 polypeptide encoded by any of the MEF2 nucleic acids defined above. The polypeptide can include at least a 54 amino acid portion of the amino acid sequence of Fig. 1A (SEQ ID NO: 1), e.g., a sequence including the conserved MADS domain, or a sequence including the MEF2 DNA binding domain. A MEF2 polypeptide can be included in a composition that additionally includes a pharmaceutically acceptable carrier.

In a third aspect, the invention includes a method of inducing the expression of muscle-specific genes of a mammal, e.g., a human, or a domestic animal. The method involves administering to the mammal a nucleic acid vector that encodes a member of the Myocyte-specific
Enhancer Factor 2 (MEF2) protein family that has transcription enhancing activity.

A preferred nucleic acid vector used in the above method of inducing the expression of muscle specific genes includes at least one of the following elements: a) a nucleotide sequence encoding at least eleven consecutive glutamine residues, or b) a nucleotide sequence encoding the amino acid sequence SEEEL (SEQ ID NO: 20). The method of inducing the expression of muscle-specific genes can further include a second nucleic acid administered to the mammal, the second nucleic acid enhancing the activity of the MEF2 protein family member. The enhancing activity can be any enhancing activity that increases MEF2 activity, i.e., by increasing the amount of MEF2 transcribed, e.g., by increasing the expression of MEF2; or by increasing the activity of an at least partially inactive form of the MEF2 protein family member. The activity of a partially inactive MEF2 protein can be increased, for example, by administering a second nucleic acid that phosphorylates MEF2, or by including a second nucleic acid that codes for a protease that deletes inactivating sequences from the primary sequence of the MEF2 polypeptide, or by administering a second nucleic acid that codes for an activator molecule, e.g., a hormone. Examples of proteins that can enhance MEF2 activity include, but are not limited to, a MyoD polypeptide, a myogenin polypeptide, a retinoblastoma polypeptide, or a homeobox protein. The invention also includes a method of inducing the expression of muscle-specific genes in a mammal, the method including administering a polypeptide expressed from any of the MEF2 nucleic acid sequences described above.

A method of alleviating symptoms of muscular dystrophy in a mammal features administering a MEF2
nucleic acid, or a member of the MEF2 protein family, to a mammal, preferably to a human diagnosed with any of the disease forms of Muscular Dystrophy, in a vector that includes means for expressing the MEF2-family member-encoding nucleic acid. Where the method of alleviating symptoms of muscular dystrophy features administering a nucleic acid, the method can further include administering a second nucleic acid, e.g., a nucleic acid encoding a dystrophin protein, to the mammal, the level of transcriptional expression of the second nucleic acid being enhanced by a member of the Myocyte-specific Enhancer Factor 2 (MEF2) protein family.

The invention also includes a method of preventing or reducing muscle atrophy in a mammal, involving administering a vector that includes a MEF2 nucleic acid of the invention, or a MEF2 polypeptide, to the mammal.

The invention also includes a method of enhancing muscle mass in a mammal, involving administering the MEF2 nucleic acid of the invention, or a MEF2 polypeptide, to the mammal. The administration can be by direct intramuscular injection.

The invention also includes a method of identifying a molecule that enhances the activity of a member of the Myocyte-specific Enhancer factor 2 (MEF2) family. The method includes providing a candidate molecule; providing a MEF2 family member of the invention in a solution; providing a MEF2 consensus nucleic acid binding sequence; and determining whether the candidate molecule enhances binding of the MEF2 family member to the MEF2 consensus binding sequence.

The invention also includes a method of identifying a molecule that enhances the activity of a member of the Myocyte-specific Enhancer factor 2 (MEF2) family. The method involves providing a candidate molecule; providing MEF2 nucleic acid of the invention,
transformed into a cell, the cell comprising a structural
gene which includes a regulatory region that includes a
MEF2 consensus binding sequence and a promoter responsive
to the consensus binding sequence; and determining
whether introduction of the candidate molecule into the
cell enhances expression of the structural gene.

The invention also includes a method of
identifying a molecule that enhances the activity of a
member of the Myocyte-specific Enhancer factor 2 (MEF2)
family. The method involves providing a candidate
molecule; providing a MEF2 nucleic acid of the invention,
transformed into a cell, the cell including a structural
gene which includes a regulatory region that includes a
MEF2 consensus binding sequence and a promoter responsive
to the consensus binding sequence; and determining
whether introduction of the candidate molecule into the
cell enhances expression of the structural gene.

"Essentially pure nucleic acid", as used herein,
is nucleic acid that is not immediately contiguous with
both of the flanking sequences with which it is
immediately contiguous (i.e., one at the 5' end and one
at the 3' end) in the naturally-occurring genome of the
organism from which the nucleic acid of the invention is
derived. The term therefore includes, for example, a
recombinant DNA which is incorporated into a vector; into
an autonomously replicating plasmid or virus; or into the
genomic DNA of a prokaryote or eukaryote, or which exists
as a separate molecule (e.g., a cDNA or a genomic DNA
fragment produced by the polymerase chain reaction or by
restriction endonuclease treatment) independent of other
nucleic acid sequences. It also includes a recombinant
DNA which is part of a hybrid gene encoding additional
colypeptide sequence.

"Homologous" refers to the sequence similarity
between two polypeptide molecules or between two nucleic
acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomeric subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology.

"A substantially pure MEF2 polypeptide" is a preparation which is substantially free of the proteins with which it naturally occurs in a cell.

The transcription factors of the invention bind and induce the expression of a number of muscle specific enhancers and promoters with the consensus sequence (C/T)(A/T)(A/T)AAATA(A/G) (SEQ ID NO: 21). These factors regulate muscle-tissue specific gene expression in skeletal, cardiac, and smooth muscle cells. In particular, applicants have isolated and characterized multiple isoforms of the MEF2 protein family. Four preferred genes encoding this family of transcription factors (aMEF2, xMEF2, dMEF2, and cMEF2) are described below. By alternative splicing, four different isoforms of dMEF2 and CM-MEF2; are produced.

MEF2 transcription factors according to the invention can be used to produce transgenic animals with increased muscle cell mass, to prevent or counteract muscle atrophy in humans or animals suffering a pathological muscular condition, or to develop pharmacological agents that regulate the expression of muscle-tissue genes.
Other features and advantages of the invention will be apparent from the following description and from the claims.

Brief Description of the Drawings

Fig. 1A is a representation of the nucleotide sequence and corresponding amino acid sequence of MEF2. Fig 1B compares amino acid sequences of a region of MEF2 with other proteins. Fig. 1C shows alternatively spliced isoforms (SEQ ID NOS: 1, 2, 6 and 7).

Fig. 2 is a representation of the nucleotide sequence and corresponding amino acid sequence of the xMEF2 isoform, a product of a related gene (SEQ ID NO: 3).

Figs. 3A-3G is an illustration of how ubiquitously expressed MEF2-related RNAs accumulate preferentially in skeletal muscle, heart, and brain.

Fig. 4 is an autoradiograph showing that xMEF2 RNAs are highly restricted to skeletal muscle, heart, and brain.

Fig. 5A through 5D are autoradiographs showing that endogenous myotube MEF2 and cloned MEF2 have identical DNA binding specificities.

Fig. 6A and 6B are electrophoretic demonstrations that skeletal, cardiac, and smooth muscle specific DNA binding activity is due to MEF2/aMEF2.

Figs. 7A and 7B are diagrammatic representations that cloned MEF2 reproduces site-dependent transcriptional activation present in skeletal, cardiac, and smooth muscle.

Fig. 8 is a diagrammatic representation that MyoD induces trans-activation in nonmuscle cells.

Fig. 9 is an illustration of the relation between the amount of injected DNA and CAT-activity.

Fig. 10 is an illustration of the time course of expression of injected gene constructs.
Fig. 11 is an illustration of the regional expression pattern of injected gene constructs throughout the left ventricular wall.

Fig. 12 is an illustration of the expression of promiscuous (MSV) or muscle-specific (-667rβ-MHC) promoter constructs in the right ventricle and in skeletal muscle.

Figs. 13A and 13B are illustrations of the correlation of CAT-to Luciferase-activity in co-injection experiments.

Fig. 14 is an illustration of the mapping of the 5' flanking region of the β-MHC gene in vivo.

Fig. 15 is a representation of the nucleotide (1-2161) and predicted amino acid (1-465) sequences of the dMEF2 cDNA. The double underlined region indicates the putative MADS domain. The region downstream of the MADS domain which is necessary for sequence specificity of the MEF2 related factors is underlined. The alternatively spliced (96nt) region at the 3' end of the cDNAs is overlined with a dashed line. (SEQ ID NO: 4)

Fig. 16 is a diagram of the various alternatively spliced dMEF2 gene products: white, untranslated sequence; checkered, MADS domain; spotted, MEF2 conserved region; diagonal stripes, dMEF2 alternative coding exons.

Fig. 17 is a sequence analysis comparing the predicted amino acid sequence of dMEF2 (SEQ ID NO: 5) and MEF2. A vertical line indicates an identical amino acid; : indicates a highly conservative substitution; and · indicates a conservative substitution.

Fig. 18 is a comparison of the MADS/MEF2 domain amino acid sequences from dMEF2 (present study) (SEQ ID NO: 5), MEF2 (Yu et al 1992) SEQ ID NO: 1), xMEF2 (Yu et al. 1992) (SEQ ID NO: 3), SRF (Norman et al. 1988) (SEQ ID NO: 11), MCM1 (Passmore et al 1988) (SEQ ID NO: 12), AGL6
(Ma et al. 1991) (SEQ ID NO: 13), AG (Yanofsky et al. 1990) (SEQ ID NO: 17), TM6 (Pnueli et al. 1991) (SEQ ID NO: 14), DEF (Sommer et al. 1990) (SEQ ID NO: 15), and AP3 (Jack et al. 1992) (SEQ ID NO: 16). The MADS domain is the checkered sequence and the MEF2 specific extension of the binding site corresponds to the spotted sequence. The overall identity between these factors is indicated at the right of each sequence. The absolutely conserved amino acids are indicated in capitals in the consensus and conservative substitutions are indicated in lower case letters. For the MADS domain, the consensus is calculated for all of the factors. For the MEF2 specific domain the consensus is calculated just for the 3 MEF2 related factors.

The two schematics show a cross section of the two amino-terminal regions which contain predicted amphipathic alpha helices (aa 20-33, and aa 60-69 respectively). The amino-terminal regions which contain predicted amphipathic alpha helices (aa 20-33), and aa 60-69 respectively). The amino-terminal end of helix1 begins at Thr-20 in the upper region of the diagram and rotates clockwise 100° per residue to Tyr-33. Helix 2 begins at Thr-60 in the upper region of the diagram and ends at tyrosine-69 in the lower region. The hydrophobic residues, which are in bold print, are clustered on one side of each alpha-helix.

Fig. 19 is a comparison by alignment of the amino acid sequences of aMEF2 (SEQ ID NO: 7), yMEF2 (SEQ ID NO: 8), CM-MEF2 (SEQ ID NO: 9), cMEF2 (SEQ ID NO: 10), and xMEF2 (SEQ ID NO: 3). Amino acids are expressed in one letter standard code.
Description of the Preferred Embodiment(s)

Methods

Library Screening

The initial MEF2 cDNA clone was obtained by screening a λGT11 expression library generated from primary human skeletal myocytes cultured from vastus lateralis with a probe containing four concatenated copies of the MEF2 site, sequences -1081 to -1059 of the mouse MCK enhancer (-1081/-1059) (Sternberg et al., Mol. Cell. Biol., 8: 2896-2909, 1988) (5'-GATCCTCCTCTAAAAATAACCCCTGTC-3') (SEQ ID NO: 22) at a specific activity of 7.8 x 10^8 cpm/μg. The screening procedures of Singh et al. (Cell, 52:415-423, 1988.) were followed with several modifications. The buffer used to blot the filters contained 5% nonfat milk powder (Carnation) in 1 x binding buffer (1 x BB: 20 mM HEPES, pH7.9, 50 mM KCl, 0.2 mM EDTA, 1mM DTT). After washing the filters twice with 0.25% milk in 1 x BB, the filters were incubated in the same buffer containing 10μg/ml poly(dI-dC)/poly(dI/dC), 10μg/ml denatured salmon sperm DNA, and the ^32^P-MEF2 probe (1.7 x 10^6 cpm/ml) at room temperature for 1 hr. The filters were then kept at 4 °C overnight with gentle agitation, followed by washing four times with 0.25% milk in 1 x BB at 4 °C for a total of 25 minutes, and subjected to autoradiography. One positive clone was purified, and the DNA insert (2.97 kb) was subcloned and completely sequenced.

For DNA hybridization screening, human adult male cardiac ventricle λZAPII (Stratagene, LaJolla, CA) and dog cardiac ventricle λgt10 (Scott et al., J. Biol. Chem., 263: 8958-8964, 1988) cDNA libraries, were screened according to standard methods (Sambrook et al., Molecular Cloning, A Laboratory Manual, Second Edition, 1989). Filters were hybridized at 37 or 42 °C in 5X SSC,
50 mM Na phosphated, pH 6.5, 1.2X Denhardt’s, 0.1% SDS, 100 μg/ml calf thymus DNA, 10% dextran sulfate, 25% or 50% formamide, and 2 x 10^6 cpm/ml probe. The probe was the 387 bp NsiI/NdeI MEF2 cDNA fragment (nt 342-728) labeled to specific activity (10^8-10^9 cpm/μg) with ^32P. Filters were then washed in 2X SSC/0.2% SDS at 25-37°C and exposed to film. The positive clones were purified, and the cDNA inserts were subcloned and sequenced.

**DNA Sequence Analysis**


**RNA Blot Analysis**

Poly-A+ RNAs from cultured cells and mouse tissues were electrophoresed (5 μg per lane) and transferred to membranes according to Sambrook et al. (Molecular Cloning, A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, 1989). The human tissue mRNA blot (2 μg per lane) was obtained from Clontech (Palo Alto, CA). Blots were hybridized in 1 M NaCl, 50 mM Tris-HCl, pH 7.5, 1% SDS, calf thymus DNA 100 μg/ml, 10% dextran sulfate, and 50% formamide with 1 x 10^6 cpm/ml of probe at 42°C. The probes were those described in Figures 3 and 4. The blots were then washed at progressively increasing stringencies up to 0.1 X SSC/0.1% SDS at 60°C. Between probes, blots were completely stripped.
Synthetic Oligonucleotides Used in Electrophoretic Mobility Shift Assay (EMSA)

All probes and competitor DNAs were double-stranded (d.s.) synthetic oligonucleotides. For each DNA, the nucleotide sequence (one strand, linker sequence shown in parenthesis), and coordinates in the respective enhancer or promoter is as follows: MEF2,

5'-(GATC)CTCGCTCTAAAAAAATAACCCGT(G)-3' (SEQ ID NO: 22)
(mouse MCK enhancer -1081/-1060, Sternberg et al., Mol. Cell. Bio., 8:2896-2909, 1988); MEF2mt, MEF2mt4, and MEF2mt6 were MEF2 mutants with point mutations shown in Table 1 (also in Cserjesi, et al., Mol. Cell. Bio. 11:4854-4862, 1991); A/Temb, 5'-(AGCTT)CGAACCTCTGCCTATTCTATATAG(G)-3' (SEQ ID NO:23)

OTF-2, 5'-(GATCC)TTCCCAATGTATTGCTGCTCACG-3' (SEQ ID NO: 25) (immunoglobulin κ light chain promoter -75/-51, Scheider et al., Cell, 52:783-793, 1987); MLC2-HF-1, 5'-(GATC)TCCCTGGGTTAAAAATAACCCCATGAC-3' (SEQ ID NO: 26) (rat cardiac myosin light-chain-2 promoter -35/-62, Zhu et al., Mol. Cell. Bio. 11, 2273-2281, 1991); MCK A/T,
5'-(GATC)GATCGATGCTGGTTATAATTAACCCACACAT-3' (SEQ ID NO: 27) (mouse MCK enhancer -1200/-1173, Sternberg et al., Mol. Cell. Bio. 8:2896-2909, 1988); cTNT A/T,
5'-(GATC)TCCGAGGGTTAAAAATAGCAAAAACCTCT-3' (SEQ ID NO: 28) (chicken cardiac troponin T gene -226/-119, Iannello et al., J. Biol. Chem. 266, 3309-3316, 1991); αMHC A/T-, 5'-(GATC)CTTTTCAGATTTAAAAATACTAAGTAA-3' (SEQ ID NO: 29) and αMHC A/T-2, 5'-(GATC)GCCCAAGACTAAAAAGGCCCTGGA-3' (SEQ ID NO: 30) (rat α myosin heavy chain gene -340/-313 and -
228/202, respectively (Mahdavi et al., PNAS, 82:2626-2630, 1984). The core sequences of these sites are shown in Table 1. (Also located on page 58)

<table>
<thead>
<tr>
<th>Probe/Competitor DNA</th>
<th>Sequence</th>
<th>MEF2 Binding</th>
<th>Sequences ID</th>
</tr>
</thead>
</table>
| MEF2                 | GTGCTGTAACAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC

Table 1. Nucleotide sequences of probes and competitor DNAs used in MEF2 binding assays. Only the core sequences of the d.s. oligonucleotides are shown. + ( ) represent positive and negative binding of the probes, respectively. (Reproduced by Cambridge and Ollman, Mol Cell Bio, 11:4854-4862, 1991.)
Preparation of nuclear extracts and EMSA

The nuclear extracts from C2C12 myoblasts, myotubes, HeLa cells, and rat primary neonatal cardiocytes were prepared as described previously (Yu et al., Mol. Cell. Bio., 2:1839-1849, 1989; Thompson et al., J. Bio. Chem., 266:22678-22688, 1991). Nuclear extracts from NIH3T3 cells, 10T1/2 cells, and smooth muscle cells were prepared according to the procedures of Schreiber et al. (Nucl. Acids. Res., 18:5496-5503, 1990). Smooth muscle cells were from a cell line derived from adult rat pulmonary arteries. The EMSA assays were carried out as described previously (Yu, et al., 1989, supra) with a few modifications. When the nuclear extracts were examined for the binding activities, the incubation mixture contained 4-7 µg extract, 0.2 ng probe, 3-3.5 µg polydI-dC/polydI-dC, and 100 ng single stranded (s.s.) synthetic oligonucleotide as nonspecific DNA competitors in the binding buffer. When the in vitro translated protein was used in the EMSA assays, the incubation mixture contained 1.5 µl translated reticulocyte lysate, 0.2 ng probe, 0.45 µg polydI-dC, and 100-150 ng s.s. oligonucleotide. The bound fraction and the free probe were separated in a 5% polyacrylamide gel (acrylamide:bis = 29.1) at 4°C.

Generation of anti-MEF2 antibodies and supershift gel retardation assays

Synthetic peptides corresponding to the partial alternative exons in MEF2 (SEQ ID NO: 1) and aMEF2 (SEQ ID NO: 7) (Fig. 1A), TPHTEEKYKKINEEF (SEQ ID NO: 31) and (C)DYFEHSP-LSED (SEQ ID NO: 32), respectively, were used to raise antibodies against MEF2 and aMEF2 (Harlow and Lane, Antibodies, A Laboratory Manual, 1988). The specificities of antisera were demonstrated by the EMSA using the in vitro translated MEF2 and aMEF2, as well as by the specific immunoprecipitation of MEF2 and aMEF2 obtained from in vitro translation and in vivo
expression. The anti-MEF2 antiserum recognized both MEF2 and aMEF2, whereas the anti-aMEF2 recognized aMEF2 only. For supershift EMSA, the procedures of Brennan and Olson (Genes & Dev. 4:582-595, 1990) were followed, using 1 µl of serum.

Construction of Plasmid DNAs

For in vitro and in vivo expression of cloned MEF2 isoforms, the cDNA inserts were subcloned into pGEM vectors (Promega Corp., Madison, WI) and pMT2 vector (Kaufman et al., Mol. Cell. Bio. 9:946-958, 1989), respectively. To generate the MHcemb-CAT reporter constructs, two copies of various oligonucleotides were inserted at -102 of the MHcemb promoter in plasmid pE102CAT (Fig. 7A) (Yu, et al., Mol. Cell. Bio., 1989 supra). Two copies of oligonucleotides were also cloned at the HindIII site of p8TCKAT (Thompson, J. Bio. Chem., 266:22678-22688, 1991) located at -109 of the HSV thymidine kinase gene promoter to generate the TK-CAT reporter constructs.

Tissue culture and transient expression assays

The tissue culture and transient expression assays were performed as described previously (Yu, et al., 1989 supra; Thompson et al., J. Bio. Chem., 266:22678-22688, 1991). Transfections were carried out using 10 µg of the individual CAT reporter plasmid, 5 µg of either pMT2-MEF2 or vector pMT2, and 3 µg of the internal control pSV-βgal. The preparation of cell extracts and the assays on the activities of CAT and β-galactosidase were reported previously (Yu, et al., 1989 supra; Thompson et al., 1991 supra). When noted, 5 µg of pMSV-myOD (Davis et al., Cell, 51:987-1000, 1987) was used. Pulmonary arterial smooth muscle cells were maintained in DME/20%FCS. For transient expression assays, these cells were allowed to grow to about 60% confluency, and transfected with various DNAs by calcium phosphate coprecipitation as
described above. Cells were glycerol shocked 18 hrs later, and re-fed with DME/20%FCS. After 24 hours, the media was changed to low serum media (DME/5% heat inactivated horse serum), and cells were harvested 48 hours later.

Results

**MEF2 and Related Isoforms Are Members of the MADS Gene Family**

Using oligonucleotides containing four concatenated copies of the MCK MEF2 binding site sequence, a total of $1.5 \times 10^6$ recombinants were screened from a λgt11 cDNA expression library generated from primary human skeletal myocytes cultured from vastus lateralis. A single positive clone was obtained, producing a protein which specifically bound the probe.

The results are shown in Fig. 1. In Fig.1A (SEQ ID NO: 1) the nucleotide (1-2968) and predicted amino acid (1-507) sequences of the MEF2 cDNA are shown in upper case letters (SEQ ID NO: 1). The aMEF2 cDNA differs from MEF2 in the alternatively spliced exon beginning at nt 673 (aa 87), which is 2 codons shorter and is indicated above the MEF2 sequence (SEQ ID NO: 7). The underlined region is highly conserved between these isoforms and the product of another gene, xMEF2 (Fig. 2) (SEQ ID NO: 3), including the MADS domain underlined in bold. The sequence of the clone containing the alternatively spliced 5' untranslated region is indicated in lower case letters (unnumbered) (SEQ ID NO: 6), with the dotted line overlying the excluded Alu repeat. In Fig. 1B the MEF2 and xMEF2 MADS domain amino acid sequences (#3-57 OF SEQ ID NO: 1, and #3-57 OF SEQ ID NO: 3, respectively) are compared to those of the plant homeotic genes *agamous* (AG, Yanofsky et al., *Nature*, 346:35-39, 1990 (SEQ ID NO: 17) and *deficiens* (DEFA, Sommer et al., *EMBO J.*, 9:605-613, 1990) (#1-55 of SEQ ID
NO: 15), the human serum response factor (SRF, Norman et al., Cell, 55:989-1003, 1988) (#1-55 of SEQ ID NO: 11), and the yeast transcription factors MCM1 (Ammerer, Genes Dev. 4, 299-312, 1990) (#1-55 of SEQ ID NO: 12) and ARG80 (Dubois et al., Mol. Gen. Genet., 207:142-148, 1987) (SEQ ID NO: 2). The first position of each is numbered. Residues conserved in MEF2, xMEF2, and at least two of the other proteins are marked (*). In Fig. 1C the various alternatively spliced isoforms of the MEF2 gene are diagrammed: white, untranslated sequence; checkered, MADS domain; black, constant regions; horizontal and vertical stripes, MEF2 and aMEF2 alternative coding exons. An additional isoform that introduces a premature stop codon is also depicted (diagonal stripes). The alternative sequence encoding the peptide SEEEEEEL (SEQ ID NO: 20), absent in RSRFC4/RSRFC9 (Pollack and Treisman, Genes Dev. 5, 2327-2341, 1991), is indicated.

The 2.97 kb insert has a long open reading frame encoding a predicted polypeptide of 507 amino acids, provisionally named MEF2, with a calculated molecular weight of 54.8 kD and isoelectric point of 7.99 (Figure 1A). The methionine initiation codon is preceded by a translation stop three codons upstream. The 3' end of the cDNA has a tract of eleven adenosines, but there is no canonical polyadenylation signal. The sequence AACAAA appears beginning 29 nt upstream, but this has been shown to be a poorly functional mutation of the consensus (reviewed in Birnstiel et al., Cell, 41:349-359, 1985). Thus, this tract of adenosines may be internally encoded in a longer 3' untranslated sequence.

The N-terminal region of the encoded MEF2 protein (amino acids 3-57) (SEQ ID NO: 1) is closely homologous to the conserved DNA binding and dimerization domains of the recently identified MADS gene family, comprising a series of yeast and human transcription factors and plant
homeotic loci (Fig. 1B; reviewed in Schwarz-Sommer et al., Science, 250:931-936, 1990; Ceon, et al., Nature, 353:31-37, 1991). A region rich in basic residues (amino acids 3-31) overlaps a relatively long predicted α-helix from amino acids 23-48. Beyond the MADS domain, there is a distinctive sequence of 27 consecutive glutamines and prolines (amino acids 420-446 (SEQ ID NO: 1) and another region rich in serine and threonine (amino acids 141-186, 43% S+T). Domains such as these are important for the transcription activation function of other factors (Courey, et al., Cell, 55:887-898, 1988; Courey, et al., Cell, 59:827-836, 1989; Mermod et al., Cell, 58:741-753, 1989). The MEF2 sequence contains numerous potential phosphorylation sites, i.e. nine for casein kinase II ([S,T]XX[D,E]) and eight for protein kinase C ([S,T]X[R,K]), that could be important for post-translational regulation (Sorger, et al., Cell, 54:855-864, 1988; Yamamoto et al., Nature 334, 494-498, 1988; Manak et al., Genes Dev. 4, 955-967, 1990; Boyel et al., Cell 64, 573-584, 1991).

Using a MEF2 cDNA subfragment (nt 342-728) encompassing the MADS domain as a probe, we also screened 1.25 x 10⁶ recombinants from an adult human cardiac ventricle λZAPII cDNA library at a range of hybridization stringencies. Sequencing of the 16 clones isolated revealed several isoforms in addition to the original MEF2 from the skeletal muscle library, that apparently arise from alternatively spliced transcripts of the same MEF2 gene (Fig. 1C).

One partial cDNA isoform (lower case in Fig. 1A) has an alternatively processed 5' untranslated sequence that excludes the segment from nt 56-262 (SEQ ID NO: 6). This deleted domain is an Alu repetitive element (Jelinek, et al., Ann. Rev. Biochem. 51:813-844, 1982). This isoform also has an additional 80 nt of untranslated
sequence at its 5' end. A second alternative splicing event results in the substitution of translated sequences: amino acids 87-132 (nt 673-810) in the original MEF2 isoform are replaced by a different peptide, shorter by two codons, in the alternative isoform named aMEF2. These alternative peptide sequences share limited homology, with 15 identical residues and 12 conservative substitutions out of 44 positions. Another cDNA clone was identified that differs entirely from MEF2 downstream from nt 672, i.e. at precisely the point of MEF2/aMEF2 divergence (see Figure 1C). In the divergent sequence of this clone, however, the translational reading frame terminates after just 12 nt and, as it begins with a possible 5' splice site (AG-GTAACA), it may be a retained intron (data not shown). While this cDNA could arise as an artifact from reverse transcription of incompletely spliced nuclear RNA, retained introns do occur in regulated alternative splicing in some systems (Breitbart et al., Ann. Rev. Biochem., 56:467-495, 1987).

MEF2 and aMEF2 are apparently isoforms of the same gene that also encodes the human SRF-related clones RSRFC4 and RSRFC9, respectively (Pollack and Treisman, Genes Dev. 5, 2327-2341, 1991). RSRFC4 and RSRFC9 correspond to the isolate without the 5' untranslated Alu sequence. However, nt 1279-1302 in Figure 1, encoding the amino acids SEEEEEL (SEQ ID NO: 20) (residues 289-296 in MEF2), are absent from RSRFC4/RSRFC9, presumably as a result of alternative RNA splicing. Also absent are two of the eleven glutamine codons (CAG) at nt 1672-1704, possibly due to alternative splicing also, most likely at adjacent splice acceptor sites (CAGCAGCAG). The RSRFC4/RSRFC9 sequence lacks a single A nucleotide among the three at nt 1892-1894, possibly a sequencing discrepancy, that produces a shifted reading frame with a different C-terminus eleven amino acids shorter than
MEF2. Furthermore, the RSRFC4/RSRFC9 sequence does not possess the transcription enhancing activities of the MEF2 factors. Other minor differences in RSRFC4/RSRFC9, either allelic or sequencing discrepancies, include the absence of two GT repeats at nt 2084-2093, and two G-T transversions at nt 1767 and nt 2655, none of which affects the protein sequence.

We also isolated clones corresponding to the MEF2 alternative isoforms from a canine heart cDNA library. These include the form with the apparent retained intron, lending credence to the hypothesis that this represents a bona fide splicing event. It is striking that the dog and human MEF2 nucleotide sequences, which are 93% conserved in translated regions, are also better than 90% conserved over the entire 5' (excluding the Alu repeat) and most of the 3' untranslated sequences. There are no long open reading frames in the untranslated regions in either species to suggest that they might actually be translated in unidentified alternatively processed isoforms. However, these highly conserved sequences may be important in regulating mRNA turnover or translation.

The cDNAs shown in Figure 1 all derive from the same MEF2 gene, as is clear from the absolute sequence identity outside the alternative regions and from the genomic structure. We also isolated the product of a second related gene, xMEF2, by low-stringency screening of the human cardiac library (Figure 2).

The nucleotide (1-1500) and predicted amino acid (1-365) sequences of the xMEF2 cDNA are shown in Fig. 2. The underlined region is highly conserved between xMEF2 and MEF2/aMEF2 (Fig. 1A), including the MADS domain underlined in bold. The remainder of the sequence is entirely divergent. The canonical polyadenylation signal is overlined. (Note that nt 1 is actually from the linker used in cloning.)
This 1.5 kb cDNA, xMEF2, has a 365 amino acid open reading frame following the methionine codon at nt 250. The predicted protein has a calculated molecular weight of 38.6 kD and an isoelectric point of 10.24. Residues 3-57 constitute a MADS domain identical to MEF2 at 50 of 55 positions (Figure 1B). The xMEF2 and MEF2 peptide sequences remain similar immediately downstream of this domain over another 29 residues, with just four conservative substitutions. The corresponding nucleotide sequences are 76% homologous over these regions. Beyond residue 86, MEF2 and xMEF2 have no substantial similarity. This point of divergence aligns precisely with the beginning of the MEF2/αMEF2 alternative peptides (see Figure 1A), consistent with it being an exon boundary. The remainder of xMEF2 is peculiarly proline-rich (22%) overall; however, it lacks a long glutamine/proline domain like that found in MEF2. There are three potential casein kinase II and seven potential protein kinase C phosphorylation sites.

It should be noted that the methionine at position 1 in xMEF2 is actually the first methionine codon within an uninterrupted long open reading frame that extends to the 5' end of this cDNA, i.e., it is unknown whether a stop codon or, alternatively, the true initiation codon, might lie further upstream. Nevertheless, the xMEF2 peptide as depicted in Figure 2 aligns exactly with MEF2 and DEF A, both of which also have N-terminal MADS domains. In addition, the sequence around codon 1 in xMEF2 has a 6 of 7 match to the initiator consensus sequence, suggesting that this is a functional translation start site (Kozak, Cell 44, 283-292, 1986). The 3' end of xMEF2, in contrast to MEF2, terminates with a canonical polyadenylation signal and poly-A tail. xMEF2 is an alternatively spliced isoform of the gene that also encodes the SRF-related clone RSRFR2 (Pollack...
and Treisman, Genes Dev. 5, 2327-2341, 1991): at nt 33/34, it lacks 178 nt of 5' untranslated sequences containing the sole upstream in frame stop codon present in RSRFR2. The protein coding sequences are identical.

MEF2-Related Transcripts Accumulate Preferentially in Muscle and Brain Tissues

To determine the tissue distribution of the cloned sequences, we probed blots of poly-A+ RNAs from a series of cell lines and human tissues with an MEF2 cDNA fragment (nt 342-728) containing the MADS sequences (Figure 3A and 3B).

Fig. 3 shows northern blots of poly-A+ RNAs from a variety of muscle and non-muscle cell lines (Fig. 3, panels A, C, E; Mb, myoblasts; Mt, myotubes; 28S and 18S ribosomal RNA positions shown) and adult human tissues (Fig. 3, panels B, D, F; RNA size markers indicated in kilobases, kb) were sequentially hybridized, stripped, and rehybridized at high stringency to a series of radiolabeled probes derived from the MEF2 cDNA, including; MADS Domain (Fig. 3, panels A,B; nt 342-728), 3'UT Sequence (panels C,D; nt 2158-2969), and Exon-Specific (panels E,F; nt 673-810). Another blot of adult mouse tissue poly-A+ RNAs was also probed with the Exon-Specific probe (panel G). The corresponding aMEF2 exon-specific probe gave identical results (not shown). Each blot has equivalent amounts of RNA per lane (see Materials and methods).

MEF2 transcripts were found in all cells and tissues examined, but were more abundant in myotubes, skeletal muscle, heart, and brain. In all samples, the predominant species is ≈6.5 kb, with a minor band at ≈3.5kb. The abundance of the longer transcript is increased relative to the shorter one in differentiated myotubes, as compared with myoblasts and non-muscle cells. Smaller bands were also detected in non-muscle
cells. Because of the possibility that the conserved MADS sequence was cross-hybridizing with transcripts from related genes (see Figure 4), we probed the same blots with a second fragment (nt 2158-2968) (SEQ ID NO: 1) comprising only the MEF2 3' untranslated sequence (Figure 3C and D). This probe showed the same distribution of 6.5 and 3.5 kb transcripts (but not the smaller bands), confirming that these species are, in fact, products of this MEF2 gene. The hybridization of this human untranslated probe to rodent RNAs at high stringency again reflects the unusual interspecies conservation of these sequences as noted above for the dog clones.

In order to investigate the possible tissue-restricted splicing of these transcripts, we generated exon-specific probes corresponding to the two alternative coding exons for MEF2 and aMEF2 (see Figure 1A) and hybridized them individually to the same mRNA blots, and to another blot with mouse tissue poly-A⁺ RNAs (Figure 3E, 3F, and 3G). Both exon-specific probes show that, while transcripts containing these exons are expressed ubiquitously at low levels, they are noticeably more abundant in myotubes, skeletal muscle, heart and brain. This enrichment is even more pronounced than is seen using either MADS or 3' untranslated probes, indicating that there is tissue-specific regulation of MEF2 splicing, or perhaps mRNA stability. That both exons give the same result (data not shown) indicates that they are regulated in parallel, and that other transcripts of this gene detected only by the common region probes must lack these exons.

Similar northern blot analysis for xMEF2, using a probe from this cDNA (nt 1-502) (SEQ ID NO: 3) that contains its MADS sequence, demonstrated that expression of the xMEF2 gene is clearly tissue-specific (Figure 4).
The same cell (panel A) and human tissue (panel C) RNA blots shown in Fig. 3, and another of rat heart poly-A⁺ RNA (panel B) were probed at high stringency with a radiolabeled fragment of the xMEF2 cDNA (nt 1-502).

Again, transcripts are abundant in myotubes, skeletal muscle, heart, and brain. The major species in myotubes form a doublet at approximately 7.5 and 6.5 kb, with a less abundant transcript at about 3.5 kb. In the tissues, only the 7.5 kb and 3.5 kb bands are seen.

These xMEF2 transcripts are present at a lower level in myoblasts (which generally include a small subpopulation of differentiated myocytes in culture) and are barely detectable in non-muscle, non-neural cells and tissues. Smaller species in HeLa and CV-1 are distinct from those seen with the corresponding MEF2 probe. It is noted that none of the cDNAs isolated, either for MEF2 or xMEF2, is as long as the transcripts for these genes in RNA blots.

Cloned and Endogenous MEF2 Have Identical DNA Binding Specificities

The presence of MEF2-related transcripts in multiple tissues contrasts sharply with the muscle specific MEF2 activity described previously (Gossett et al., Mol. Cell. Bio., 2:5022-5033, 1989; Cserjesi and Olson, Mol. Cell. Bio. 11, 4854-4862, 1991). We undertook experiments to compare the DNA binding specificity of cloned proteins with that of the endogenous muscle activity.

Electrophoretic mobility shift assay (EMSA) confirmed specific binding of the MCK MEF2 site in C2C12 myotube nuclear extract (Figure 5A; probe and competitor oligonucleotide sequences are shown in Table 1), as demonstrated by others (Gossett et al., Mol. Cell. Bio. 2:5022-5033, 1989).
In Fig. 5A, C2C12 myotube nuclear extract was assayed for binding to the radiolabeled MEF2, CARG, and MEF2 mutant probes (specified at bottom) in the absence (-) or presence (+) of a 100- or 250-fold molar excess of unlabeled competing oligonucleotide (specified at top), with sequences shown in Table 1. Bound probe (B) was separated from free probe (F) by EMSA and detected by autoradiography. Lanes 1 and 12 show probe without extract. In Fig. 5B, in vitro translated MEF2 protein from the cloned MEF2 cDNA was similarly assayed for DNA binding. Controls showing probe alone (P), bound in myotube nuclear extract(C2), and not bound in unprogrammed rabbit reticulocyte lysate (RL) are included for comparison (lanes 1-3). In Fig. 5C, in vitro translated proteins from the three corresponding cDNAs (indicated at top) were each assayed for binding to a series of known or potential MEF2 sites from muscle gene regulatory regions shown in Table 1. MCK MEF2 is the MEF2 site, and RRL is unprogrammed rabbit reticulocyte lysate. The EMSA autoradiograms are cropped to show only the bound probes (arrowheads). In Fig. 5D the DNA binding domain of MEF2 was identified using EMSA in which full length in vitro translated MEF2 and a series of C-terminal deletions (d1-d4) were tested for binding to the MEF2 probe. Truncated cDNA templates are diagrammed at bottom; boxes represent coding and lines untranslated (UT) sequences; restriction enzyme cleavage sites are marked for HindIII (H), sacI (S), ndeI (N), and nheI (Nhe), producing the N-terminal peptide lengths indicated. The autoradiogram shows free probe (F) separated from that bound by MEF2 (B), d1 (B1), d2 (B2), and d3 (B3), while d4, cleaved immediately downstream from the MADS sequences, fails to bind. Unbound probe (P) and unprogrammed lysate (RL) controls are included.
The MEF2 site probe was bound (B) by an activity in this extract (lane 2). This interaction was competed by excess unlabeled probe (lane 3) but not by the mutated MEF2 site (lanes 4 and 5), confirming that the interaction is specific. The A/Temb site, a cis element in the embryonic myosin heavy chain (MHCemb) promoter important for its muscle specific activity (Bouvagnet et al., Mol. Cell. Bio., 7:4377-4389, 1987; Y.-T.Y. and B.N.-G., in preparation), was a less effective competitor (lanes 6 and 7). Unrelated A/T-rich sequences including CArG, which is a target for another MADS protein SRF (Boxer et al., Mol. Cell. Bio. 9:515-522, 1989), and the OTF-2 site (Scheidereit et al., Cell, 51:783-793, 1987), did not compete at all for MEF2 binding (lanes 8-11); nor did the MEF2 oligonucleotide compete for CArG binding in complementary experiments (lanes 13-15), consistent with previous reports (Gossett et al., Mol. Cell. Bio., 2:5022-5033, 1989). Further, the extract bound MEF2 mutant site mt4, but not mt6, distinguishing between ubiquitous and muscle specific binding (lanes 16-18), as shown previously (Cserjesi and Olson, Mol. Cell. Bio. 11, 4854-4862, 1991). These data confirm that the MEF2 site is specifically bound by a myotube nuclear factor distinct from known ubiquitous binding activities.

Cloned MEF2 exhibited the same DNA binding specificity as the endogenous myotube activity in similar EMSAs using cDNA-encoded in vitro translated MEF2 (Fig. 5B). The mobility of the complex formed by this protein with the MEF2 probe was identical to that in the myotube extract (compare lanes 4 and 2). Competition for this binding by the same series of oligonucleotides used in Fig. 5A showed that the relative affinity of the cloned MEF2 for these sites exactly recapitulates that of the endogenous activity (lanes 5-10). In vitro translated MEF2 bound mt4, but not mt6, again identical to the
endogenous muscle specific binding activity (lanes 11-13). The same binding specificity was also reproduced by
in vitro translated aMEF2, the alternative isoform (data not shown). Thus, these cloned factors have a DNA
binding specificity indistinguishable from that of endogenous muscle MEF2.

MEF2 and aMEF2, but Not xMEF2, Bind Multiple Cardiac and Skeletal Muscle Gene Promoter Elements in vitro

The promoters or enhancers of many muscle-specific
genes contain essential A/T rich elements that conform
fully or partially to the MEF2 site consensus (Cserjesi,
et al., Mol Cell Bio, 11:4854-4862, 1991). We used
oligonucleotide probes corresponding to these sequences
(see Table 1) in EMSAs to determine the relative
affinities of the in vitro translated MEF2-related
isoforms (Fig. 5C). Both MEF2 and aMEF2 bound all of the
known or potential MEF2 sites tested, including, in
decreasing order of affinity: the cardiac myosin light
chain 2 promoter HF-1 element; the original MCK enhancer
MEF2 site; a second A/T rich element in the MCK enhancer;
and A/T rich sequences from the promoters for cardiac
troponin T, cardiac α-myosin heavy chain (two distinct
sites), and MHCemb. Remarkably, the myosin light chain 2
HF-1 and α-myosin heavy chain A/T-1 sites have identical
core sequences (TTAAAAATAA) (SEQ ID NO: 33); however, the
former was bound avidly while the latter was bound
poorly, implicating the flanking sequences in site
specification.

In every instance, aMEF2 bound several fold more
effectively than MEF2; thus, the alternative peptides,
which lie well outside the shared MADS domain, must
modulate the binding properties of these proteins.
xMEF2, with a nearly identical MADS domain, bound none of
these sequences in vitro, due either to the few amino
acid substitutions in the N-terminal region or to the
completely divergent C-terminus. Its capacity to activate transcription via these MEF2 sites, however, indicates that it may well bind in vivo (see below).

The MADS Homology Region Alone Is Not Sufficient for DNA Binding

Earlier studies with SRF and MCM1 have shown that the DNA binding function of each factor resides in a domain that includes the MADS homology (Norman, C., et al., Cell, 55:989-1003, 1988; Christ, C., et al., Genes Dev., 5:751-763, 1991). To ascertain whether the same might be true for MEF2, we constructed a set of progressive C-terminal deletions of cloned MEF2 and assayed DNA binding by gel shift (Fig. 5D). Truncated in vitro translation products containing 322, 201, and 104 N-terminal residues retained the capacity to bind DNA. Further deletion to 58 amino acids, at the boundary of the conserved MADS domain, eliminated binding. Therefore, as in other proteins in this family, the DNA binding function of MEF2 includes the MADS homology, but as many as 46 additional residues C-terminal to it are also required. Indeed, as noted above, differences in this region are responsible for the different DNA binding affinities of MEF2 and aMEF2 (see Fig. 5C).

Skeletal as Well as Cardiac and Smooth Muscle Specific DNA Binding Activity Is Due to MEF2/aMEF2

The DNA binding specificity of cloned MEF2 and aMEF2, which faithfully reproduces that of endogenous muscle MEF2 activity, stands in contrast to the ubiquitous distribution of MEF2-related transcripts. In order to investigate the tissue specificity of the MEF2 protein isoforms, we compared nuclear extracts from a variety of cell types in EMSAs with the MCK MEF2 probe (Fig. 6A).

In Fig. 6A nuclear extracts from C2C12 (C2) and Sole8 myoblasts (nb) and myotubes (mt), rat primary
cardiocytes (Card), rat pulmonary artery smooth muscle cells, C3H10T1/2 fibroblasts (10T1/2), HeLa cells, and NIH3T3 cells untransfected (3T3) or transiently transfected with MyoD (3T3+MyoD) were used in EMSA assays in which free MEF2 probe (F) was separated from specifically bound probe (B), or from the nonmuscle complex (H) which migrated more slowly (lower band in HeLa is a nonspecific artifact). In Fig. 6B Antisera raised against cloned MEF2 isoforms demonstrated that these proteins are responsible for the muscle specific MEF2 binding activity shown by EMSA. Immune sera included Anti-MEF2, specific for MEF2 and aMEF2, and Anti-aMEF2, specific for aMEF2. Controls included the corresponding preimmune sera (Pre-MEF2, Pre-aMEF2) or unrelated antisera (Rabbit S, Anti-100kd). Extracts specified in Fig. 6A were also used here, in addition to those of COS cells and rat liver tissue.

The differentiation of skeletal myoblasts to myotubes in both C2C12 (Lanes 2 and 3) and Sol8 (lanes 5 and 6) cells was accompanied by a marked increase in binding (B). Similarly, NIH3T3 fibroblasts normally devoid of this activity developed MEF2 binding upon transient transfection with MyoD (lanes 8 and 9). These data are consistent with previous work documenting the induction of this activity during myogenesis and in response to myogenin (Gossett, et al., Mol Cell Bio, 9:5022-5033, 1989; Cserjesi, et al., Mol Cell Bio, 11:4854-4862, 1991). It is striking that smooth muscle cells and primary cardiocytes which lack known myogenic bHLH factors, also contained specific MEF2 binding activity (lanes 4 and 7). In contrast, cells outside these muscle lineages showed only a slower-migrating complex (H) distinct from the muscle specific complex (C3H10T1/2 fibroblasts and HeLa, lanes 10 and 11; see also below).
Tissue specific expression of MEF2 isoforms was further demonstrated using antisera to define the proteins in MEF2 DNA binding complexes (Fig. 6B). Anti-MEF2 recognizes both MEF2 and aMEF2, while anti-aMEF2 is specific for aMEF2 (see Methods). Both antibodies produced a "supershift" of bound probe, confirming the presence of these factors in C2C12 myotube (lanes 2-8), cardiocyte (lanes 16-18), and smooth muscle cell (not shown) extracts, while preimmune and unrelated controls had no effect. In contrast, the slower-migrating H complex lacks these MEF2 proteins and was not supershifted in HeLa (lanes 9-13), COS (lanes 14 and 15), and liver (lanes 19-21) extracts, nor in C3H10T1/2 or CACO (colon carcinoma) cells (data not shown), confirming that ubiquitous binding of the probe is not due to the cloned factors. A fraction of the H complex from liver extract seems to be supershifted; whether a small amount of MEF2 is expressed in liver tissue or possibly arises from vascular smooth muscle in the organ remains to be determined.

Thus, MEF2 DNA binding activity is found in skeletal, cardiac, and smooth muscle lineages. Note that vascular smooth muscle could account for MEF2-related transcripts in non-muscle tissues, but not in cultured cells. The presence of MEF2 RNAs in cells and tissues outside these lineages indicates that post-transcriptional mechanisms are required to produce absolute tissue specificity of MEF2 DNA binding. Some of this regulation may come from preferential splicing of the MEF2- and aMEF2-specific alternative exons (see Fig. 3), but translational or post-translational mechanisms are likely to operate as well. The antibody supershifts demonstrate unambiguously that tissue specific MEF2 DNA binding activity is directly attributable to the cloned MEF2 gene products. It is particularly interesting here
that anti-aMEF2, which is specific for only one (aMEF2) of the alternative isoforms, supershifted virtually all of the bound probe in these assays. Either these complexes comprise aMEF2 alone, or MEF2:aMEF2 heterodimers that are shifted intact by this antibody.

**The Cloned Factors Are MEF2 Site-Dependent Transcriptional Activators**

A considerable body of data shows that the MEF2 site is critical for tissue specific transcription conferred by muscle gene promoters and enhancers (Gossett, et al., *Mol Cell Bio*, 9:5022-5033, 1991; Zhu et al., *Mol Cell Bio*, 11:2273-2281, 1991; Wentworth, et al., *PNAS*, 88:1242-1246, 1991). Our results to this point correlate tissue specific binding at the MEF2 site with the cloned gene products. To determine whether these proteins are functional transcription factors, we examined their capacity to trans-activate promoters containing MEF2 sites.

MEF2 cDNAs, subcloned into the pMT2 eukaryotic expression vector, were cotransfected with various reporter plasmids in nonmuscle cells. As diagrammed in Fig. 7A, the reporter constructs comprise the bacterial chloramphenicol acetyl transferase (CAT) gene linked to the basal MHCemb promoter (pE102-CAT; Bouvagnet, et al., *Mol Cell Bio*, 7:4377-4389; Yu, et al., *Mol Cell Bio*, 9:1839-1849, 1989), or the HSV thymidine kinase promoter (p8TK-CAT; McKnight, et al., *Science*, 217:316-324, 1982). Each promoter was tested with or without two copies of intact or mutated MEF2 binding sites (M). Parallel experiments in HeLa (Fig. 7A), CV-1, NIH3T3, and C3H10T1/2 (not shown) cells all gave similar results. The MEF2 expression vector pMT2-MEF2 produced marked transcriptional activation of reporters containing the MCK MEF2 binding site (p8TKCAT-MEF2x2,pE102CAT-MEF2x2) or the related A/Temb site from the MHCemb promoter.
pE102CAT-ATembx2). Control experiments with reporter constructs containing MEF2 site mutants (p8TKCAT-MEF2mtx2, pE102CAT-MEF2mtx2) or no MEF2 binding sites (p8TKCAT, pE102CAT) showed that trans-activation by MEF2 depends absolutely on the presence of intact binding sites. The enhanced responsiveness of the MHcemb promoter over the thymidine kinase promoter (26 fold versus 5-6 fold) suggests that MEF2 may interact synergistically with other transcription factors that bind the MHcemb promoter. The pE175CAT reporter containing the native MHcemb promoter, including the single endogenous A/Temb site (-162 to -150) was also activated by MEF2, albeit at a lower level.

Fig. 7A: The various chloramphenicol acetyltransferase (CAT) reporter genes, with and without duplicated wild type or mutated MEF2 binding sites (M), are diagrammed here described in detail in the text. The coordinates of the MHcemb (pE102-CAT) and thymidine kinase (p8TK-CAT) promoters are indicated. The pE175CAT reporter, not diagrammed, is described in the text. HeLa cells were cotransfected individually with these constructs and either the MEF2 cDNA expression plasmid (pMT2-MEF2) or vector control (pMT2), and the results displayed graphically.

Fig. 7B: The same cotransfection experiments were conducted in C2C12 myoblasts and myotubes, rat primary cardiocytes, and rat pulmonary smooth muscle cells.

This is consistent with previous results showing that duplicated MEF2 binding sites are more effective than a single site (Gosset, et al., Mol Cell Bio 2:5022-5033, 1989). Together, these results document that the cloned MEF2 proteins by themselves are sufficient to produce both specific DNA binding and trans-activation in nonmuscle cells. Therefore, the cloned sequences encode
the endogenous factors responsible for MEF2 activity in vivo.

When we performed analogous experiments with aMEF2 and xMEF2 expression constructs, transcription activation by aMEF2 was consistently as good or better than that conferred by MEF2, correlating with the relative binding affinities of the two isoforms (see above, Fig. 5C). xMEF2, which gave no detectable DNA binding in vitro, also conferred lower but reproducible trans-activation in these cotransfection experiments. We infer either that xMEF2 binds DNA in vivo as a heteromeric complex with other unidentified MEF2-related isoforms or unrelated factors, or, less likely, that it potentiates other transcription factors without contacting the DNA itself. Alternatively, the discrepancy between xMEF2 in vitro binding and in vivo trans-activation may be due to the difference between the single copy MEF2 site in the binding probe and the duplicated copies in the reporter genes.

**Skeletal, Cardiac, and Smooth Muscle Cells contain Saturating Levels of Endogenous MEF2 Trans-Activating Factors**

The presence of trans-activating MEF2 activity in skeletal myotubes is well established (Gosset, et al., Mol Cell Bio 9:5022-5033, 1989). Here we have found that specific MEF2 DNA binding activity is present not only in skeletal muscle, but also cardiac and smooth muscle cells, raising the question as to whether all three muscle lineages express endogenous MEF2 transcriptional activity. To investigate this, we performed a series of cotransfection experiments in all three muscle cell types (Fig. 7B). Again, the pE102CAT reporter without binding sites was inactive. Undifferentiated C2C12 myoblasts behaved much as nonmuscle cells (above, Fig. 7A) in that transcription of pE102CAT -MEF2x2 was
increased significantly when cotransfected with pMT2-MEF2. In fused myotubes, however, this reporter construct was already fully active without cotransfection, and there was no appreciable further stimulation when pMT2-MEF2 was added. As shown, the results in primary cardiocytes and pulmonary arterial smooth muscle cells were the same as those in skeletal myotubes, i.e., these cell types, in contrast to nonmuscle cells, already contain saturating levels of MEF2 activity, presumably from endogenous MEF2 itself and/or from its related isoforms.

There is, therefore, an exact correlation between the tissue specific MEF2 DNA binding activity demonstrated in skeletal, cardiac, and smooth muscle (see Fig. 6), and functional trans-activation in these same cell types. Particularly striking is the presence of MEF2 activity in all three muscle lineages. Despite certain phenotypic similarities between smooth and striated muscle tissues, common mechanisms for specific gene regulation in these tissue types have not been previously described.

MEF2 is Induced by MyoD but, Alone, is Not Myogenic

It is clear from these data that the appearance of MEF2 activity is correlated with muscle differentiation.

Given the well-known capacity of MyoD to produce myogenic conversion, it was of interest to investigate the potential interrelationship between MEF2 and MyoD in this process. MyoD, as well as myogenin, induces MEF2 DNA binding activity in transfected fibroblasts (see Fig. 6A; Cserjesi, et al., Mol. Cell. Bio. 11:4854-4862, 1991). We tested whether this coincided with the development of MEF2 trans-activation and, further, whether MEF2 alone is sufficient to generate the muscle phenotype.

MEF2 trans-activation was induced by MyoD in transiently transfected NIH3T3 cells (Fig. 8).
In Fig. 8, NIH3T3 fibroblasts were transiently cotransfected with a MyoD cDNA expression plasmid and the pE102CAT reporter, with or without MEF2 binding sites (see Fig. 7), and assayed for CAT activity following incubation in either low (5% heat-inactivated equine) or high (10% fetal bovine) serum conditions.

Indeed, pE102CAT-MEF2x2 was transcribed at a high level in these cells. This activity was independent of serum concentration in these cultures, indicating that the fully differentiated muscle phenotype associated with serum withdrawal is not required for MEF2 activity in the presence of exogenous MyoD. However, transfected MyoD alone was not sufficient to produce MEF2 activity in HeLa cells (data not shown) which are resistant to myogenin conversion (Weintrab, et al., PNAS, 86:5434-5438, 1989). These results seem to contrast with similar experiments, using myogenin instead of MyoD, where MEF2 DNA binding activity in transfected C3H10T1/2 cells required serum withdrawal, and where this same activity was induced in transfected CV-1 cells which are also resistant to myogenic conversion (Cserjesi, et al., Mol. Cell. Bio. 11:4854-4862, 1991). Cell type differences may be responsible for these apparent discrepancies.

Finding that MEF2 activity was induced by MyoD, we sought to determine whether ectopic expression of MEF2 alone might induce the muscle program in otherwise nonmyogenic cells. In both transient and stable transfection of C3H10T1/2 fibroblast, however, MEF2 failed to induce the muscle phenotype as characterized by myotube formation and striated myosin heavy chain expression (data not shown).

These results define a hierarchy for myogenesis in which MEF2 lies downstream of the muscle specific bHLH factors. MEF2 is induced by MyoD but is not, by itself, myogenic. It is clear, therefore, that MEF2 is not the
sole proximate effector of myogenic conversion by MyoD. Other muscle specific factors must be induced in parallel. Furthermore, the presence of MEF2 activity in cardiac and smooth muscle, in which MyoD and its cognates have not been detected, must be taken as evidence for the existence of alternate pathways for MEF2 induction.

Isolation and Characterization of Other MEF2 Family Members

Genomic southern blotting with a probe from the MEF2 DNA binding domain indicated the existence of several genes containing homology to the probe. These observations led us to postulate that a family of transcription factors containing this conserved domain may be present in muscle in an analogous manner to the MyoD family, and that this protein family may be important for muscle gene regulation based on the functional presence of the MEF2 binding site in many muscle specific genes.

We have screened a skeletal muscle cDNA library at low stringency with a conserved DNA binding domain probe from the first MEF2 related gene isolated in our laboratory, with the purpose of identifying additional members of the putative MEF2 family of transcription factors. We report the isolation and characterization of cDNA's encoding a new MEF2 related factor which is homologous to the initial MEF2 gene but is derived from a separate gene. The products of this gene termed dMEF2, activate transcription. The methods used to isolate dMEF2 can similarly be used to isolate other MEF2 family members.

DMEF2 has a similar binding specificity to the previously isolated MEF2 related factors. Immunofluorescence studies indicate that dMEF2 is developmentally up-regulated in the myoblast to myotube transition and is also present in a subset of neuronal
cell nuclei. There is strict tissue specific transcriptional regulation of this gene, in comparison to
the more ubiquitous expression of the other MEF2 related factors.

CDNA library screening was performed as described above. Screening of the λ+10 library was performed with
random primed $^{32}P$ labeled CDNA (380 bp NsiI-Ddel fragment) from MEF2 that had a specific activity of $1\times10^9$cpm/$\mu$g of
DNA.

Plasmids and transfections

For in vitro transcription, translation and sequencing the cloned dMEF2 CDNA's were subcloned into
pGEM vectors (Promega Corp., Madison, WI). For in vivo expression, the CDNA's were subcloned into pMT2 vector.

The MHC emb CAT reporter construct consisted of 2 copies of the MCK MEF2 sites inserted in a concatemerised
orientation at the -102 position of the MHC emb promoter in plasmid PE102 CAT, as described above. The
oligonucleotide binding sites were also cloned into the
HindIII site of p8TKCAT (Thompson et al. 1991.

*J.Bio.Chem.* 266:22678-22688) and at -109 of the HSV TK
promoter. Transient transfection assays were carried out
as previously described. Briefly, Hela cells were grown
to ~60% confluence, and transfected with the various DNA
expression constructs by calcium-phosphate
coprecipitation. The cells were glycerol shocked 18h
later. After 24 hrs., the media was switched to low
serum media (DME/5% heat inactivated horse serum), cells
were harvested 48hrs. later. Each plate of cells (~5x10^6
cells) was transfected with the following DNA's: 5 $\mu$g of
the appropriate CAT reporter construct, 5 $\mu$g of the pMT2-
dMEF2 construct or the pMT2 vector alone, and 3 $\mu$g of the
pSV $\beta$-gal which served as an internal control for the
transfection efficiency. For the COS cell transfections
20 $\mu$g of the expression construct was used. Cell
extracts were prepared and CAT activity was determined by previously published procedures.

In vitro Transcription and Translation

For in vitro translation MEF2-pGEM7 zF(+)

constructs were linearized with either BamH1 or Pst1 for the full length and truncated translation products respectively. The resulting RNA was translated in vitro using a rabbit reticulocyte lysate according to the manufacturer's suggested conditions (Promega). The in vitro translation products were analyzed by the incorporation of $[^{35}S]$ methionine and a 3 µl aliquot was electrophoresed on a 12% SDS-polyacrylamide gel. After the proteins were resolved the gel was exposed to Enlightning (DuPont) for 30 mins., dried, and autoradiographed.

Cloning of dMEF2

A human adult skeletal muscle cDNA library, constructed in the phage lambda gt10, was screened by low stringency hybridization with a DNA probe which contains the MEF2 DNA-binding domain. Three phage were chosen for further analysis from 67 positives isolated from the $1.5 \times 10^6$ screened, which contained overlapping cDNAs with substantial homology to the DNA binding domain of MEF2. The open reading frame encoded by these cDNA's is highly conserved in the DNA binding domain (~74% identity at the nucleotide level, 99% at the amino acid level) when compared to the other MEF2 factors, but diverges outside of this conserved domain.

The complete sequence of the longest cDNA insert (1.9kb), designated as dMEF2, has one single continuous open reading frame, as shown in Fig. 15. The sequence contains an in frame methionine with upstream stop codons which fits the consensus as a strong initiation site. The dMEF2 cDNAs encode a 465 amino acid polypeptide (isoelectric point - 8.69), with a predicted Mr of 50.3
kd. Amino acid alignment of the predicted amino acid sequences of dMEF2 and MEF2 reveals an overall identity of 66% (Fig. 17), although the conservation at the N-terminus is much greater (83 of 84 residues). In the region where MEF2 and dMEF2 are strongly conserved there exists a striking homology to a number of protein factors that belong to the MDS protein family. dMEF2 contains an 84 amino acid amino (N)-terminus which is highly conserved with the other MEF2 related factors isolated thus far (Fig. 18). The amino-terminal part of this structural motif (aa3-60) contains the MADS box homology in common with the other MADS factors (Fig. 18). The carboxy (C) terminal end (aa 60-86) of this domain diverges from the other MADS factors but is highly conserved in the MEF2 family (Fig. 18), conferring a binding specificity which is sequence specific but distinct from the other MADS box proteins.

After residue 86, dMEF2 and MEF2 diverge considerably (Fig. 17). This diversity after residue 86 corresponds with the divergence between MEF2 and aMEF2 and the existence of an exon boundary at this point. In addition, dMEF2 lacks the glutamine/proline rich region which exists in the C-terminus of MEF2, a region which is a known motif in some transcription factors. Two of the dMEF2 cDNA's are identical except that a 96nt segment (nucleotides 1737-1833) is absent and represents a bona fide splicing variant (Fig. 16).

A prediction of the amino acid secondary structure of the dMEF2 molecule reveals that the binding domain contains a short alpha-helical region (amino acids (aa)1-6) followed by a turn and an extended alpha helix (aa 20-48). The N terminal part of this helix (aa 20-33) is highly hydrophilic and has a high surface probability indicating that it may be involved in dimerization and/or binding to DNA. This region is predicted to be an
amphipathic alpha helix in which the hydrophobic residues are clustered on one side of the helix, a molecular arrangement which stabilizes a coiled-coil structure (Fig. 18). In several other proteins it has been shown that coiled motifs of this nature are important for dimer formation and transcriptional activation (Johnson et al. 1989. *Annu.Rev.Biochem.* 58:799-839; Rasmussen et al. 1991. *Proc.Natl.Acad.Aci.USA* 88:561-564). The C-terminal part of the second helix (aa 34-48) is very hydrophobic indicating that it is probably oriented to the interior facing region of the alpha helix. There is an additional alpha helix within the MEF2 specific region of the binding site from aa 60-69 which is also predicted to be an amphipathic alpha helix (Fig. 18). It is possible that this region is responsible for the binding specificity which distinguishes the MEF2 related factors from the other MADS proteins. There are potential glycosylation sites at aa 49 and 283.

**DNA Binding Site Specificity**

In order to investigate if the dMEF2 protein binds to the MEF2 DNA binding site, protein-DNA interactions were assessed using electrophoretic mobility shift assays. The binding of in vitro translated dMEF2 to a double-stranded (ds) oligodeoxyribonucleotide, comprised of the previously characterized MCK enhancer MEF2 site (Cserjesi et al. 1991. *Mol.Cell.Bio.* 11:4854-4862), was tested. The probe used in the electrophoretic mobility shift assay was a 27bp double stranded, single core recognition motif for the MEF2 site labelled by phosphorylation using T4 polynucleotide kinase and gamma-\(^{32}\)P-labeled MEF2 site ds oligonucleotides and the resulting protein-DNA complex was resolved by gel electrophoresis followed by autoradiography. The specificity of the protein-DNA complex observed between dMEF2 and the labelled MCK MEF2 site was determined by
using various unlabelled synthetic oligonucleotides as competitors. The results of these experiments are consistent with the known specificity of the consensus binding site [CTA(AT)₄ TAG], in that mutant 4 which has a single base change at one of the variant positions in the consensus does bind and effectively compete the specific complex. Conversely, mutants 1 and 6, which have mutations in the invariant region of the binding site, do not effectively compete indicating that they are not bound by dMEF2 with appreciable affinity. As expected, the CArG box binding site, which is a high affinity binding site for the MADS protein SRF, does not complete the binding.

We tested if the presence or absence of the peptide encoded by the 96 nucleotide alternate region in the cDNA’s would influence the DNA binding affinity. However, there was no observable difference between the DNA binding of in vitro translated proteins either with or without this region. In addition, a truncated version of the protein (Amino terminal 178aa, truncated at the PstI site) retained its DNA binding capacity. The data indicate that both the long and truncated forms bind DNA with similar affinity. When the long and truncated forms were co-translated we were not able to observe an intermediary complex which would indicate homodimerization. Taken together, these observations demonstrate a similar in vitro DNA binding specificity that is shared by dMEF2 and the other MEF2 related factors so far isolated.

Transcriptional activation by dMEF2

To determine if dMEF2 could function as an activator of transcription, the dMEF2 cDNA’s were subcloned into an eucaryotic expression vector (pMT2). The dMEF2 containing expression constructs were co-transfected with various reporter constructs containing a
heterologous promoter site and two concatenated copies of the MEF2 high affinity binding site. All transfections were carried out in Hela cells. The reporter constructs used are comprised of the bacterial chloramphenicol acetyl transferase (CAT) gene fused to either: 1) the basal MHC emb promoter (pE102 CAT); 2) the HSV thymidine kinase promoter (TK-CAT); or 3) the SV40 major late promoter (A10-CAT). Control transfections with reporter constructs without the MEF2 binding sites present were not transactivated by the expression constructs indicating that transactivation of the reporter constructs was dependent on the presence of the intact MEF2 binding sites. An interesting result from these experiments is that the most potent transactivation of the reporter constructs was observed with the muscle specific promoter when compared to the two non-muscle specific promoter elements tested. Thus, the cellular context of the promoter element may be important for transactivation by dMEF2.

Deletion of the Carboxy-Terminal Third of dMEF2

Using the transcriptional activation assay described above, we assayed the effect of the presence or absence of the carboxyterminal third of MEF2. We found that the presence of a C-terminal portion constituting about one-third of the molecule substantially reduces or inactivates transcriptional enhancement.

Method of Screening for Molecules that Enhance the activity of a MEF2 protein.

To test for molecules that enhance the activity of the MEF2 proteins we are using in vitro and in vivo assays. The in vitro assay is a modification of the DNA binding assay (retardation gels) described above. The different isoforms of MEF2 produced by expression in bacteria, animal cells, or by in vitro translation are diluted in a progressive fashion until the amount of
protein present in the assay is insufficient, on its own, to generate a retardation of the DNA probe added to the assay. This DNA probe contains the MEF2 DNA consensus binding site, as described above. The different molecules, including other proteins, cell extracts and different types of bacteria, animal cells, or by in vitro translation are diluted in a progressive fashion until the amount of protein present in the assay is insufficient, on its own, to generate a retardation of the DNA probe added to the assay. This DNA probe contains the MEF2 DNA binding site as described above. The different molecules, including other proteins, cell extracts and different types of bacterial or fungal broths are then added to the assay and tested for the appearance of a MEF2 retardation complex. This assay has proven successful in identifying a homeobox-containing protein (mHOX) as an enhancer of MEF2 activity.

An in vivo assay follows the same principle. Limiting amounts of a mammalian expression plasmid driving MEF2 cDNA sequences corresponding to the different isoforms are transfected in limiting amounts into a variety of host cells that do not endogenously have MEF2 activity. In practice we have used HeLa cells and fibroblast cell lines. A concentration of the plasmid that in itself is insufficient to activate a reporter construct that drives a marker enzyme such as CAT (Chloranfenicol acetyl transferase), β-galactosidase, luciferase or any other marker, whose expression is dependent on a intact MEF2 DNA binding site, is used. This plasmid is cotransfected together with the test expression plasmids. The enhancement in the expression of the reporter plasmid is an indication of the enhancing effect of mHOX. The same assay will be used to monitor the effect of cell extract, broths, etc.
on the cells that contain the MEF2 expression plasmid together with the MEF2 reporter constructs.

Use

A MEF2 Transcription factor can be used to produce transgenic animals with increased muscle cell mass, to prevent or counteract muscle atrophy in humans or animals suffering a pathological muscular condition, or to develop pharmacological agents that regulate the expression of muscle-specific genes.

Biological Activity Assay for MEF2 Transcription Enhancement

Transgenic Animals

The transgenic animals being prepared are those that gain muscular function by overexpressing the MEF2 isoforms. The transgenic animals are prepared by pronuclei injection using standard protocols as described by Hogan, B., Constantini, F., and Lacy, E. (1986) Manipulating the Mouse Embryo: A Lab. Manual (CSHL, CSH, NY). These protocols with the necessary modifications will be used to produce transgenic animals of commercially and/or scientifically useful species.

The transgenic animals are being made using complete coding sequences. As the regions important for function modified molecules will be used that produce an enhanced level of activity. The expression of the MEF2 sequences can be targeted to different tissues and stages of development through the used of tissue-and developmental-specific promoter. The embryonic heavy chain promoter can target these sequences to the early developmental stages up to the perinatal age and the β myosin heavy chain promoter that can target the expression of the gene to the slow muscle fiber and the cardiac tissue. These promoters have been isolated and characterized (Strehler, E.E., et al. 1986. J. Mol. Biol.
The certainty that the transgene will be expressed in the transgenic mammal is illustrated by the ability of a MEF2 construct to be expressed in a whole animal. This has been demonstrated by direct injection of the DNA constructs into skeletal and cardiac muscle of interact dogs using a modification of the direct DNA injection described by Wolff (1991 Biotechniques 11:474-485). Expression is also illustrated by the direct intramuscular injection experiments described above. Using this methodology we have shown that it is possible to produce high level expression in cardiac and skeletal muscle of the injected MEF2. This expression lasts for at least 30 days after injection.

**Therapy**

**Administration of a Therapeutic composition by Intramuscular Injection**

The regulated expression of MEF2 genes in vivo was investigated by injecting the gene into the heart of a large mammal in situ. In so doing, a methodology suitable for expressing MEF2 genes in large mammals was developed. The method involves injection of plasmid DNA into canine myocardium.

**Methods**

MSV-CAT was created by fusing the coding sequence of the chloramphenicol acetyl transferase (CAT) gene (Gorman et al. 1982. Mol.Cell.Biol. 2:1044-1051) to the long terminal repeat of the mous sarcoma virus (MSV). RSV-Luciferase was described previously (DeWet et al. 1987. Mol.Cell.Biol. 7:725-737). The series of deletions of the 5′ flanking region of the β-MHC included the -3300rβ-MHC-CAT, -667rβ-MHC-CAT, -354rβ-MHC-CAT and -215rβ-MHC constructs, which are genomic fragments of the rat β-MHC gene from -3300 base pairs (b.p.), -667 b.p., -

Adult mongrel dogs of either sex weighing between 20 and 26 kg were used for these experiments. Dogs were premedicated with xylazine (10mg/kg i.m.) and general anesthesia was induced with thiamylal (10-20mg/kg i.v.) and maintained with halothane (0.5-1.5 vol.%). Observing sterile technique, the pericardium was opened. Circular plasmid DNA resuspended in 20% sucrose and 1 x phosphate buffered saline was injected through a 30 ga needle inserted perpendicular to the epicardium. CAT-assays were performed as previously described (Seed et al. 1988. Gene 67:271-277). Luciferase-assays were also performed as described elsewhere (Brasier et al. 1989. BioTechniques 7(10):1116-1122). All data are reported mean ± standard error of the mean (SEM).

Results

Reporter constructs utilizing the chloramphenicol acetyl transferase (CAT) gene under the control of muscle-specific (β-myosin heavy chain gene (β-MHC)) or promiscous (MSV) promoters were injected into the canine myocardium. Up to 30 separate injection sites were used per left ventricle with no mortality and only transient tachyarrhythmias. There was a linear dose-response relationship between the level of gene expression and the quantity of plasmid DNA injected between 10μg and 200μg. There was no regional variation in expression of injected reporter genes throughout the left ventricular wall. Using both the MSV and a muscle-specific β-MHC promoter reporter gene expression was 1 to 2 orders of magnitude
greater in the heart than in the skeletal muscle. Expression in the left ventricle was 3-fold higher than in the right ventricle. CAT-activity was detected at 3, 7, 14 and 21 days post-injection (p.i.) with maximal expression at 7 days p.i. Statistical analysis of co-injection experiments revealed that co-injection of a second gene construct (RSV-luciferase) is useful to control for transfection efficiency in vivo. Detection of regulatory sequences by injection of reporter constructs containing serial deletions of the β-MHC gene 5' flanking region revealed a pattern of expression that is in general agreement with results obtained in cell culture studies.

Fig. 9 shows a dose-response relation between the amount of injected DNA and CAT-activity. Scatter plot of CAT-activity (in counts per minute/100) per injection site versus total amount of DNA (MSV-CAT) per injection site. Means (±SEM) are shown as solid squares, n=4 for each dose. Linear regression function is shown (P<0.001).

Fig. 10 shows a time course of expression of injected gene constructs. CAT-activity (in counts per minute/1000 versus days post injection for promiscuous (MSV, solid bars) and muscle specific (-667rβ-MHC, hatched bars) promoters driving the CAT reporter gene. Mean ± SEM, n=5 for each time point (*P<0.01 compared with day 7).

Fig. 11 shows a regional expression pattern of injected gene constructs throughout the left ventricular wall. 24 injections of -667rβ-MHC-CAT were performed with 4 columns around the left ventricle each comprising 6 injection sites ranging from base to apex (see cartoon). Means ± SEM of each column are shown on the left hand panel 9 (n=6). Means ± SEM of each row are shown on the right hand panel (n=4).
Fig. 12 shows an expression of promiscuous (MSV) or muscle-specific (-667rβ-MHC) promoter constructs in the right ventricle and in skeletal muscle. Values (mean ± SEM) are depicted as percent of expression of the same construct in the left ventricle (= 100%, solid bars). Open bar is right ventricle (n=10 for MSV, n=8 for -667rβ-MHC). Hatched bar is skeletal muscle (n=10 for MSV, n=9 for -667rβ-MHC).

Fig. 13 shows the correlation of CAT-to Luciferase-activity in co-injection experiments. Scatter plot of CAT-activity (counts per minute) versus luciferase-activity (light units). 100 µg of a tissue-specific (-667rβ-MHC-CAT, Fig. 5a) or promiscuous (MSV, Fig. 5B) reporter gene construct were co-injected with 20 µg of a control gene construct (RSV-Luciferase). The regression functions are as indicated.

Fig. 14 shows the mapping of the 5' flanking region of the β-MHC gene in vivo. A series of deletions of the upstream region of the rat β-MHC gene ranging from -3300 to -215 relative to the transcription start site were cloned in front of the CAT gene and injected into the canine myocardium. For comparison -607α-MHC-CAT and -256 ApoAI-CAT were also injected. 100 µg of reporter gene construct were co-injected with 20 µg of a control gene construct (RSV-Luciferase). CAT-activity was corrected for luciferase-activity and is expressed in percent of MSV-CAT. Open bars are β-MHC-CAT constructs (n=6-10). Hatched bar is the α-MHC-CAT construct (n=10). Solid bar is the Apo AI-CAT construct (n=10). See text for statistical analysis.

Other Modes of Administration of a Therapeutic Composition

The MEF2 polypeptides of the invention can be administered to a mammal, particularly a human, by any appropriate method: e.g., orally, parenterally,
transdermally, or transmucosally. Administration can be in a sustained release formulation using a biodegradable biocompatible polymer, or by on-site delivery using micelles, gels or liposomes. Therapeutic doses can be, but are not necessarily, in the range of 0.001 – 100.0 mg/kg body weight, or a range that is clinically determined as appropriate by those skilled in the art. With the availability of the cloned gene, a substantially pure MEF2 polypeptide can be produced in quantity using standard techniques known to one skilled in the art (see, e.g., Scopes, R. Protein Purification: Principles and Practice, 1982 Springer Verlag, NY).

The nucleic acids of the invention can be administered to a mammal, preferably a human, or a domesticated animal, by techniques of gene therapy. An appropriate recombinant vector, e.g., an attenuated virus, is administered to a patient in a pharmaceutically-acceptable buffer (e.g., physiological saline). The therapeutic preparation is administered in accordance with the condition to be treated. For example, retroviral vectors, can be used as a gene transfer delivery system for a MEF2 polypeptide. Numerous vectors useful for this purpose have been described (Miller, 1990 Human Gene Therapy 1:5-14; Friedman, 1989 Science 244:1275-1281); Eglitis et al. 1988 Biotechniques 6:608-614; Tolstoshev et al. 1990 Current Opinion in Biotechnology 1:55-61; Sharp, 1991 The Lancet 337:1277-1278; Cornetta et al., 1987 Nucleic Acid Research and Molecular Biology 36:311-322; Anderson 1984 Science 226:401-409; Moen, 1991 Blood Cells 17:407-416; and Miller et al. 1989 Biotechniques 7:980-990). Retroviral vectors are particularly well developed and have been used in a clinical setting (Rosenberg et al. 1990 N. Engl. J. Med. 323:370).
The retroviral constructs, packaging cell lines and delivery systems that may be useful for this purpose include, but are not In many cases where it is necessary for the MEF2 polypeptide to enter the nucleus, it may be necessary to employ an attenuated viral vector that naturally replicates and is expressed in the nucleus. Alternatively, the nucleic acid vector can include a nuclear localization region, e.g., two consensus regions consisting of basic amino acids separated approximately 10 "spacer" amino acids. This region is likely to be responsible for directing the transport of this protein from the cytoplasm, where it is produced, to the cellular nucleus (Dingwall, C. and Laskey, R., 1991. Trends in Biochemical Sciences. 16:478-481).

The retroviral constructs, packaging cell lines and delivery systems which may be useful for this purpose include, but are not limited to, one, or a combination of, the following: Moloney murine leukemia viral vector types; self inactivating vectors; double copy vectors; selection marker vectors; and suicide mechanism vectors.

Non viral methods for the therapeutic delivery of nucleic acid encoding a MEF2 polypeptide

Nucleic acid encoding MEF2, or a fragment thereof, under the regulation of the a muscle-cell specific promoter, and including the appropriate sequences required for autonomous replication or for insertion into genomic DNA of the patient, may be administered to the patient using the following gene transfer techniques: microinjection (Wolff et al., Science 247:1465 (1990)); calcium phosphate transfer (Graham and Van der Eb, Virology 52:456 (1973); Wigler et al., Cell 14:725 (1978); Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413 (1987)); lipofection (Felgner et al., Proc. Natl.

**Muscle Cell Specific Expression**

In any of the modes of administration of MEF2 nucleic acids discussed above, e.g., administration by transgenics or gene therapy, the specific expression of MEF2 can be localized to muscle tissue by including the promoters of any of the following genes in the regulatory sequences of the construct to be administered: the MyoD family of genes; myogenin; creatine kinase; the myosin heavy chain gene family; the myosin light chain family; troponins; and tropomyosins.

**Regulation of, and by, the MEF2 family proteins**

The MEF2 genes can be induced in a family of transcription factors called the myogenic determination genes. We have tested two of these for their ability to induce MEF2. Both MyoD1 and myogenin are able to induce the MEF2 genes. On the other hand, MEF2 is able to induce the expression of myogenin. These results indicate that there is a feedback loop by which the two families of myogenic regulators (the MyoD family and the MEF2 family) regulate each other. MEF2 upregulates many known muscle specific genes described to date. These include, but are not limited to, creatine kinase, the myosin heavy chain gene family, the myosin light chain family, troponins, tropomyosins, and various ion channels. A MEF2 protein or nucleic acid of the invention can be administered to a mammal to upregulate, or mask a symptomatic defect in, any of these genes, or
any other as yet uncharacterized genes that include a MEF2 consensus DNA binding sequence in its 5' regulatory sequences.

Other Embodiments

Other embodiments are within the following claims. For example, the invention includes any protein that is substantially homologous to a member of the human MEF2 protein family, and possesses the transcriptional enhancer activity of the MEF2 family. Also included are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions (e.g., washing at 2xSSC at 40 °C with a probe length of at least 40 nucleotides) to a naturally occurring MEF2 family nucleic acid (for other definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6, hereby incorporated by reference); and polypeptides or proteins specifically bound by antisera to a member of the MEF2 protein family, especially by antisera to the active site or binding domain of a member of the MEF2 protein family. The term also includes chimeric polypeptides that include biologically active fragments of the MEF2 protein family.

The invention also includes any biologically active fragment or analog of a member of the MEF2 protein family. By "biologically active" is meant possessing in vivo or in vitro transcriptional activity which is characteristic of the MEF2 -amino acid polypeptide shown in Fig. 2. Since a member of the MEF2 protein family exhibits a range of physiological properties and since such properties may be attributable to different portions of the MEF2 molecule, a useful MEF2 fragment or MEF2 analog is one that exhibits a biological activity in any biological assay for MEF2 activity, as described above.
Most preferably a MEF2 protein fragment or analog possesses 10%, preferably 40%, or at least 90% of the activity of a member of the MEF2 protein family, in any \textit{in vivo} or \textit{in vitro} MEF2 activity assay.

Preferred analogs include MEF2 (or biologically active fragments thereof) whose sequences differ from the wild-type sequence only by conservative amino acid substitutions, for example, substitution of one amino acid for another with similar characteristics (e.g., valine for glycine, arginine for lysine, etc.) or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not abolish the polypeptide's biological activity.

Other useful modifications include those which increase peptide stability. Such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) or D-amino acids in the peptide sequence.

Analogs can differ from a naturally occurring member of the MEF2 protein family in amino acid sequence or in ways that do not involve sequence, or in both. Analogs of the invention will generally exhibit at least 70%, more preferably 80%, more preferably 90%, and most preferably 95% or even 99%, homology with a segment of 20 amino acid residues, preferably more than 40 amino acid residues, or more preferably the entire sequence of a naturally occurring MEF2 polypeptide sequence.

Alterations in primary sequence include genetic variants, both natural and induced. Also included are analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., \( \beta \) or \( \gamma \) amino acids. Alternatively, increased stability may be conferred by cyclizing the peptide molecule, or by exposing the polypeptide to phosphorylation-altering
enzymes, e.g., kinases or phosphatases. Other useful modifications also include in vivo or in vitro chemical derivatization of polypeptides, e.g., acetylation, methylation, phosphorylation, carboxylation, or glycosylation; glycosylation can be modified, e.g., by modifying the glycosylation patterns of a polypeptide during its synthesis and processing or in further processing steps, e.g., by exposing the polypeptide to glycosylation affecting enzymes derived from cells that normally provide such processing, e.g., mammalian glycosylation enzymes; phosphorylation can be modified by exposing the polypeptide to phosphorylation-altering enzymes, e.g., kinases or phosphatases.

In addition to substantially full-length MEF2 polypeptides, the invention also includes biologically active fragments of the MEF2 polypeptides. As used herein, the term "fragment", as applied to a polypeptide, will ordinarily be at least about 20 residues, more typically at least about 40 residues, or preferably at least about 60 residues in length. Fragments of a MEF2 polypeptide can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of a member of the MEF2 protein family can be assessed by methods known to those skilled in the art as described herein. Also included are MEF2 polypeptides containing residues that are not required for biological activity of the peptide, or that result from alternative mRNA splicing or alternative protein processing events.

What is claimed is:
<table>
<thead>
<tr>
<th>Probe/Competitor DNA</th>
<th>Sequence</th>
<th>MEF2 Binding</th>
<th>SEQUENCE ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEF2</td>
<td>5’-CGCTCTAAATACCT-3’</td>
<td>+++</td>
<td>SEQ ID NO: 34</td>
</tr>
<tr>
<td>MEF2mt</td>
<td>CGCTCTAAGGCTAACCCT</td>
<td>-</td>
<td>SEQ ID NO: 35</td>
</tr>
<tr>
<td>MEF2mt4</td>
<td>CGCTCTATAAACCTTCT</td>
<td>+++</td>
<td>SEQ ID NO: 36</td>
</tr>
<tr>
<td>MEF2mt6</td>
<td>CGCTCTAACCCTTCT</td>
<td>-</td>
<td>SEQ ID NO: 37</td>
</tr>
<tr>
<td>5'/Temb</td>
<td>ATTCTATATACTTTTC</td>
<td>+</td>
<td>SEQ ID NO: 38</td>
</tr>
<tr>
<td>ArG</td>
<td>GGGGACCAATAAGGCCA</td>
<td>-</td>
<td>SEQ ID NO: 39</td>
</tr>
<tr>
<td>OTF-2</td>
<td>CCAATGATTTGCATGCTC</td>
<td>-</td>
<td>SEQ ID NO: 40</td>
</tr>
<tr>
<td>MLC2 HF-1</td>
<td>GGGGTAAAAATACCC</td>
<td>+++</td>
<td>SEQ ID NO: 41</td>
</tr>
<tr>
<td>MCK A/T</td>
<td>CTGGTTATATACCA</td>
<td>++</td>
<td>SEQ ID NO: 42</td>
</tr>
<tr>
<td>CTNT A/T</td>
<td>CGGGTTAAATAGCACAA</td>
<td>++</td>
<td>SEQ ID NO: 43</td>
</tr>
<tr>
<td>aMHC A/T-1</td>
<td>CAGATTAAATAACTAA</td>
<td>+</td>
<td>SEQ ID NO: 44</td>
</tr>
<tr>
<td>aMHC A/T-2</td>
<td>AGGAACAAAGGCCCC</td>
<td>+</td>
<td>SEQ ID NO: 45</td>
</tr>
<tr>
<td>Consensus</td>
<td>CTAATAAATTAG</td>
<td></td>
<td>SEQ ID NO: 18</td>
</tr>
</tbody>
</table>

*Table 1. Nucleotide sequences of probes and competitor DNAs used in MEF2 binding assays. Only the core sequences of the d.s. oligonucleotides are shown. (+) and (-) represent positive and negative binding of the probes, respectively (see Figure 5). Nucleotides in bold print conform to the consensus sequence of the MEF2 site as reported by Cserjesi and Olson, Mol Cell Bio, 11:4854-4862, 1991.*
SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Bernardo Nadal-Ginard

(ii) TITLE OF INVENTION: MYOCYTE-SPECIFIC TRANSCRIPTION ENHANCING FACTOR 2

(iii) NUMBER OF SEQUENCES: 45

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Fish & Richardson
(B) STREET: 225 Franklin Street
(C) CITY: Boston
(D) STATE: Massachusetts
(E) COUNTRY: U.S.A.
(F) ZIP: 02110-2804

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
(B) COMPUTER: IBM PS/2 Model 50Z or 55SX
(C) OPERATING SYSTEM: MS-DOS (Version 5.0)
(D) SOFTWARE: WordPerfect (Version 5.1)

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: 07/939,898
(B) FILING DATE: 04 SEP 1992
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: John W. Freeman
(B) REGISTRATION NUMBER: 29,066
(C) REFERENCE/DOCKET NUMBER: 00108/088001

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (617) 542-5070
(B) TELEFAX: (617) 542-8906
(C) TELEX: 200154

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2968
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GAATTTCCTG CAAGGATCAT ATCTAAGTGC ACTTTTTGCT GATACTTCCAT TTCTAGACAT 60
TGAGTCTCAC TCTACCCCCC AGGCTGAAAT GCCAGTTGTG GATCTCGGTT CACTGCAACC 120
TCCGCTTCCA GGTTCAAGTG ATTCTCCTAC CTCAGCCTCC CGAGTAGCTG GGATTACAGG 180
CGCCTGCCAC CATGCTTGGG TGATATTTAT ATTTTTAGTA GAGATGGAGT TTCACCATGT 240
TGCCCAAGCT GTTTCTGAAC TCTGGACCTC AGATCTTGTGA AAAATTTCCA GCTGTAAGCC 300
TTGACTAGA AGCTGAAATA ACAGAGACGT GCTACGATGC ATTAGGTAT TCGAGAAAAAT 360
TAACTTTTGA ATTAATATT TGGAATATAAA GGAAATAAAGG AAAGTTGACT GAAA 414

ATG GGG CGG AAG AAA ATA CAA ATC ACA CGC ATA ATG GAT GAA AGG AAC 462
Met Gly Arg Lys Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Asn
1 5 10

CGA CAG GTC ACT TTT ACA AAG AGA AAG TTT GGA TTA ATG AAG AAA GCC 510
Arg Glu Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala
20 25 30

TAT GAA CTT AGT GTG CTC TGT GAC TGT GAA ATA GCA CTC ATC ATT TTC 558
Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe
35 40 45

AAC AGC TCT AAC AAA CTG TTT CAA TAT GCT AGC ACT GAT ATG GAC AAA 606
Asn Ser Ser Asn Lys Leu Phe Glu Tyr Ala Ser Thr Asp Met Asp Lys
50 55 60

GTT CTT CTC AAG TAT ACA GAA TAT AAT GAA CCT CAT GAA AGC AGA ACC 654
Val Leu Leu Lys Tyr Thr Glu Tyr Asn Glu Pro His Glu Ser Arg Thr
65 70 75 80

AAC TCG GAT ATT GTT GAG GCT CTG AAC AAG AAG GAA CAC AGA GGG TGC 702
Asn Ser Asp Ile Val Glu Ala Leu Asn Lys Glu His Arg Gly Cys
85 90 95

GAC AGC CCA GAC CCT GAT ACT TCA TAT GTG CTA ACT CCA CAT ACA GAA 750
Asp Ser Pro Asp Pro Thr Ser Tyr Val Leu Thr Pro His Thr Glu
100      105     110

GAA AAA TAT AAA AAA ATT AAT GAG GAA TTT GAT AAT ATG ATG CGG AAT 798
Glu Lys Tyr Lys Lys Ile Asn Glu Glu Phe Asp Asn Met Met Arg Asn
115     120     125

CAT AAA ATC GCA CCT GGT CTG CCA CCT CAG AAC TTT TCA ATG TCT GTC 846
His Lys Ile Ala Pro Gly Leu Pro Pro Gln Asn Phe Ser Met Ser Val
130      135     140

ACA GTT CCA GTG ACC ACC CCC AAT GCT TTG TCC TAC ACT AAC CCA GGG 894
Thr Val Pro Val Thr Ser Pro Asn Ala Leu Ser Tyr Thr Asn Pro Gly
145     150     155     160

AGT TCA CTG GTG TCC CCA TCT TTG GCA GCC AGC TCA ACG TTA ACA GAT 942
Ser Leu Val Leu Val Pro Ser Leu Ala Ala Ser Ser Thr Leu Thr Asp
165     170     175

TCA AGC ATG CTC TCT CCA CCT CAA ACC ACA TTA CAT AGA AAT GTG TCT 990
Ser Met Leu Ser Pro Pro Gln Thr Thr Leu His Arg Asn Val Ser
180      185     190

CCT GGA GCT CCT CAG AGA CCA CCA AGT ACT GCC AAT CCA GGT GGG ATG 1038
Pro Gly Ala Pro Gln Arg Pro Ser Thr Gly Asn Ala Gly Gly Met
195     200     205

TTG AGC ACT ACA GAC CTC ACA GTG CCA AAT GGA GCT GGA AGC AGT CCA 1086
Leu Ser Thr Thr Asp Leu Thr Val Pro Asn Gly Ala Gly Ser Ser Pro
210     215     220

GTG GGG AAT GGA TTT GTA AAC TCA AGA GCT TCT CCT AAT TTG ATT GGA 1134
Val Gly Asn Gly Phe Val Asn Ser Arg Ala Ser Pro Asn Leu Ile Gly
225     230     235     240

GCT ACT GGT GCA AAT AGC TTA GCC AAA GTC ATG CCT ACA AAG TCT CCC 1182
Ala Thr Gly Ala Asn Ser Leu Gly Val Met Pro Thr Lys Ser Pro
245     250     255

CCT CCA CCA GTT GTG GTT GAT CTT GGA ATG AAC AGT AGG AAA CCA GAT 1230
Pro Pro Pro Gly Gly Gly Asn Leu Gly Met Asn Ser Arg Lys Pro Asp
260     265     270

CTT CGA GTT GTC ATC CCC CCT TCA AGC AAG GCC ATG ATG CCT CCA CTA 1278
Leu Arg Val Val Ile Pro Pro Ser Ser Lys Gly Met Met Pro Pro Leu
275     280     285

TCG GAG GAA GAG GAA TTG GAG TTG AAC ACC CCA AGG ATC AGT AGT TCT 1326
Ser Glu Glu Glu Glu Leu Glu Leu Asn Thr Gln Arg Ile Ser Ser Ser
290     295     300

CAA GCC ACT CAA CCT CTT GCT ACC CCA GTC GTG TCT GTG ACA ACC CCA 1374
Gln Ala Thr Gln Pro Leu Ala Thr Pro Val Val Ser Val Thr Thr Pro
305 310 315 320

AGC TTG CCT CCG CAA GGA CTT GTG TAC TCA GCA ATG CCG ACT GCC TAC 1422
Ser Leu Pro Pro Gln Gly Leu Val Tyr Ser Ala Met Pro Thr Ala Tyr
325 330 335

AAC ACT GAT TAT TCA CTG ACC AGC GCT GAC CTG TCA GCC CTT CAA GGC
Asn Thr Asp Tyr Ser Leu Thr Ser Ala Asp Leu Ser Ala Leu Gln Gly
340 345 350

TTC AAC TCG CCA GGA ATG CTG TCG CTG GGA CAG GTG TCG GCC TGG CAG 1518
Phe Asn Ser Pro Gly Met Leu Ser Leu Gly Gln Val Ser Ala Trp Gln
355 360 365

CAG CAC CAC CTA GGA CAA GCA GCC CTC AGC TCT CTT GTC GCT GGA GGG
Gln His His Leu Gly Gln Ala Ala Leu Ser Ser Leu Val Ala Gly Gly
370 375 380

CAG TTA TCT CAG GCT TCC AAT TTA TCC ATT AAT ACC AAC CAA AAG TTC 1614
Gln Leu Ser Gln Gly Ser Asn Leu Ser Ile Asn Thr Asn Gln Asn Ile
385 390 395 400

AGC ATC AAG TCC GAA CCG ATT TCA CCT CCT CGG GAT CGT ATG ACC CCA
Ser Ile Lys Ser Pro Ile Ser Pro Pro Arg Asp Arg Met Thr Pro
405 410 415

TCG GCC TTC CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CAG CCC CCG
Ser Gly Phe Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Pro
420 425 430

CCA CCA CCG CAG CCC CAG CCA CAA CCC CCG CAG CCC CAG CCC CAG CCA
Pro Pro Pro Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Arg Gln
435 440 445

GAA ATG GGG CGC TCC CCT GTG GAC AGT CTG AGC AGC TCT AGT AGC TCC 1806
Glu Met Gly Arg Ser Pro Val Asp Ser Leu Ser Ser Ser Ser Ser Ser
450 455 460

TAT GAT GCC AGT GAT CGG GAT GAT CCA CGG GGC GAC TCC CAT TCT CCA 1854
Tyr Asp Gly Ser Asp Arg Glu Asp Pro Arg Gly Asp Phe His Ser Pro
465 470 475 480

ATT GTG CTT GGC CGA CCC CCA AAC ACT GAG GAC AGA GAA AGC CCT TCT 1902
Ile Val Leu Gly Arg Pro Pro Asn Thr Glu Asp Arg Glu Ser Pro Ser
485 490 495

GTA AAG CGA ATG AGG ATG GAC GCG TGG GTG ACC TAA 1938
Val Lys Arg Met Arg Met Asp Ala Trp Val Thr
500 505

GGCTTCAAAG CTGATGTTTG TACTTTTGTC TTACTGCAGT GACCTGCCCT ACATATCTAA 1998
ATCGGTAAAT AAGGACATGA GTTAAATATA TTTATATGTA CATACATATA TATATCCCTT 2058
TACATATATA TGTATGTGGG TGTTGATGTG TGTTATATGTG TGGGTGTGTG TTACATACAC 2118
AGAATCAGGC ACTTACCTGC AAACCTCCTG TAGGTCTGCA GATGTGTGTC CCATGGCAGA 2178
CAAAGCACCC TGTAGGCACA GACAAGCTG GCACCCCTCTT GGACTACCTG TTTCTGTAAG 2238
ATAACCAGTT TTTGCAGAGA AACGTGTACC CATATATAAT TCTCCCACAC TAGCTTGCAG 2298
AAACCTAGAG GGGCCCTAC TTGTATTTAT TAACCTGTGCA GTGACTGTAG TTACTTAAGA 2358
GAAAAATGCTT TGTAGAAACAG AGCAGTAGAA AAGCAGGAAC CAAGGAAACCA ATACTGTACA 2418
TAAAATGTCA TTTATATTTT CCCACCTGGC ATGGGTGTCT GTTGCAAAGG GGTGCATGAG 2478
AAAGGGCTGT TGATATATAA AAACAACAAA ACAAAAAAGC CCCACACATA ACTGTTTGC 2538
ACGTGCAAAA ATGTATTGGG TCAAGAAGTG ATCTTTAGCT AATAAGAAAA GAGAATAGAA 2598
AACACGCTAG ATGATTTTAG AAAATACTAG CCTAGAAATA TAGGCAATTT ACAAAGAAA 2658
ATTAATATAT TAAGTTTATAA TTGGAATATG TCGAGAAGTT CTGTATATTT TCTATATTAA 2718
AAAAATGAA AAACCTGATT TAGCTCATGT ATATTTTATA TGAAGAAAAA CACCTTATG 2778
AATTGATGAC TATATATAAA ATTATATCTA CTACCTTTGA ACACATCTTG CTATGAAATT 2838
TTTATATAG CCAAAGCTAT ATGTTGTAAC TTTTTTTATG AGAAATAGCTT TATCTTGGTT 2898
TAACTCTTTA GTTTATATTT AAGAGGGAAGA AACAATAATA TCTTGCAAAGC AGAACCTTTA 2958
AAAAA AAAAAA 2968

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 55
(B) TYPE: amino acid
(C) STRANDEDNESS: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Arg Arg Lys Gln Pro Ile Arg Tyr Ile Glu Asn Lys Thr Arg Arg His
5 10 15

Val Thr Phe Ser Lys Arg Arg His Gly Ile Met Lys Lys Ala Tyr Glu
20 25 30
Leu Ser Val Leu Thr Gly Ala Asn Ile Leu Leu Leu Ile Leu Ala Asn
35
40
45
Ser Gly Leu Val Tyr Thr Phe
50

<table>
<thead>
<tr>
<th>INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) SEQUENCE CHARACTERISTICS:</td>
</tr>
<tr>
<td>(A) LENGTH: 1500</td>
</tr>
<tr>
<td>(B) TYPE: nucleic acid</td>
</tr>
<tr>
<td>(C) STRANDEDNESS: double</td>
</tr>
<tr>
<td>(D) TOPOLOGY: linear</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:</th>
</tr>
</thead>
<tbody>
<tr>
<td>CGGACGCGGA GATGCGACTC AAGGGGAAGA AAGCGCCCGTG AAGAACCTGG TGGACAGCAG</td>
</tr>
<tr>
<td>CGTCTACTTC CGCAGCGTGG AGGGCTGCT CAAACAGGCC ATCAGCAGC GGGACCATAT</td>
</tr>
<tr>
<td>GAATGACGAT GCCGAGGGCC ACAGGGCGGG GGAACCCACCC CCGCCCTCTT CAGCCGTGATC</td>
</tr>
<tr>
<td>CTGGAAAGAGA CTCGCGGCCC CCCAGCCTCC GCGAACCACG ACAAGATCA TTCCACTCAG</td>
</tr>
<tr>
<td>CCTGGGACG</td>
</tr>
<tr>
<td>ATG GGG AGG AAA AAA ATC CAG ATC TCC CGC ATC CTG GAC CAA AGG AAT</td>
</tr>
<tr>
<td>Met Gly Arg Lys Lys Ile Gln Ile Ser Arg Ile Leu Asp Gln Arg Asn</td>
</tr>
<tr>
<td>1 5 10 15</td>
</tr>
<tr>
<td>CGG CAG CTG ACG TTC ACC AAG CGG AAG TTC GGG CTG ATG AAG AAG GCC</td>
</tr>
<tr>
<td>Arg Gin Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala</td>
</tr>
<tr>
<td>20 25 30</td>
</tr>
<tr>
<td>TAT GAG CTG AGC GTG CTC TGT GAC TGT GAG ATA GCC CTC ATC ATC TTC</td>
</tr>
<tr>
<td>Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe</td>
</tr>
<tr>
<td>35 40 45</td>
</tr>
<tr>
<td>AAC AGC GCC AAC CGC CTC TCC CAG TAT GCC AGC ACG GAC ATG GAC CGT</td>
</tr>
<tr>
<td>Asn Ser Ala Asn Arg Leu Phe Gin Tyr Ala Ser Thr Asp Met Asp Arg</td>
</tr>
<tr>
<td>50 55 60</td>
</tr>
<tr>
<td>GTG CTG CTG AGG TAC ACA GAG TAC AGC GAG CCC CAC GAG AGC CGC ACC</td>
</tr>
<tr>
<td>Val Leu Leu Lys Tyr Thr Glu Tyr Ser Glu Pro His Glu Ser Arg Thr</td>
</tr>
<tr>
<td>65 70 75 80</td>
</tr>
<tr>
<td>AAC ACT GAC ATC CTC GAG ACG CTG AAG CGG AGG GCC ATT GCC CTC GAT</td>
</tr>
<tr>
<td>Asn Thr Asp Ile Leu Glu Thr Leu Lys Arg Arg Gly Ile Gly Leu Asp</td>
</tr>
<tr>
<td>85 90 95</td>
</tr>
</tbody>
</table>
GGG CCA GAG CTG GAG CGG CAT GAA GGG CCT GAG GAG CCA GGA GAG AAG 585
Gly Pro Glu Leu Glu Pro Asp Glu Gly Pro Glu Leu Pro Gly Glu Lys 100 105 110
TTT CGG AGG CTG GCA GCC GAA GGG GGT GAT CGG GCC TTC CCC CGA CCC 633
Phe Arg Arg Leu Ala Gly Glu Gly Gly Asp Pro Ala Leu Pro Arg Pro 115 120 125
GGG CTG TAT CCT GCA GCT CCT GCT ATG CCC AGC CCA GAT GTG GTA TAC 681
Arg Leu Tyr Pro Ala Ala Pro Ala Met Pro Ser Pro Asp Val Val Tyr 130 .135 140
GGG GCC TTA CCC CCA CCC GCC TGT GAC CCC AGT GGG CTT GGG GAA GCA 729
Gly Ala Leu Pro Pro Pro Gly Cys Asp Pro Ser Gly Leu Gly Glu Ala 145 150 155 160
CTG CCC GCC CAG AGC GCC CTG GCA TCT CCC TTC CGA CCA GCA GCC CCC AAA 777
Leu Pro Ala Glu Ser Arg Pro Ser Pro Phe Arg Pro Ala Ala Pro Lys 165 170 175
GCC CCC GCC CCC CGT CTG GCC CTG CAC CTT CTC TCC TCA CCA AGC CAC CTC 825
Ala Gly Pro Pro Gly Leu Val His Pro Leu Phe Ser Pro Ser His Leu 180 185 190
ACC AGC AAG ACA CCA CCC CCA CTG TAC CTG CCG ACG GAA GGG CGG AGG 873
Thr Ser Lys Thr Pro Pro Pro Pro Tyr Leu Pro Thr Glu Gly Arg Arg 195 200
TCA GAC CTG CCT GGT GCC CTG GCT GGG CCC CCA GGG GGA CTA AAC ACC 921
Ser Asp Leu Pro Gly Gly Leu Ala Gly Pro Arg Gly Gly Leu Asn Thr 210 215 220
TCC AGA AGC CTC TAC AGT GCC CTG CAG AAC CCC TGC TCC ACT GCA ACT 969
Ser Arg Ser Leu Tyr Ser Ser Gly Leu Gln Asn Pro Cys Ser Thr Ala Thr 225 230 235 240
CCC GGA CCC CCA CTG GGG AGC TTC CCC TTC CTC CCC GGA GCC CCC CCA 1017
Pro Gly Pro Pro Leu Gly Ser Phe Pro Phe Leu Pro Gly Gly Pro Pro 245 250 255
GTG GGG GCC GAA GCC TGG GCC AGG AGG GTC CCC CAA CCC GGG GGC CCT 1065
Val Gly Ala Glu Ala Trp Ala Arg Arg Val Pro Gln Pro Ala Ala Pro 260 265 270
CCC CGC CGA CCC CCC CAG TCA GCA TCA AGT CTG AGC GCC TCT CTC CGG 1113
Pro Arg Pro Pro Gln Ser Ala Ser Ser Leu Ser Ala Ser Leu Arg 275 280 285
CCC CCG GGG GCC CCC GGG ACT TTC CTA AGA CCT CCC CCT ATC CCT TGC 1161
Pro Pro Gly Ala Pro Ala Thr Phe Leu Arg Pro Ser Pro Ile Pro Cys 290 295 300
TCC TCG CCC GGT CCC TGG CAG AGC CTC TGC GGC CTG GCC CCG CCC TGC
Ser Ser Pro Gly Pro Trp Gln Ser Leu Cys Gly Leu Gly Pro Pro Cys
305       310       315       320

GCC GGC TGC CCT TGG CCG ACG GCT GGC CCC GGT AGG AGA TCA CCC GGT
Ala Gly Cys Pro Trp Pro Thr Ala Gly Pro Gly Arg Arg Ser Pro Gly
325       330       335

GCC ACC AGC CCA GAG CGC TTC CCA GGT ACG GCG AGG GCA CTG GGG GAC
Gly Thr Ser Pro Glu Arg Ser Pro Gly Thr Ala Arg Ala Arg Gly Asp
340       345       350

CCC ACC TCC CTC CAG GCC TCT TCA GAG AAG ACC CAA CAG TGA
Pro Thr Ser Leu Gln Ala Ser Ser Glu Lys Thr Gln Gln
355       360       365

CGCCCCCCTC CGCGGTGGGG GCTTGGAGGT GGCGCGCTGG ACTCAATCCA CCCTGGGAGG
1305
CTCTTTTCCT TCTTCCTATT TGTTGTATA TCCACAATA AAACGCGCGT GGCGTGCTTG
1467
GACCAGAAAA AAAAAAAAAA AAAAAAAAAA AAA
1500

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER:  4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH:  2161
(B) TYPE:  nucleic acid
(C) STRANDEDNESS:  double
(D) TOPOLOGY:  linear

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AAGATGCATG GTGAAATGGT GAGTCAGAA AGGGGCCCTG CATTGATAA AAATGAAAAA 60
AACAAATAAA AAATAGCATG AAAGAAACTA GTATATACAA TGGATGTCAG TGGACCAAAT 120
AGATTGCTAA TGATTTAAA ACAATTTAGG GTGTTGCAAT GTGATATGTT CTAACCCCAC 180
AGGTATCTCT TTTGACAGCT GACCTTAAAC TTATAAAATG TAAGCAGAGT AAAAGAAAAC 240
AGANAGAANA TAGTTACTCA AATGTCGAAAT CACCAAAATA TACCCCCCTC CCGCTATTTA 300
GATAACAATAA CTTCTGCTAT TACCATAATA TTATATATAT TAGAAAGCTA TACACAAGCA 360
TGTTAATTCTT AAAGATT TCTTAATTT TTTATAAATTA TATAATAATTA GAAATACACA 420
CATTTCAAAA ACAAACTTCT ACAAGAGAA AAGATTTATC TTGTTAAGGA AAGCATGGAG 480
TTCTTCATGG CTTAGGGTAG TGCTTTCTAT ACACAAAGTC CTTTTTGTTT TTTTACAGGA
CTGTTPAAAAA TATTAGCGAC GCTATCAAGG AAAAAATACA TAATTTCAAGG GACGAGAGAA
AGAAAAAGGA GAAAAAATA CATAAATTTCG GGGACGAGAG AGAGAAAGAA AAGGGGGGACT

ATG GGG AGA AAA AAG ATT CAG ATT AGG ATT ATG GAT GAA CGT AAC
Met Gly Arg Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Asn
1 5 10 15

AGA CAG GTG ACA TTG ACA AAG AGG AAA TTT GGG TTG ATG AAG AAG GCT
Arg Glu Val Thr Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala
20 25 30

TAT GAG CTG AGC GTG CTG TGT GAC TGT GAG ATT GCG CTG ATC ATC TTC
Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe
35 40 45

AAC AGC ACC AAC AAG CTG TTC CAG TAT GCC AGC ACC GAC ATG GAC AAA
Asn Ser Thr Asn Lys Leu Phe Glu Tyr Ala Ser Thr Asp Met Asp Lys
50 55 60

GTG CTT CTC AAG TAC ACG GAG TAC AAC GAG CCC CAT GAG AGC CGG ACA
Val Leu Leu Lys Tyr Thr Glu Thr Tyr Asn Glu Pro His Glu Ser Arg Thr
65 70 75 80

AAC TCA GAC ATC CTG GAG AGC TTG AGA AAG AAG GGC CTT AAT GGC TGT
Asn Ser Asp Ile Val Glu Thr Leu Arg Lys Gly Leu Asn Gly Cys
85 90 95

GAC AGC CCA GAC CCC GAT GCG GAC GAT TCC GTA GGT CAC AGC CCT GAG
Asp Ser Pro Asp Pro Asp Ala Asp Asp Val Gly His Ser Pro Glu
100 105 110

TCT GAG CAC ATT AGG AAT ATT AAC GAA GAT ATT GAT CTA ATG ATC
Ser Glu Asp Lys Tyr Arg Lys Asn Glu Asp Ile Asp Leu Met Ile
115 120 125

AGC AGG CAA AGA TTG TGT GCT GTC GTC GAT ACC TCC GAC ATG CCA
Ser Arg Gln Arg Leu Cys Ala Val Pro Pro Asn Phe Glu Met Pro
130 135 140

GTC TCC ATC CCA GTG TCC AGC CAC AAC AGT TTG GTG TAC AGC AAC CCT
Val Ser Ile Pro Val Ser Ser His Asn Ser Leu Val Tyr Ser Asn Pro
145 150 155 160

GTC AGC TCA CTG GGA AAC CCC AAC CTA TTG CCA GTG GCT CAC CCT TCT
Val Ser Ser Leu Gly Asn Pro Asn Leu Leu Pro Leu Ala His Pro Ser
165 170 175

CTG CAG AGG AAT AGT ATG TCT CCT GGT GTA ACA CAT CGA CCT CCA AGT
Leu Gln Arg Asn Ser Met Ser Pro Gly Val Thr His Arg Pro Pro Ser
180 185 190
GCA GGT AAC ACA GGT GGT CTG ATG GGT GGA GAC CTC ACG TCT GGT GCA
Ala Gly Asn Thr Gly Gly Leu Met Gly Gly Asp Leu Thr Ser Gly Ala
195  200  205

GGC ACC AGT GCA GGG AAC GGG TAT GCC AAT CCC CGA AAC TCA CCA GGT
Gly Thr Ser Ala Gly Asn Gly Tyr Gly Asn Pro Arg Asn Ser Pro Gly
210  215  220

CTG CTG GTC TCA CCT GGT AAC TTG AAC AAG AAT ATG CAA GCA AAA TCT
Leu Leu Val Ser Pro Gly Asn Leu Asn Lys Asn Met Gln Ala Lys Ser
225  230  235  240

CCT CCC CCA ATG AAT TTA GGA ATG AAT AAC CGT AAA CCA GAT CTC CGA
Pro Pro Pro Met Asn Leu Gly Met Asn Arg Lys Pro Asp Leu Arg
245  250  255

GTT CTT ATT CCA CCA GGC AGC AAG AAT ACG ATG CCA TCA GTG AAT CAA
Val Leu Ile Pro Pro Gly Ser Lys Asn Thr Met Pro Ser Val Asn Gln
260  265  270

AGG ATA AAT AAC TCC CAG TCG GCT CAG TCA TTG GCT ACC CCA GTG GTT
Arg Ile Asn Ser Gln Ser Ala Gln Ser Leu Ala Thr Pro Val Val
275  280  285

TCC GTA GCA ACT CCT ACT TTA CCA GGA CAA GGA ATG GGA GGA TAT CCA
Ser Val Ala Thr Pro Thr Leu Pro Gly Gln Gly Met Gly Gly Tyr Pro
290  295  300

TCA GCC ATT TCA ACA ACA TAT GGT ACC GAG TAC TCT CTG AGT AGT GCA
Ser Ala Ile Ser Thr Thr Tyr Gly Thr Glu Tyr Ser Leu Ser Ser Ala
305  310  315  320

GAC CTG TCA TCT CTG TCT GGG TTT AAC ACC GCC AGC GCT CTT CAC CTT
Asp Leu Ser Leu Ser Leu Ser Gly Phe Asn Thr Ala Ser Ala Leu His Leu
325  330  335

GTT TCA GTA ACT GCC TGG CAA CAG CAA CAC CTA CAT AAC ATG CCA CCA
Gly Ser Val Thr Gly Trp Gln Gln Gln His Leu His Asn Met Pro Pro
340  345  350

TCT GCC CTC AGT CAG TTT GGA GCT TGG ACT AGC ACT CAT TTA TCT CAG
Ser Ala Leu Ser Gln Leu Gly Ala Cys Thr Ser Thr His Leu Ser Gln
355  360  365

AGT TCA AAT CTC TCC CTG CCT ACT CAT ACG TCT AAG TCA
Ser Ser Asn Leu Ser Leu Pro Ser Thr Gln Ser Leu Asn Ile Lys Ser
370  375  380

GAA CCT GTT CCT CCT AGA GAC CGT ACC ACC CCT TGC AGA TAC
Glu Pro Val Ser Pro Pro Arg Asp Arg Thr Thr Thr Pro Ser Arg Tyr
385  390  395  400
- 69 -

CCA CAA CAC ACG CGC CAT CAG GCG GGA AGA TCT CCT GTT GAC AGC TTG 1908
Pro Gln His Thr Arg His Glu Ala Gly Arg Ser Pro Val Asp Ser Leu 405 410 415

AGC AGC TCT AGC AGT TCG TAC GAC GGG AGC GAC CGA GAG GAT CAC CGG 1956
Ser Ser Cys Ser Ser Ser Tyr Asp Gly Ser Asp Arg Glu Asp His Arg 420 425 430

AAC GAA TTC CAC TCC CCC ATT GGA CTC ACC AGA CCT TCG CCG GAC GAA 2004
Asn Glu Phe His Ser Pro Ile Gly Leu Thr Arg Pro Ser Pro Asp Glu 435 440 445

AGG GAA AGT CCC TCA GTC AAG CGC ATG CGA CTT TCT GAA GGA TGG GCA 2052
Arg Glu Ser Pro Ser Val Lys Arg Met Arg Leu Ser Glu Gly Trp Ala 450 455 460

ACA 2055
Thr 465

TGATCAGATT ATTACTTACT AGTTTTTTTT TTTCTCTTGC AGTGTGTTG TGTTTACCT 2115

TAATGGGGAA GGGGTCGA TATGCATTAT ATGTGCCGGT TGTTGGA 2161

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 465
(B) TYPE: amino acid
(C) STRANDEDNESS: linear
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

Met Gly Arg Lys Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Asn 5 10 15
Arg Gln Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala 20 25 30
Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe 35 40 45
Asn Ser Thr Asn Lys Leu Phe Glu Tyr Ala Ser Thr Asp Met Asp Lys 50 55 60
Val Leu Leu Lys Tyr Thr Glu Tyr Asn Glu Pro His Glu Ser Arg Thr 65 70 75 80
Asn Ser Asp Ile Val Glu Thr Leu Arg Lys Lys Gly Leu Asn Gly Cys
85 90
Asp Ser Pro Asp Pro Asp Ala Asp Asp Ser Val Gly His Ser Pro Glu
100 105 110
Ser Glu Asp Lys Tyr Arg Lys Ile Asn Glu Asp Ile Asp Leu Met Ile
115 120 125
Ser Arg Gln Arg Leu Cys Ala Val Pro Pro Pro Asn Phe Glu Met Pro
130 135 140 145
Val Ser Ile Pro Val Ser Ser His Asn Ser Leu Val Tyr Ser Asn Pro
145 150 155
Val Ser Ser Leu Gly Asn Pro Asn Leu Leu Pro Leu Ala His Pro Ser
160 165 170 175
Leu Gln Arg Asn Ser Met Ser Pro Gly Val Thr His Arg Pro Pro Ser
180 185 190
Ala Gly Asn Thr Gly Gly Leu Met Gly Gly Asp Leu Thr Ser Gly Ala
195 200 205
Gly Thr Ser Ala Gly Asn Gly Tyr Gly Asn Pro Arg Asn Ser Pro Gly
210 215 220
Leu Leu Val Ser Pro Gly Asn Leu Asn Lys Asn Met Gln Ala Lys Ser
225 230 235
Pro Pro Pro Met Asn Leu Gly Met Asn Asn Arg Lys Pro Asp Leu Arg
240 245 250 255
Val Leu Ile Pro Pro Gly Ser Lys Asn Thr Met Pro Ser Val Asn Gln
260 265 270
Arg Ile Asn Asn Ser Gln Ser Ala Gln Ser Leu Ala Thr Pro Val Val
275 280 285
Ser Val Ala Thr Pro Thr Leu Pro Gly Gln Gly Met Gly Gly Tyr Pro
290 295 300
Ser Ala Ile Ser Thr Tyr Gly Thr Glu Tyr Ser Leu Ser Ser Ala
305 310 315
Asp Leu Ser Ser Leu Ser Gly Phe Asn Thr Ala Ser Ala Leu His Leu
320 325 330 335
Gly Ser Val Thr Gly Trp Gln Gln Gln His Leu His Asn Met Pro Pro
340 345 350
Ser Ala Leu Ser Gln Leu Gly Ala Cys Thr Ser Thr His Leu Ser Gln
355 360 365
Ser Ser Asn Leu Ser Leu Pro Ser Thr Glu Ser Leu Asn Ile Lys Ser
370
Glu Pro Val Ser Pro Pro Arg Asp Arg Thr Thr Thr Pro Ser Arg Tyr
385
Pro Gln His Thr Arg His Glu Ala Gly Arg Ser Pro Val Asp Ser Leu
400
Ser Ser Cys Ser Ser Ser Tyr Asp Gly Ser Asp Arg Glu Asp His Arg
420
Asn Glu Phe His Ser Pro Ile Gly Leu Thr Arg Pro Ser Pro Asp Glu
435
Arg Glu Ser Pro Ser Val Lys Lys Met Arg Leu Ser Glu Gly Trp Ala
450
Thr
465

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 371
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TGGCGGGCCC GGGCTGCGGC GTGTGCGCGC CCGCCAGCTG CTCCGGAGAT ACGGAATTGC
60
ATTTTGTGAA AAAAGAACAA GAATTTATCTG CAAGGATCAT ATCTAAGTGC ACTTTTTGCT
120
GATACTTCAT TTCTAATCTT GTAGAAAATT TCAGCTGTAG CCCTTGGACT AGAAGCTGAA
180
ATAACAGAAG CTGTGTACGA TGCATTAGGG TATTGAGAA AATTAACCTTT TGAATTAAT
240
ATTTGGAATA TAAGGAATA AGGAAAAGTG ACTGAAAATG GGGCGGAAGA AAATAACAAAT
300
CACACGCATA ATGGATGAAA GGAACCGACA GGTCACTTTT ACAAGAGAAA AGTGGGATT
360
AATGAAGAAAA G
371

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2950
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAATTTTTCTG CAAGGATCAT ATCTAAATGC ACTTTTTGCT GATACTTCAT TTCTAGACAT 60
TGAGTCTCAC TCTACCCCCC AGGCTGAAGT GCAGTGTTGT GATCTCGGTT CACTGCAACC 120
TCCGCTCTCA GGTTCAGTGT ATTCTCGTAC CTCAGCCTCC CGAGTACCTG GGATTACAGG 180
CCGCTGCCAC CATGGCTGGC TGATATTTAT ATTTTTAGTA GAGATGGAGT TTCACCAGCT 240
TGGCCAGGCT GGTCTCGAAC TCTGGACCTC AGATCTTGTG AAAATTTCA GCTGTAGCC 300
TTGGAACGAGA AGCTGAAATA ACAGAAGCTG TGTACGATGC ATTAAGGTAT TGAAGAAAAT 360
TAACTTTTGA ATTAAATATT TGGAGTATAAA GAAATAAAGG AAAGTTGACT GAAA 414

ATG GGG CCG AAG AAA ATA CAA ATC ACA GCC ATA ATG GAT GAA AGG AAC 462
  Met Gly Arg Lys Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Asn
1      5     10    15

CGA CAG GTC ACT TTT ACA AAG AGA AAG TTT GGA TTA ATG AAG AAA GCC 510
  Arg Gln Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys lys Ala
20     25    30

TAT GAA CTT AGT GTC TCT TGT GAC TGT GAA ATA GCA CTC ATC ATT TTC 558
  Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe
35     40    45

AAC AGC TCT AAC AAA CTT CAA TAT GCT AGC ACT GAT ATG GAC AAA 606
  Asn Ser Ser Asn Lys Leu Phe Gln Tyr Ala Ser Thr Asp Met Asp Lys
50     55    60

GTT CTT CTC AAG TAT ACA GAA TAT AAT GAA CCT CAT GAA AGC AGA ACC 654
  Val Leu Leu Lys Tyr Thr Glu Tyr Asn Glu Pro His Glu Ser Arg Thr
65     70    75    80

AAC TCG ACT TTA AGA AAG AAA GGC CTT AAT GGT TGT GAG AGC CCT GAT 702
  Asn Ser Thr Leu Arg Lys Gly Leu Asn Gly Cys Glu Ser Pro Asp
85     90    95

GCT GAC GAT TAC TTT GAG CAC AGT CCA CTC TCG GAG GAC AGA TTT AGC 750
  Ala Asp Asp Tyr Phe Glu His Ser Pro Leu Ser Glu Asp Arg Phe Ser
100   105   110

AAA CTA AAT GAA GAT AGT GAT TTT ATT TCC AAA CGA GGC CCT CCT GGT 798
  Lys Leu Asn Glu Asp Ser Asp Phe Ile Phe Lys Arg Gly Pro Pro Gly
115   120   125
CTG CCA CTT CAG AAC TTT TCA ATG TCT GCA ACA CTT CCA GTG ACC AGC
  Leu Pro Pro Glu Asn Phe Ser Met Ser Val Thr Val Pro Val Thr Ser
  130 135 140

CCC AAT GCT TTG TCC TAC ACT AAC CCA GGG AGT TCA CTG GTG TCC CCA
Pro Asn Ala Leu Ser Tyr Thr Asn Pro Gly Ser Ser Leu Val Ser Pro
  145 150 155 160

TCT TTG GCA GCC AGC TCA ACG TTA ACA GAT TCA AGC ATG CTC TCT CCA
Ser Leu Ala Ala Ser Ser Thr Leu Thr Asp Ser Ser Met Leu Ser Pro
  165 170 175

CCT CAA ACC ACA TTA CAT AGA AAT GTG TCT CCA GGA GCT CCA CAG AGA
Pro Glu Thr Thr Leu His Arg Asn Val Ser Pro Gly Ala Pro Glu Arg
  180 185 190

CCA CCA AGT ACT GGC AAT GCA GTT GGG ATG TTG AGC ACT ACA GAC CTC
Pro Pro Ser Thr Gly Asn Ala Gly Gly Met Leu Ser Thr Thr Asp Leu
  195 200 205

ACA GTG CCA AAT GGA GCT GCA AGT CCA GTG GGG AAT GGA TTG GTA
Thr Val Pro Asn Gly Ala Gly Ser Ser Pro Val Gly Asn Gly Phe Val
  210 215 220

AAC TCA AGA GCT TCT CCA AAT TTT ATG GGT ACT CTC GTT GCA AAT AGC
Asn Ser Arg Ala Ser Pro Asn Leu Ile Gly Ala Thr Gly Ala Asn Ser
  225 230 235 240

TTA GGC AAA GTC ATG CCT ACA AAG TCT CCC CCA CCA GGT GGT CCT
Leu Gly Lys Val Met Pro Thr Lys Ser Pro Leu Pro Pro Gly Gly Gly
  245 250 255

AAT CTT GGA ATG AAC AGT ATT AAA CCA GAT CTT GCA GGT GTC ATC CCC
Asn Leu Gly Met Asn Ser Arg Lys Pro Asp Leu Arg Val Val Ile Pro
  260 265 270

CCT TCA AGC AAG GGC ATG ATG CCT CCA CTA TCG GAG CAA GAG TTG GAG
Pro Ser Ser Lys Gly Met Met Pro Pro Leu Ser Glu Glu Glu Glu Leu
  275 280 285

GAG TTG AAC ACC CAA AGG ATC AGT AGT TCT CCA GCC ACT CAA CCT CTT
Glu Leu Asn Thr Glu Arg Ile Ser Ser Ser Glu Ala Thr Glu Pro Leu
  290 295 300

GCT ACC CCA GTC GTG TCT GTG ACA ACC CCA AGC TTG CTC CGG CAA GGA
Ala Thr Pro Val Val Ser Val Thr Thr Pro Ser Leu Pro Pro Glu Gly
  305 310 315 320

CTT GTG TAC TCA GCA ATG CCG ACT GCC TAC AAC ACT GAT TAT TCA CTG
Leu Val Tyr Ser Ala Met Pro Thr Ala Tyr Asn Thr Asp Tyr Ser Leu
  325 330 335
- 74 -

 ACC AGC GCT GAC CTG TCA GCC CTT CAA GGC TTC AAC TCG CCA GGA ATG 1470
 Thr Ser Ala Asp Leu Ser Ala Leu Gln Gly Phe Asn Ser Pro Gly Met 340 345 350

 CTG TCG CTG GGA CAG GTG TCG GCC TGG CAG CAC CAC CTA GGA CAA 1518
 Leu Ser Leu Gly Gln Val Ser Ala Trp Gln Gln His His Leu Gly Gln 355 360 365

 GCA GCC CTC AGC TCT CTT GGT GCT GGA GGG CAG TTA TCT CAG GGT TCC 1566
 Ala Ala Leu Ser Ser Leu Val Ala Gly Gly Gln Leu Ser Gln Gly Ser 370 375 380

 AAT TTA TCC ATT AAT ACC AAC CAA AAC ATC AGC ATC AAG TCC GAA CCG 1614
 Asn Leu Ser Ile Asn Thr Asn Gln Asn Ile Ser Ile Lys Ser Glu Pro 385 390 395 400

 ATT TCA CCT CCT CGG GAT GTT ATG ACC CCA TCG GCC TCC CAG CAG CAG 1662
 Ile Ser Pro Pro Arg Asp Arg Met Thr Pro Ser Gly Phe Gln Gln Gln 405 410 415

 CAG CAG CAG CAG CAG CAG CAG CAG CCC CCA CCC CCA CCC CCA CCC CCA 1710
 Gln Gln Gln Gln Gln Gln Gln Gln Pro Pro Pro Pro Pro Pro Pro 420 425 430

 CCA CAA CCC CCC CAG CAG CAG CCC CAG CCA CAG GAA ATG GGG CCG TCC CCT 1758
 Pro Gln Pro Pro Gln Pro Pro Arg Gln Glu Met Gly Arg Ser Pro 435 440 445

 GTG GAC AGT CTG AGC AGC TCT AGT AGC TCC TAT GAT GGC AGT GAT CGG 1806
 Val Asp Ser Leu Ser Ser Ser Ser Ser Ser Tyr Asp Gly Ser Asp Arg 450 455 460

 GAG GAT CCA CGG GCC TTC CAT TCT CCA ATT GTG CTT GGC CGA CCC 1854
 Glu Asp Pro Arg Gly Asp Phe His Ser Pro Ile Val Leu Gly Arg Pro 465 470 475 480

 CCA AAC ACT GAG GAC AGA CAA AGC CCT TCT GTA AAG CGA ATG AGG ATG 1902
 Pro Asn Thr Glu Asp Arg Glu Ser Pro Ser Val Lys Arg Met Arg Met 485 490 495

 GAC GCG TGG GTG ACC TAA 1920
 Asp Ala Trp Val Thr 500

 GGCTTCCAAG CTGATGTTTG TACTTTTGTG TTACTGCAGT GACCTGCCCT ACATATCTAA 1980

 ATCGGGTAAT AAGGACATGA GTTAAATATA TTATATGTA CATACATATA TATATCCCTT 2040

 TACATATATA TGATATGGGG TGTGAGTGTG TGTGATATGTG TGGGTGTTGTG TTACATACAC 2100

 AGAATCAGGC ACTTACCTGC AAACCTCCTG TAGGTCTGCA GATGTGTGTG CCAAGCCAGA 2160
(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 507
(B) TYPE: amino acid
(C) STRANDEDNESS: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Gly Arg Lys Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Arg Asn

5 10 15

Arg Gln Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala

20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe

35 40 45

Asn Ser Ser Asn Lys Leu Phe Gln Tyr Ala Ser Thr Asp Met Asp Lys

50 55 60
Val Leu Leu Lys Tyr Thr Glu Tyr Asn Glu Pro His Glu Ser Arg Thr
65  70  75  80
Asn Ser Asp Ile Val Glu Ala Leu Asn Lys Glu His Arg Gly Cys
85  90  95
Asp Ser Pro Asp Pro Asp Thr Ser Tyr Val Leu Thr Pro His Thr Glu
100 105 110
Glu Lys Tyr Lys Lys Ile Asn Glu Glu Phe Asp Asn Met Met Arg Asn
115 120 125
His Lys Ile Ala Pro Gly Leu Pro Pro Gln Asn Phe Ser Met Ser Val
130 135 140
Thr Val Pro Val Thr Ser Pro Asn Ala Leu Ser Tyr Thr Asn Pro Gly
145 150 155 160
Ser Ser Leu Val Ser Pro Ser Leu Ala Ala Ser Ser Thr Leu Thr Asp
165 170 175
Ser Ser Met Leu Ser Pro Pro Gln Thr Thr Leu His Arg Asn Val Ser
180 185 190
Pro Gly Ala Pro Gln Arg Pro Pro Ser Thr Gly Asn Ala Gly Gly Met
195 200 205
Leu Ser Thr Thr Asp Leu Thr Val Pro Asn Gly Ala Gly Ser Ser Pro
210 215 220
Val Gly Asn Gly Phe Val Asn Ser Arg Ala Ser Pro Asn Leu Ile Gly
225 230 235 240
Ala Thr Gly Ala Asn Ser Leu Gly Lys Val Met Pro Thr Lys Ser Pro
245 250 255
Pro Pro Pro Gly Gly Gly Asn Leu Gly Met Asn Ser Arg Lys Pro Asp
260 265 270
Leu Arg Val Val Ile Pro Pro Ser Ser Lys Gly Met Met Pro Pro Ile
275 280 285
Ser Glu Glu Glu Leu Glu Leu Asn Thr Gln Arg Ile Ser Ser Ser
290 295 300
Gln Ala Thr Gln Pro Leu Ala Thr Pro Val Val Ser Val Thr Thr Pro
305 310 315 320
Ser Leu Pro Pro Gln Gly Leu Val Tyr Ser Ala Met Pro Thr Ala Tyr
325 330 335
Asn Thr Asp Tyr Ser Leu Thr Ser Ala Asp Leu Ser Ala Leu Gln Gly 340 345 350
Phe Asn Ser Pro Gly Met Leu Ser Leu Gly Gln Val Ser Ala Trp Gln 355 360 365
Gln His Leu Gly Gln Ala Ala Leu Ser Ser Leu Val Ala Gly Gly 370 375 380
Gln Leu Ser Gln Gly Ser Asn Leu Ser Ile Asn Thr Asn Gln Asn Ile 385 390 395 400
Ser Ile Lys Ser Glu Pro ile Ser Pro Pro Arg Asp Arg Met Thr Pro 405 410
Ser Gly Phe Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Pro 420 425 430
Pro Pro Pro Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Gln Pro Arg Gln 435 440 445
Glu Met Gly Arg Ser Pro Val Asp Ser Leu Ser Ser Ser Ser Ser Ser 450 455 460
Tyr Asp Gly Ser Asp Arg Glu Asp Pro Arg Gly Asp Phe His Ser Pro 465 470 475 480
Ile Val Leu Gly Arg Pro Pro Asn Thr Glu Asp Arg Glu Ser Pro Ser 485 490 495
Val Lys Arg Met Arg Met Asp Ala Trp Val Thr 500 505

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 473
(B) TYPE: amino acid
(C) STRANDEDNESS: linear
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

Met Gly Arg Lys Lys Ile Gln Ile Thr Arg Ile Met Asp Glu Arg Asn 5 10 15
Arg Gln Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala 20 25 30
Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe
35   40 45
Asn Ser Thr Asn Lys Leu Phe Gln Tyr Ala Ser Thr Asp Met Asp Lys
50  55  60
Val Leu Leu Lys Tyr Thr Glu Tyr Asn Glu Pro His Glu Ser Arg Thr
65  70  75  80
Asn Ser Asp Ile Val Glu Thr Leu Arg Lys Lys Gly Leu Asn Gly Cys
85  90  95
Asp Ser Pro Asp Pro Asp Ala Asp Asp Ser Val Gly His Ser Pro Glu
100 105 110
Ser Glu Asp Tyr Arg Lys Ile Asn Glu Asp Ile Asp Leu Met Ile
115 120 125
Ser Arg Gln Arg Leu Cys Ala Val Pro Pro Pro Asn Phe Glu Met Pro
130 135 140
Val Ser Ile Pro Val Ser Ser His Asn Ser Leu Val Tyr Ser Asn Pro
145 150 155 160
Val Ser Ser Leu Gly Asn Pro Asn Leu Leu Pro Leu Ala His Pro Ser
165 170 175
Leu Gln Arg Asn Ser Met Ser Pro Gly Val Thr His Arg Pro Pro Ser
180 185 190
Ala Gly Asn Thr Gly Gly Leu Met Gly Gly Asp Leu Thr Ser Gly Ala
195 200 205
Gly Thr Ser Ala Gly Asn Gly Tyr Gly Asn Pro Arg Asn Ser Pro Gly
210 215 220
Leu Leu Val Ser Pro Asn Leu Asn Lys Asn Met Gln Ala Lys Ser Pro
225 230 235 240
Pro Pro Met Asn Leu Gly Met Asn Asn Arg Lys Pro Asp Leu Arg Val
245 250 255
Leu Ile Pro Pro Gly Ser Lys Asn Thr Met Pro Ser Val Ser Glu Asp
260 265 270
Val Asp Leu Leu Leu Asn Gln Arg Ile Asn Asn Ser Gln Ser Ala Gln
275 280 285
Ser Leu Ala Thr Pro Val Val Ser Val Ala Thr Pro Thr Leu Pro Gly
290 295 300
Gln Gly Met Gly Gly Tyr Pro Ser Ala Ile Ser Thr Thr Tyr Gly Thr
305 310 315 320
- 79 -

Glu Tyr Ser Leu Ser Ser Ala Asp Leu Ser Ser Leu Ser Gly Phe Asn
325 330 335

Thr Ala Ser Ala Leu His Leu Gly Ser Val Thr Gly Trp Gln Gln Gln
340 345 350

His Leu His Asn Met Pro Pro Ser Ala Leu Ser Gln Leu Gly Ala Cys
355 360 365

Thr Ser Thr His Leu Ser Gln Ser Ser Asn Leu Ser Leu Pro Ser Thr
370 375 380

Gln Ser Leu Asn Ile Leu Lys Ser Glu Pro Val Ser Pro Pro Arg Asp
385 390 395 400

Arg Thr Thr Thr Pro Ser Arg Tyr Pro Gln His Thr Arg His Glu Ala
405 410 415

Gly Arg Ser Pro Val Asp Ser Leu Ser Ser Cys Ser Ser Ser Tyr Asp
420 425 430

Gly Ser Asp Arg Glu Asp His Arg Asn Glu Phe His Ser Pro Ile Gly
435 440 445

Leu Thr Arg Pro Ser Pro Asp Glu Arg Glu Ser Pro Ser Val Lys Arg
450 455 460

Met Arg Leu Ser Glu Gly Trp Ala Thr
465 470 473

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 521
(B) TYPE: amino acid
(C) STRANDEDNESS: linear
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gly Arg Lys Lys Ile Gln Ile Gln Arg Ile Thr Asp Glu Arg Asn
5 10 15

Arg Gln Val Thr Phe Thr Lys Arg Lys Phe Gly Leu Met Lys Lys Ala
20 25 30

Tyr Glu Leu Ser Val Leu Cys Asp Cys Glu Ile Ala Leu Ile Ile Phe
35 40 45
- 81 -

Asp Tyr Gln Leu Thr Ser Ala Glu Leu Ser Ser Leu Pro Ala Phe Ser
   340  345  350

Ser Pro Gly Gly Leu Ser Leu Gly Asn Val Thr Ala Trp Gln Gln Pro
   355  360  365

Gln Gln Pro Gln Gln Pro Gln Gln Pro Gln Gln Pro Gln Gln Pro
   370  375  380

Pro Gln Pro Gln Gln Pro Gln Gln Pro Gln Gln Pro Gln Gln Gln
   385  390  395  400

Pro Pro Gln Gln Ser His Leu Val Pro Val Ser Leu Ser Asn Leu
   405  410  415

Ile Pro Gly Ser Pro Leu Pro His Val Gly Ala Ala Leu Thr Val Thr
   420  425  430

Thr His Pro His Ile Ser Ile Lys Ser Glu Pro Val Ser Pro Ser Arg
   435  440  445

Glu Arg Ser Pro Ala Pro Pro Pro Ala Val Phe Pro Ala Ala Arg
   450  455  460

Pro Glu Pro Gly Asp Gly Leu Ser Ser Ser Pro Ala Gly Gly Ser Tyr Glu
   465  470  475  480

Thr Gly Asp Arg Asp Gly Arg Gly Asp Phe Gly Pro Thr Leu Gly
   485  490  495

Leu Leu Arg Pro Ala Pro Glu Pro Ala Glu Gly Ser Ala Val Lys Arg
   500  505  510

Met Arg Leu Asp Thr Thr Trp Thr Leu Lys
   515  520

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 11:

   (i) SEQUENCE CHARACTERISTICS:

   (A) LENGTH: 84
   (B) TYPE: amino acid
   (C) STRANDEDNESS: linear
   (D) TOPOLOGY: linear

   (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

   Arg Val Lys Ile Lys Met Glu Phe Ile Asp Asn Lys Ile Arg Arg Tyr
   5  10  15
Thr Thr Phe Ser Lys Arg Lys Thr Gly Ile Met Lys Lys Ala Tyr Glu
  20  25  30
Leu Ser Thr Leu Thr Gly Thr Gln Val Leu Leu Leu Val Ala Ser Glu
  35  40  45
Thr Gly His Val Tyr Thr Phe Ala Thr Arg Lys Leu Gln Pro Met Ile
  50  55  60
Thr Ser Glu Thr Gly Lys Ala Leu Ile Gln Thr Cys Leu Trp Ser Pro
  65  70  75  80
Asp Ser Pro Pro
  84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRAIGHTNESS: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

Arg Arg Lys Ile Glu Ile Lys Phe Ile Glu Asn Lys Thr Arg Arg His
  5  10  15
Val Thr Phe Ser Lys Arg Lys His Gly Ile Met Lys Lys Ala Glu Pro
  20  25  30
Leu Ser Val Leu Thr Gly Thr Gln Val Leu Leu Leu Val Val Ser Glu
  35  40  45
Thr Gly Leu Val Tyr Thr Phe Ser Thr Pro Lys Phe Glu Pro Ile Val
  50  55  60
Thr Gln Gln Glu Gly Arg Asn Leu Ile Gln Ala Cys Leu Asn Ala Pro
  65  70  75  80
Asp Asp Glu Glu
  84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 13:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

Arg Gly Arg Val Glu Met Lys Arg Ile Glu Asn Lys Ile Asn Arg Gln 5
Val Thr Phe Ser Lys Arg Arg Asn Gly Leu Leu Lys Ala Tyr Glu 10
25
30
Leu Ser Val Leu Cys Asp Ala Glu Val Ala Leu Ile Ile Phe Ser Ser
35
40
45
Arg Gly Lys Leu Tyr Glu Phe Gly Ser Val Gly Ile Glu Ser Thr Ile
50
55
60
Glu Arg Tyr Asn Arg Cys Tyr Asn Cys Ser Leu Ser Asn Asn Lys Pro
65
70
75
80
Glu Glu Thr Thr
84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Pro Gly Lys Ile Glu Ile Lys Lys Ile Glu Asn Ser Thr Asn Arg Gln 5
Val Thr Phe Cys Lys Arg Arg Asn Gly Ile Phe Lys Arg Lys Glu 10
20
25
30
Leu Thr Val Leu Cys Asp Ala Lys Ile Ser Leu Ile Met Ile Ser Ser
35
40
45
Thr Arg Lys Tyr His Glu Tyr Thr Ser Pro Asn Thr Thr Thr Lys Lys
50
55
60
Met Ile Asp Gln Tyr Gln Ser Ala Leu Gly Val Asp Ile Trp Ser Ile
65
70
75
80
His Tyr Glu Lys
84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRANDEDNESS: linear
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Arg Gly Lys Ile Gln Ile Lys Arg Ile Glu Asn Gln Thr Asn Arg Gln
5 10 15
Val Thr Tyr Ser Lys Arg Arg Asn Gly Leu Phe Lys Lys Ala His Glu
20 25 30
Leu Ser Val Leu Cys Asp Ala Lys Val Ser Ile Ile Met Ile Ser Ser
35 40 45

Thr Gln Lys Leu His Glu Tyr Ile Ser Pro Thr Thr Thr Ala Thr Lys Gln
50 55 60
Leu Phe Asp Gln Tyr Gln Lys Ala Val Gly Val Asp Leu Trp Ser Ser
65 70 75 80
His Tyr Glu Lys
84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRANDEDNESS: linear
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Arg Gly Lys Ile Gln Ile Lys Arg Ile Glu Asn Gln Thr Asn Arg Gln
5 10 15
Val Thr Tyr Ser Lys Arg Arg Asn Gly Leu Phe Lys Lys Ala His Glu
20 25 30
Leu Thr Val Leu Cys Asp Ala Arg Val Ser Ile Ile Met Phe Ser Ser
35 40 45
Ser Asn Lys Leu His Glu Tyr Ile Ser Pro Asn Thr Thr Thr Lys Glu
50 55 60
Ile Val Asp Leu Tyr Gln Thr Ile Ser Asp Val Asp Val Trp Ala Thr
65 70 75 80
Gln Tyr Glu Arg
84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84
(B) TYPE: amino acid
(C) STRANDEDNESS: linear

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

Arg Gly Lys Ile Glu Ile Lys Arg Ile Glu Asn Thr Thr Asn Arg Gln
5 10 15
Val Thr Phe Cys Lys Arg Arg Asn Gly Ile Leu Lys Ala Tyr Glu
20 25 30
Leu Ser Val Leu Cys Asp Ala Glu Val Leu Ala Ile Val Phe Ser Ser
35 40 45
Arg Gly Arg Leu Tyr Glu Tyr Ser Asn Asn Ser Val Lys Gly Thr Ile
50 55 60
Glu Arg Thr Lys Lys Ala Ile Ser Asp Asn Ser Asn Thr Gly Ser Val
65 70 75 80
 Ala Glu Ile Asn
84

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

CTAAAATAAA

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 19

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xii) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

CTAATATATA TTAG

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 9
(B) TYPE: amino acid
(C) STRANDEDNESS: 
(D) TOPOLOGY: linear

(xiii) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Ser Glu Glu Glu Glu Glu Leu Glu Leu

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xiv) SEQUENCE DESCRIPTION: SEQ ID NO: 21

YTWWAAATAR

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 22:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27
(B) TYPE: nucleic acid
- 87 -

(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GATCCTCGCT CTAAAATAA CCCTGTM

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 23:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

AGCTTCGGAC CCTGCTCATT TCTATATATA G

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 24

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

AGCTTGGGGA CCAAATAAGG CAAGGTG

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(x) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GATCCTTCCC AATGATTTGC ATGCTCTCAC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 26

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

GATCTCCCTG GGGTTAAAAA TAACCCCCATG AC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GATCGATCGA TGCGTGGTTA TAATTAACCC AGACAT

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GATCTCCGAC GGGTTAAAAA TAGCAAAACT CT

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

GATCCCTTTTC AGATTTAAAA TAACTAAGGT AA
(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

GATCGCCCCAA GGACTAAAAA AAGGCCCTGG A 31

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

Thr Pro His Thr Glu Glu Lys Tyr Lys Lys Ile Asn Glu Glu Phe Cys Thr
1 5 10 15

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Cys Asp Tyr Phe Glu His Ser Pro Leu Ser Glu Asp Arg
5 10

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

TTAAAAATAAA

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CGCTCTAAAAA ATAACCCCT

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CGCTCTAAGG CTAACCCCT

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 36:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

CGCTCTATAAA ATAACCCCT

18
(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GCCTCTAAAC ATAACCT

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

ATTTCTATAT ATACTTTC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

GGGGACCAAA TAAGGCAA

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 40:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

CCAATGATTT GCATGCTC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 41:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GGGTTAAAAA ATAACCC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 42:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

CTGGTTATAA TTAACCC

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 43:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

CGGGTTAAAA ATAGCAAA

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 44:
(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

CAGATTAAAA ATAACTAA 18

(2) INFORMATION FOR SEQUENCE IDENTIFICATION NUMBER: 45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

AGGACTAAAA AAAGGCC 18
Claims

1. An essentially pure nucleic acid encoding a member of the Myocyte-specific Enhancer Factor 2 (MEF2) protein family which has myocyte transcription enhancing activity.

2. The nucleic acid of claim 1, wherein said nucleic acid comprises at least one of the following elements: a) a nucleotide sequence encoding at least eleven consecutive glutamine residues, or b) a nucleotide sequence encoding the amino acid sequence SEEEEEL (SEQ ID NO: 1).

3. The nucleic acid of claim 1, wherein said nucleic acid encodes at least a 54 amino acid portion of the amino acid sequence of Fig. 1A.

4. The nucleic acid of claim 1, wherein said MEF2 family member is a mutant of a wild-type protein, said wild-type protein comprising an inactivation domain, said nucleic acid being deleted for sequences encoding said inactivation domain.

5. The nucleic acid of claim 1, wherein said MEF2 is an isoform of the MEF2 sequence shown in Fig. 1.

6. The nucleic acid of claim 1, wherein said MEF2 is selected from the group consisting of aMEF2, xMEF2, dMEF2, and cMEF2.

7. The nucleic acid of any of claims 1–6 for use in therapy.
8. The nucleic acid of any of claims 1-6 for use in the treatment or prevention of muscular dystrophy or muscle atrophy in a mammal.

9. The nucleic acid of any of claims 1-6 for use in enhancing muscle mass in a mammal.

10. The nucleic acid of any of claims 1-9 in combination with a pharmaceutically acceptable carrier.

11. A nucleic acid vector comprising the nucleic acid of any of claims 1-6.

12. The vector of claim 11, wherein said vector comprises a transcription regulatory sequence positioned and oriented to regulate expression of said nucleic acid encoding said MEF2 family member.

13. A cell comprising the vector of claim 12.

14. A substantially pure MEF2 polypeptide encoded by the nucleic acid of any of claims 1-6.

15. A substantially pure MEF2 polypeptide encoded by the nucleic acid of any of claims 1-6 for use in therapy.

16. A substantially pure MEF2 polypeptide encoded by the nucleic acid of any of claims 1-6 for use in the treatment or prevention of muscular dystrophy or muscle atrophy in a mammal.

17. A substantially pure MEF2 polypeptide encoded by the nucleic acid of any of claims 1-6 for use in enhancing muscle mass in a mammal.
18. The polypeptide of any of claims 14-17 in combination with a pharmaceutically acceptable carrier.

19. A transgenic non-human mammal comprising a first transgene, said first transgene comprising the nucleic acid of any of claims 1-6.

20. The transgenic mammal of claim 19, further comprising a second transgene, said second transgene comprising a promoter and regulatory DNA positioned to effect expression of a structural gene, said promoter and regulatory DNA being characterized in that said expression is enhanced by said MEF2 protein family member.

21. The transgenic mammal of claim 19, further comprising a second transgene, said second transgene enhancing the activity of said MEF2 protein family member.

22. The transgenic mammal of claim 8, wherein said enhancing is selected from the group consisting of a) increasing the expression of said MEF2 protein family member, and b) increasing the activity of an at least partially inactive form of said MEF2 protein family member.

23. The transgenic mammal of claim 21, wherein said second transgene encodes a MyoD polypeptide, a myogenin polypeptide, a retinoblastoma polypeptide, or a homeobox protein.

24. The transgenic mammal of claim 19, wherein said first transgene is expressed by a tissue-specific promoter.
25. The transgenic mammal of any of claims 19-24, wherein said first transgene is introduced into said mammal, or an ancestor of said mammal, at an embryonic stage.

26. The transgenic mammal of any of claims 19-24, wherein said transgene is introduced into a somatic cell or into a somatic tissue of said mammal.

27. A method of identifying a molecule that enhances the activity of a member of the Myocyte-specific Enhancer factor 2 (MEF2) family, comprising
    providing a candidate molecule;
    combining the candidate molecule with a polypeptide comprising an activity reducing domain of the carboxy terminal one-third of an active MEF2 family member according to claim 22, said polypeptide being characterized in that it lacks MEF2 transcription enhancing activity, and said domain being characterized in that deletion of said domain from said MEF2 family member enhances MEF2 activity of said family member;
    determining whether said candidate molecule binds to said polypeptide.

28. A method of identifying a molecule that enhances the activity of a member of the Myocyte-specific Enhancer factor 2 (MEF2) family, comprising
    providing a candidate molecule;
    providing a MEF2 family member according to claim 22 in a solution,
    providing a MEF2 consensus nucleic acid binding sequence
    determining whether said candidate molecule enhances binding of said MEF2 family member to said MEF2 consensus binding sequence.
29. A method of identifying a molecule that enhances the activity of a member of the Myocyte-specific Enhancer factor 2 (MEF2) family, comprising
  providing a candidate molecule;
  providing nucleic acid according to claim 1,
transformed into a cell said cell comprising a structural gene which comprises a regulatory region that includes a MEF2 consensus binding sequence and a promoter responsive to said consensus binding sequence;
determining whether introduction of said candidate molecule into said cell enhances expression of said structural gene.
MADS Domains

MEF2  3  RKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
MEF2  3  RKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY

AG   52  RKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
DFA  3  RKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
SRF  143 RVKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
MCM  18  RVKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
ARG2  80  RVKKERTRMDEERNRQVTFTKFKGLMKKAYELSVLCDEZAL12FTNNESS1LPQY
Figure 1C

MEF2 - Related Gene Products

y/aMEF2  cMEF2  dMEF2  xMEF2
Probe = xMEF2
Figure 5B
Figure 5C

- MLC2 HF-1
- MCK MEF2
- MCK A/T
- cTNT A/T
- α MHC A/T-1
- A/Temb
- α' MHC A/T-2
Figure 6A

N. Extract  pC2 nb  C2 nt  Card  Sol 9 nb  Sol 9 nt  Smooth M.  3T3  3T3 + MyoD  10T 1/2  HeLa

F → 1 2 3 4 5 6 7 8 9 10 11
Figure 9

CAT Activity per Injection Site (CPM/100)

\[ y = 0.2x + 10.8, r^2 = 0.5 \]
Figure 10

CAT-activity (cpm/1000)

Days Post-Injection:
- 3 days
- 7 days
- 14 days
- 21 days

* Indicates a significant difference.
Figure 13

- **Muscle Specific Promoter: r667-CAT**
  
  \[ y = 0.8x + 113, \quad r^2 = 0.8, \quad P < 0.01 \]

- **Promiscuous Promoter: MSV-CAT**
  
  \[ y = 3.5x + 665, \quad r^2 = 0.9, \quad P < 0.005 \]
d-MEF2 Alternatively Spliced Isoforms
Figure 19 Cont.

451

501
**INTERNATIONAL SEARCH REPORT**

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(S): Please See Extra Sheet.
US CL: Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S.: 435/6, 7.1, 69.1, 172.3, 240.2, 320.1; 530/358; 536/23.5, 24.1, 24.31; 800/2, DIG 2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

GENBANK, EMBL, BIOSIS, CAS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Molecular and Cellular Biology, Vol. 9, No. 8, issued August 1989, Johnson et al., &quot;Muscle Creatine Kinase Sequence Elements Regulating Skeletal and Cardiac Muscle Expression in Transgenic Mice&quot;, pages 3393-3399, see entire document.</td>
<td>19-26</td>
</tr>
</tbody>
</table>

**X** Further documents are listed in the continuation of Box C.  **See patent family annex.**

<table>
<thead>
<tr>
<th>*</th>
<th>Special categories of cited documents:</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;A&quot;</td>
<td>document defining the general state of the art which is not considered to be part of particular relevance</td>
</tr>
<tr>
<td>&quot;E&quot;</td>
<td>earlier document published on or after the international filing date</td>
</tr>
<tr>
<td>&quot;L&quot;</td>
<td>document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td>
</tr>
<tr>
<td>&quot;O&quot;</td>
<td>document referring to an oral disclosure, use, exhibition or other means</td>
</tr>
<tr>
<td>&quot;P&quot;</td>
<td>document published prior to the international filing date but later than the priority date claimed</td>
</tr>
</tbody>
</table>

**Date of the actual completion of the international search**

23 NOVEMBER 1993

**Name and mailing address of the ISA/US Commissioner of Patents and Trademarks**

Box PCT

Washington, D.C. 20231

**Authorized officer**

Jacqueline M. Stone

Authorized officer

**Telephone No.:** (703) 308-0196

**Form PCT/ISA/210 (second sheet)(July 1992)**
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Molecular and Cellular Biology, Volume 11, No. 10, issued October 1991, Cserjesi et al., &quot;Myogenin Induces the Myocyte-Specific Enhancer Binding Factor MEF-2 Independently of Other Muscle-Specific Gene Products&quot;, pages 4854-4862, see entire document.</td>
<td>14-29</td>
</tr>
</tbody>
</table>
A. CLASSIFICATION OF SUBJECT MATTER:
IPC (S):
C12N 5/10, 15/12, 15/90, 15/63; G01N 33/50; A01K 67/00; C07H 21/04; C07K13/00

A. CLASSIFICATION OF SUBJECT MATTER:
US CL:
435/6, 7.1, 69.1, 172.3, 240.2, 320.1; 530/358; 536/23.5, 24.1, 24.31; 800/2, D1G 2