## Cloning and characterization of four murine homeobox genes

(mouse embryo/transcription factors/Uncx-4.1/OG-2/OG-9/OG-12)

ALESSANDRA CECILIA ROVESCALLI, SADAMITSU ASOH\*, AND MARSHALL NIRENBERG<sup>†</sup>

Laboratory of Biochemical Genetics, National Heart, Lung, and Blood Institute, National Institutes of Health, Building 36, Room 1C-06, 36 Convent Drive, MSC 4036, Bethesda, MD 20892-4036

Contributed by Marshall Nirenberg, July 17, 1996

ABSTRACT Four novel murine homeobox genes, Uncx-4.1, OG-2, OG-9, and OG-12, were cloned and partially sequenced. The amino acid sequence of the mouse Uncx-4.1 homeodomain is closely related to the sequence of the unc-4 homeodomain of Caenorhabditis elegans. However, the OG-2, OG-9, and OG-12 homeodomains are relatively diverged and are not closely related to any previously described homeodomain. Northern blot analyses revealed multiple bands of Uncx-4.1, OG-2, OG-9, and OG-12 poly(A)<sup>+</sup> RNA in RNA from mouse embryos and adults that change during development and showed that each gene is expressed in a tissue-specific manner. OG-12 cDNAs were cloned that correspond to two alternatively spliced species of OG-12 mRNA. Three major bands of Uncx-4.1 poly(A)<sup>+</sup> RNA were found only in RNA from adult mouse brain, but an additional band was observed in RNA from all of the other tissues tested. Major bands of OG-9 and OG-2 poly(A)<sup>+</sup> RNA were found only in RNA from striated muscle; however, trace bands were detected in RNA from other tissues.

The homeodomain is a 60-amino acid residue portion of a protein, encoded by a homeobox gene, that binds to DNA and regulates gene expression (for recent reviews, see refs. 1 and 2). Approximately 140 murine homeobox genes have been cloned thus far and 30 additional kinds of homeobox genes have been cloned from other organisms (3). The mouse genome contains four chromosomal clusters of Hox homeobox genes, thought to have evolved by successive gene duplications of an ancestral Ant-Ubx cluster of homeobox genes (4). Both the amino acid sequences of homeodomains and the order of Hox genes in each chromosomal cluster have been conserved during evolution. The order of the genes within a cluster of genes corresponds to the order of most anterior borders of expression of the genes along the rostrocaudal axis of the embryo (2). In addition, many divergent homeobox genes have been cloned that reside at other chromosomal locations (for reviews, see refs. 2 and 3). The most anterior border of expression of *Hox* homeobox genes is within the hindbrain; however, some other homeobox genes are expressed in discrete areas, some times more rostrally than Hox genes (3).

Many murine homeobox genes have been cloned by screening mouse embryo cDNA libraries at reduced stringency with probes that correspond to known homeobox genes from other organisms. However, murine homeobox genes with nucleotide sequences that diverge from those of the probes may be difficult to detect with this strategy. Therefore, we screened a mouse genomic DNA library with a mixture of oligodeoxynucleotide probes that correspond to the complement of all nucleotide sequences encoding the amino acid sequence KVWFQNR that is found in the third  $\alpha$ -helix of some homeodomains (amino acid residues 46–52). In this report, we describe four novel murine homeobox genes, *Uncx-4.1*, *OG-9*,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

OG-12, and OG-2 that are expressed in mouse embryos and in some adult mouse tissues.

## **METHODS AND MATERIALS**

**Oligodeoxynucleotides.** Oligodeoxynucleotides were synthesized with an Applied Biosystems DNA synthesizer model 380B and purified either by precipitation with 1-butanol (5) or by OPC column chromatography (Perkin–Elmer/Applied Biosystems Division) as described by the manufacturer. Oligodeoxynucleotides used as probes for DNA hybridization were labeled by phosphorylation with  $[\gamma^{-32}P]ATP$  (ICN; >7000 Ci/mmol; 1 Ci = 37 GBq) catalyzed by T4 polynucleotide kinase.

Library Screening and Gene Cloning. A genomic DNA library from BALB/cAn mouse liver DNA partially digested with EcoRI and cloned in the EcoRI site of EMBL-4 (average size of DNA insert, 12–18 kb) was used. Approximately 10<sup>6</sup> plaque-forming units were screened in duplicate. Hybridization and the final wash were performed at 55.5°C in the presence of 3 M tetramethylammonium chloride (6).

**DNA Sequencing.** Cloned genomic DNA fragments cleaved by restriction enzymes were subcloned in pBluescript II KS+ (Stratagene) and were sequenced either manually with the Sequenase II kit (Amersham/United States Biochemical) or with a Perkin–Elmer/Applied Biosystems DNA sequencer (model 373A) using *Taq* DNA polymerase and fluorescent dideoxynucleotides (Perkin–Elmer/Applied Biosystems). Mixtures of oligodeoxynucleotides corresponding to conserved amino acid sequences in homeodomains, M13 forward and reverse primers, and sequence-specific primers were used to sequence both strands of DNA. The Wisconsin Sequence Analysis Package (Genetics Computer Group, Inc., Madison, WI), Version 8, was used for sequence analysis and DNA segment assembly.

Northern Blot Hybridization. Multiple tissue Northern blots containing  $poly(A)^+$  RNA from mouse embryos at different stages of development or  $poly(A)^+$  RNA from different adult mouse tissues were purchased from CLONTECH and hybridized according to the manufacturer's instructions. [<sup>32</sup>P]labeled DNA and RNA probes were synthesized, respectively, by random-priming of denatured DNA templates catalyzed by the Klenow fragment of DNA polymerase I (Pharmacia) or by reverse transcription of linearized recombinant DNA catalyzed by T3, T7, or SP6 RNA polymerase (Ambion, Austin, TX).

Abbreviation: SH3, src homology domain 3.

Data deposition: The sequences reported in this paper have been deposited in the GenBank data base [accession nos. U65069, U65070, U65067, U65068, U65071, and U65072 (for Uncx-4.1, OG-2, OG-9, OG-12, OG-12a, and OG-12b, respectively)].

<sup>\*</sup>Present address: Division of Biochemistry, Institute of Gerontology, Nippon Medical School, 1–396 Kosugi-cho, Nakahara-ku, Kawasaki 211, Japan.

<sup>&</sup>lt;sup>†</sup>To whom reprint requests should be addressed.

## **RESULTS AND DISCUSSION**

Genomic DNA Clones. Sixty-nine recombinant DNA clones were obtained by screening a mouse genomic DNA library with a mixture of 128 species of 20-nt antisense oligodeoxynucleotides, corresponding to the complement of all synonym codons for the amino acid sequence KVWFQNR, which is a conserved sequence within the third  $\alpha$ -helix of some homeodomains (amino acid residues 46-52). Sixteen kinds of mouse genomic DNA clones were identified by Southern blot analysis of cloned recombinant DNA incubated with EcoRI, EcoRI/BamHI, EcoRI/SstI, EcoRI/HindIII, and EcoRI/ XhoI). At least one representative clone from each class was subcloned and partially sequenced. Nucleotide sequence analysis showed that 51 of the 69 recombinant DNA clones obtained were homeobox genes. Five novel murine homeobox genes were found: Uncx-4.1, OG-9, OG-12, and OG-2 are described in this report; OG-22, which is the mouse homologue of the rat cart-1 homeobox gene (7), will be described elsewhere (S.A., A.C.P. and M.N., unpublished results). In addition, 4 previously described murine homeobox genes were cloned: Hoxa-2 (8), Emx-2 (9), Otp (10), and K-2/Mhox (11, 12)

Nucleotide sequence analysis showed that an intron is present in the genomic DNA between the codons corresponding to the 46th and 47th amino acid residues of homeodomains of six genes (OG-9, OG-12, OG-2, Uncx-4.1, Otp, and K2/ MHox). The results show that 19 of the 20 nucleotide residues of some oligonucleotide probes hybridized correctly to complementary nucleotide sequences in homeoboxes of cloned DNAs. One mismatched base was found at the 3' ends of oligodeoxynucleotide probes hybridized to genomic DNA. Consequently, the 3'-terminal thymidine residues of probes

1	CGG	AG.	ACC	SCA(	ccc	GC1	ACC	AAC	TTI	TAC	CGG	CT	GGC	AGC	TG	GA	GGZ	\GC'	rgg.	AGA	AGG	CGT	TC
1	R	R	R	т	R	Т		a 1	r I	т	G	W	Q	L		E	E	L	E	ĸ	A	F	
61	AAT	GA	GAG	CC2	ACT	ACO	CG	GAC	GTC	TT	TAT	GC	GCG	AGG	CG	CT	GGC	:GC'	rgc	GCC	TGG.	ACC	TG
21	N	E	S	H	Y	1	•	D 7	v	F	M	R	D	A	•	L	A	L	R	L	D	L	
121	GTC	GA	GTC	cco	GAG	TT	CAG	GTA	AAG	SAT	TGG	TG'	rcg	CCC	AA	GG	AGI	CCC	GAG.	ATC	CCG	CAC	TA
41	v	E	S	R	v	(	2			-													
181	GGC	GT	GCA	GA	GCG	CGC	TG	AGG	CAG	GGG	CCG	CC	TTT	GTI	CT	AA	GTO	GGC	GGG	GAG	GTG	GTT	TC
241	AGG	CT	CCG	GCG	CTG	CTO	GCC	CAT	GGI	GG	GGA	TT	ATG	AGG	GT	TG	AAC	CCC	CTT	CTC	CCA	TCT	GG
301	AGG	TG	AGT	AA	CTT	GAC	GAG	TGC	CGA	GG	TGA	TT	GGG	GAG	AG	AT	CC.			11-		-CC	AG
361	AAG	CT	GCA	ACO	CCG	AGT	TAT	CAG	CAT	CC	AAG	GG	CTG	GTG	GA	GA	GCI	TAG	CCC	TGG	ATC	CGG	CT
421	GCT	CCO	CGC	TG	GAG	CCC	AG	GTT	GTI	CA	GAC	CT	GTC	TGG	GG	CG	CGC	TT	GGG.	AGG	CCC	GCG	GC
481	CCC	CG	CGG	AG	GGC	TAC	SCC	AGC	CTG	AT	TGT	GC	CTT	CCC	CC	AT	CCC	CCC	CTT	TGT	CCG	GCT	GT
541	TCT	AG	GTC	TG	TT	CCZ		ATC	GCC	'GG	GCC	AA	ATG	GAG	AA	AG	AAC	GAG	SAA	CAC	CAA	AAA	GG
47	-		v	W	F	Q	N	R	F	ι.	A	ĸ	W	R	K		K	E	N	Т,	3K	ĸ	G
601	GCC	CG	GGC	CGG	GCC	GGC	CCZ	ACA	ACT	CG	CAC	CCO	GAC	CAC	GT	GC	AGC	GGG	GGA	GCC	CAT	GGA	CC
67	P	• (	G	R	P	A	н	N	S	5	H	P	Т	т	C		S	G	E	P	М	D	P
661	CAG	AG	GAG	AT	CGC	TCO	CA	AGG	AAC	TG	GAG	AA	GAT	GGA	GA	AG	AAC	AA	ACG	CAA	ACA	CGA	GA
								b	asi	c	and	a	ciđ	ic	re	si	due	35					
87	E		E	I	A	R	K	E	L	,	E	K	М	E	K		K	K	R	K	H	E	K
721	AGA	AA	CTG	CTC	CAA	GAC	TCI	AGA	GCC	GC	CAC	CTO	GCA	CTC	GC	CC	GGI	GGG	CCT	GTC	CCT	GCA	CA
107	K	1	L	L	к	s	3 0	s	23F	2	н	L	H	S	P		G	G	L	S	L	н	S.
781	GCG	CG	CCC	AG	CTC	CGZ	CAG	GCG.	ACA	GC	GGT	GG	CGG.	AGG	CC	TG	TCI	CCZ	AGA	GCC.	ACC	CGA	GC
																			S	H3	bin	din	g
127	A	. 1	P	S 5	s	D	s	E D	S	5	G	G	G	G	L		S	P	Е	P	P	Е	P
841	CGC	CA	CCG	CCC	GAC	CGC	CG	ČGG.	ACA	AG	GGT	CC	rgg.	AGC	GC	AC	GGC	TC	rgg	CAT	CGC	GGG	TT
147	F		P	P	т.	A	A	D	F	٢.	G	P	G	A	н		G	S	G	I	A	G	s
901	CCG	CT	cco	GT	ACC	TCO	TG	GCG	AGC	CA	ccc	GCO	GCC	TGG	CA	CC	TGO	GA	rcc	CGC	CTT	CTA	CC
						SH	b b	ind	ing	T													
167	A	. 1	P	v	Ρ	P	G	E	F	>	P	A	Ρ	G	т	6	С	D	P	A	F	Y	P
961	CGA	GC	CAA	AG	AAG	CGC	GCG	CCG	GCI	CG	CAG	CC	ACG	G 9	95								
187	S	23	2	R	s,	G	A	G	5	5	Q	P	R	1	197	7							

FIG. 1. Partial nucleotide sequence of *Uncx-4.1* genomic DNA and predicted amino acid sequence. The nucleotide sequences and predicted amino acid sequences of portions of the homeodomain are enclosed in the dark-shaded boxes. Other regions of interest are indicated by light shading. The broken line indicates a gap in the nucleotide sequence. Consensus sequences for intron ends are underlined and sites for splicing are indicated by arrowheads. Amino acid residues that are potential sites for phosphorylation are shown in white on black backgrounds. The numbers correspond to the following enzymes that may catalyze phosphorylation (14): 1, tyrosine protein kinase; 2, protein kinase C; 3, cGMP-dependent protein kinase; 4,  $Ca^{2+}/calmodulin-dependent protein kinase; 5, casein kinase I or II.$ 

hybridized to 3'-splice site sequences of introns, CAG or TAG (13), instead of the first base of the Lys codon, AAG.

Partial Nucleotide Sequence of Uncx-4.1. In Fig. 1 are shown the partial nucleotide sequence and predicted amino acid sequence of the novel mouse homeobox gene, Uncx-4.1. Nucleotide residues 1-138 are part of an exon that encodes amino acid residues 1-46 of the homeodomain. An intron, which was sequenced only partially, is present within the homeobox between codons for the 46th and 47th homeodomain amino acid residues. The RNA splice sites shown in Fig. 1 were confirmed by nucleotide sequence analysis of Uncx-4.1 cDNA, which will be reported elsewhere (A.C.R., R. Ray, and M.N., unpublished results). Exon 2 encodes 151 amino acid residues, including homeodomain residues 47-60. Following the homeodomain, the deduced amino acid sequence contains a region rich in basic and acidic amino acid residues and tandem consensus amino acid sequences for binding to src homology 3 (SH3) domains (15-17). The amino acid sequence of the mouse Uncx-4.1 homeodomain is 88% identical to the unc-4 homeodomain of Caenorhabditis elegans (18).

**Partial Nucleotide Sequence of** OG-9. The nucleotide sequence of a fragment of the putative novel OG-9 homeobox gene, 1271 nt long, and the deduced amino acid sequence of the protein are shown in Fig. 2. Sites for RNA splicing have been confirmed by nucleotide sequence analysis of partial OG-9 cDNAs (M. Cinquanta, A.C.R., and M.N., unpublished results). The first 70 nt corresponds to part of an intron, which is followed by an exon encoding 52 amino acid residues, including amino acid residues 1–46 of the OG-9 homeodomain. The homeobox is interrupted by a small intron, which is followed by an exon that encodes 127 amino acid residues, starting with Val-47 of the homeodomain. Ser and Thr residues



FIG. 2. Partial nucleotide sequence of OG-9 genomic DNA and predicted amino acid sequence. Homeodomain regions are enclosed in dark-shaded boxes. Consensus sequences for intron ends are underlined and sites for RNA splicing are indicated by arrowheads. Amino acid residues that are potential sites for phosphorylation are shown in white on black backgrounds. The numbers correspond to phosphorylations that may be catalyzed by the following enzymes (14): 1, protein kinase C; 3, cGMP-dependent protein kinase; 4, Ca<sup>2+</sup>/calmodulindependent protein kinase; 5, casein kinase I or II. A putative polyadenylylation signal is underlined and in-frame termination codons are shown in boldface type.

## Biochemistry: Rovescalli et al.

121	GGTGGTGCGTGCTCCGACGCACGTGTGGTCCTTCCGTGGTG <u>CCCTTTCAGTGTCCCCTGAACTGAA</u>
1	D R K D D A K G M E D E G Q T K I K
241	AGCAGAGGCGAAGTCGGACCAATTTTACCCTGGAACAACTCAACGAGGCTGGAGAGGCTTTTCGATGAGACCCACTATCCAGACGCTTTCATGCGCGAGGAATTGAGCCAGCGACTGGGGC
19	QRR S <sub>123</sub> R T <sub>234</sub> N F 3LEQLNELERLFDETH PDAFMREEL S <sub>23</sub> Q RLG I 1
361	TCTCTGAGGCCCGAGTACACGCCCCATTGCGAGGGGGGGG
59	S, E A R V Q
481	AAAAGGGGGGGACATCTTTTTTTTTTTAAAGAT GAGAAAGTCTAAACTTTTTGTAGTGAGGGCTTTTATGACACTAATTACCACCCCCATGTCCCCTTCCCCCACCAAAAT
601	${\tt AAGAGTCATTATTGCTTTACTGGTTTGCTTTTCTTCAAGATGAAGCTAATTAAT$
721	ACATTGGTGTCAGGTACCAGGGAGTGTTTTGGACCCCTGGGACTCTGTAATGTGGAAAGATGAAACCGGAACTAGTAGTAGTAGTAGTAGAACCAGTTTACCAAGGAACCAATTTACCTTAAAACGAGTCTAGGAA
841	ATCTTTCTATTTGTTTTTTTTTTTGTATTTTGATTTGAT
961	CTGCCCCCCCCATTGTACTACTATAATTTAAGATGTTTTGAAATATCACAGGTTTATGGT <u>TTTTTTTTCCCCAACTTTTAG</u> GTTTGGTTCAAAATCGAAGAGCTAAGTGTAGAAAACAG
65	V W F Q N R R A K C R K Q
1081	GAAAATCAACTTCACAAAGGTATATCTCATAATTCATTTATCATTAAGTAGAAGTGGGTTTGGGGTATAGCTATAACTAAGACAAAGAACATCCTAAGGATGACAATGGGACATGGACATGATTT
78	ENQLHKG
1201	AAGAAAAATTTCTGGGCAGTTAATGAAAGTTCTACTACTTAGAAGCAAACTTCAACTTATTTAACTATCACAATAACTTAGAAAAATGACAGTCCAGAAATGACTTGCTACCACTTGGC
	이 같은 것 같은
1321	accttggaaattatttatgtatgaaatgtttacctcacatctcatg <u>tttcatctcccctttttttgtttgttttgt</u>
85	VLIGAA <mark>Is</mark>
1441 <b>92</b>	GCCAGTTTGAAGCTTGTAGAGTTGCAACCTATGTCAACGTAGGTGCTTTAAGGATGCCATTTCAGC <u>AGGCAAG</u> TACAACCTAGCTACAAGCTTTTTTTTTTTTTT
1561	${\tt CTGGATGTTTGATGCTGAACAGACCCTGTTTCTTTGCTTCAATTGAAGAACCAATTTCAAGACCTATATAGAGCACCCCCTTTAAAAAAACCTATACAGTGTCAAGGCCTT$
1681	GAATATTTAGTCCTTTCTGTCTTCATAGGAGCCTGAGGGGAGCAACCTGTCCTTTAAAAACTGCATTTACGATGGCACCTGGGGAGAGGGAGAAACAGCGTTTGAGTAGGAACAACCA
1801	ATACAGAGGCTGGTCCAAGGGTGTTATAAAAGCAATCCTCTAAATAGTCATTCTAGATAGTTTTTTTT
1921	ATTCTCAATTCAGACTTATTTTTAAAAAGCTATCCTAAAAAATAAACAACAACAACAAAAAAGTTCATCTCTAATTAAAGAGGGAGG
2041	AGAARTCCA X X RECOMMENDAMENTAL TETRAGATAACGETTTAAAGAARGCCTGCAGCARGATGATGATGATGATGATGATGATGATGATGATGATGATG
2281	
2401	AAGAGAGCCAGAACTTACTAGGAAACTGGCCCTACCTAAGAAGGTAGAGATCTTCAGGCCAGCAAAATCAGCCTCACCTAGAAGGTGGTTAGAAACCTAGTCTGAGTCCCAACACACAC
2521	ACCTACTGACTCAGAATCTGTATTTAAACATGATCCCTACAGCAGGAAAAGTTAAGACCTGCAGAATTCAATAGCTCCCCTAAACGACAAGAAAAAGTTATATTTTTGTTTTAGTAGTG
2641	TGTCCAGGGGTATGTGACAAAGACAGGAGAGAGGATGGCAGTCCTTAACCCCAACCTCTCCAAGATGAGGGGGGGG
2761	${\tt CCCGAAGGGACCACCTCATTCTTGCAGTCAGGGACTAGAAACCCCAAGGAAACCATGCTCGTCGTCGTCGTGGAGGTAGAGATGGGGCACGACTAGAGTACCAATGTGGGGTCAGGGACACTAGAAACCCCAAGGAAACCATGCTCGTCGTCGTCGTCGTGGTGGTGGGGGAGACACTAGAGTACCAATGTGGGGGTGTTCCCAGGTAGGGACACTAGAAACCACGCACG$
2881	AAAGCCTATAAACGTCCTTCTCCAAACAGATACTTTGCCAAATGTTGCTTCCCTGTCGCGCACCCCAATTCTCCCCATAGCCAAGCTGGACGCCTTTAATGTTCTGTTTCTATTCTGTTGT
	이 이 가지 🔶 같이 같이 많은 것 같은 것 같은 것 같은 것이 있는 것 🚽 이 가지 않는 것 같은 것이 가지 않는 것 같이 가지 않는 것 같이 가지 않는 것 같이 같이 많이 있는 것 같이 많이 있다.
3001	CACCCTAGGATAGTCATTGCAACGTGACGCCCTTGTCCTTTCAGGCTCAGGCGCAGCTGGACGGCGCGCGC
114	D S H C N V T P L S F Q V Q A Q L Q L D S A V A H A H H H L H P H L A A H A
114	V Q A Q L Q L D S A V A H A H H H L H P H L A A H A
3121	CGCCTTACATGATGTTCCCGGCACCGCCCTTCGGACTGCCGCCGCGCGCG
	SH3-binding P-loop
152	FIMMFFAFFFGLFLATLAADSASAAB5 VVAAAAAAAKT 123 523 K
140	PYMMFPAPPFGLPLATLAADSASAA S5 VVAAAAAAKT 1223 S23KM
3241	ACTCCAGCATCGCGGATCTCAGACTGAAAGCTAAAAAGCACGCGGCCCCCCGGGTCTG <u>TGACCCCGGCCCACGGCCCACGGTCGGGGGCCCACGGCCCCACGGCCCCCCAAGCGCCCCCGGCGCCCCCC</u>
192	≌234 <sup>©</sup> 235 <sup>I</sup> A D L R L K A K K H A A A L G L * <b>210</b>
180	S <sub>23</sub> S <sub>235</sub> I A D L R L K A K K H A A A L G L * <b>198</b>
3361	TCCGCGACCGGCTTCTCCCCGCACCCGCTTCTGACCGTCGCCCAGGCCTGTCCCCTCCCCGCTGACTGCCCCCTTTTCTTCTGCACCCTGGATCC 3455

FIG. 3. Composite partial nucleotide sequence of OG-12 genomic DNA and two species of OG-12 cDNA and predicted amino acid sequences. The broken line indicates a portion of DNA whose nucleotide sequence was not determined. Consensus sequences for intron ends are underlined and sites for splicing are indicated by arrowheads. The nucleotide sequences and amino acid sequences of portions of the homeodomain are enclosed in dark-shaded boxes. Two species of OG-12 cDNA were cloned that correspond to nt 189–380, 1042–1099, 1420–1508, and 3009- or 3045–3302. The termination codon is underlined and is represented by an asterisk. Putative sites for binding to an SH3 domain or for binding to a nucleotide (P-loop) are indicated by light shading. Putative sites for phosphorylation are showed in white on black backgrounds. The numbers correspond to the following enzymes that may catalyze phosphorylation (14): 1, tyrosine protein kinase; 2, protein kinase C; 3, cGMP-dependent protein kinase; 4, Ca<sup>2+</sup>/calmodulin-dependent protein kinase; 5, casein kinase I or II.

are abundant in the C-terminal region of OG-9 protein, comprising 24% of the total amino acid residues from the end of the homeodomain to the C-terminal residue of the protein.

OG-12 Homeobox Gene. In Fig. 3 are shown a composite nucleotide sequence obtained from an OG-12 genomic DNA clone and two species of OG-12 cDNA (clones A and B), which correspond to different alternatively spliced species of mRNA, and the deduced partial amino acid sequences of OG-12 proteins. Nucleotide residues 1-188 correspond to part of an intron, which is followed by an exon that encodes 18 amino acid residues preceding the homeodomain and homeodomain amino acid residues 1-46. The homeobox is interrupted by an intron, approximately 2 kb long. A small exon follows that encodes only 20 amino acid residues, including homeodomain amino acid residues 47-60. Sequence comparisons between OG-12 genomic DNA and OG-12 cDNA clones A and B, which were isolated from a 12-day mouse embryo library (Novagen) show that the region of the OG-12 gene encoding the Cterminal region of OG-12 protein contains two additional exons: exon 3, encoding 29 amino acid residues (nt 1420-1508 in Fig. 3), which is followed by an intron, approximately 1.5 kb long. Alternative 3' splice sites were found at nt 3008 and 3044.

Hence, OG-12 protein corresponding to OG-12A cDNA has 12 more amino acid residues than the protein corresponding to OG-12B cDNA. Both OG-12 proteins have a putative region for binding to an SH3 domain (15–17) and a potential site for the binding of a nucleotide phosphoryl group, similar to the P-loop motif of ATP/GTP binding proteins (19).

Nucleotide Sequence of the OG-2 Homeobox Gene. The partial nucleotide sequence of a genomic DNA clone for the putative novel OG-2 homeobox gene, and the predicted amino acid sequence of OG-2 protein are shown in Fig. 4. The predicted OG-2 homeodomain is encoded by three exons; putative splice sites are found between homeodomain amino acid residues 10-11 and 46-47. At least two more exons seem likely after the homeobox in OG-2 genomic DNA: a short exon encoding 29 amino acid residues, rich in Pro (nt 1180–1266), followed by a longer exon with two alternative 3' sites for RNA splicing (nt 1495 or 1542). Intron-exon junctions shown for OG-2 genomic DNA were confirmed by nucleotide sequence analysis of mouse embryo OG-2 cDNA clones, which correspond to alternatively spliced forms of OG-2 mRNA that will be reported elsewhere (F. Addivinola, A.C.R., and M.N., unpublished results). The C-terminal region of OG-2 protein

1	TCCCTAGAAGTGACCTGTCAATTTCGGAAAAAAGACCCCGAACCCTGTACCGCTCAGGTTAGTCCTCTGAGGTGGAGGGAAGAGTATATATA
1	S L E V T C Q F R K K T R T L Y R S
121	GTTGGAGGGTGCAGTCACTTTGCCCTGAATGTTTCTCTTTGTCTATGGCAATGTCAATACTTTTGACCCCTGAAGGGAAGACCAGGCAGG
241	GATTCAGAAATTTATATATCACATGGGAGAAGGCCTGGCTCTTGGGTTGGGGCCAATGGCCTTATCTGCAGAAGGCACAAAGAGCATCTCCTTCTCCCTGCTATGG
	▼
361	TGGCTGGCCTCTTCCATTACGTCTCTCTCTCTCTTTTTTTT
19	DOLEELERIFOEDH M, PDS, DKRHEIS,
481	CCCAGATGGTGGGGGTAACCCCCCCAAAGAATCATGGTAAAAGGGACTGGATAATTGGTCAGAAGAGTAATCTGGAAGCTGCTGGTTAGAACTCTCTAGGAGAGGGCCATGTGCTCAGCCTC
44	OMVGVTPORIM
601	AGCTGAGATAGAGGCTGAAGTTAGTCTCAGATGCAAGGGTGTCCCAAGTGTTCCAAGGGCCCACAGCTTTAACCAATTAAAAGAGAAAGATACCACAGAGATCTTTTTATACTTCTGAGCT
	be for the set of the
721	ABGAGGTAGCTGAAACCCAGGAAAGTGTCCTCCACAATCATGCCTAAGCCTTTGTCTCTGCCAGCATTCTAAGTGTGGGTGCCCTTCCTCATTCCTAGTGTGGTTTCAGAACCGCAG
55	V W F Q N R R
841	GGCAAAGTGGAGAAAAGTGGAGAAACTGAACGAGAAGGAAACTAAGAATGGTCCTGCAGCCCCCAGTGCTGACAGCAGCACCAGGTGAGAATGGTCTCTGGTTGTTGGTGTCCACA
62	A K W R K V E K L N E K E . K N G P A A P S. A D S S O H R
961	CCAAGCAGAAGTGGAGTCAGCATGGGAGACCTGGGGTCTTGAAGGAGAACGTTGGGTATGGAGAGTAAGAGCTAGGGATATTGACAAATTCTTTAAGGGATCTGGGCTGGGCTGGAGACA
1081	agagerageragerageragerageragerageragerag
01	
1201	
00	
1321	
1921	
1441	AGCTGTCTCCCCATCTTCTTATGCTGAGT <u>GCCTCTTTTGTCATATTTTCCTAGA</u> GCCCCCCATGCTGCTGACCGACTCTGACCGACTCTGA <u>CCCCCTTTCAGA</u> ATAATGAGGGTGCTGAGA
120	P P M L L T S E Q M <sub>5</sub> L T P F Q N N E G A E R
1561	GGGTGGCAGTGACTCCACCACTCCCCAGTCCCCCCCCCTTTCGAAGGGCCAACCTTCCTCTTCGGGTCCTGTTCAAACTCCTCAAGTACTGCCTCCAATGAGGGACGTTCCTGGCA
	SH3 binding
142	VAVW, PPLLS, PPPIRRANLPLPLGPVQW, PQVLPPMRDVPGS
1981	GTGACAGCATATACAAGGACAAGGCCTATGTGTCCTGGGGCACTAGGTATGGGACCT <u>GA</u> 1739
182	DSIYKDKAYV 🛃 W G 🖬 RYG 🖬 🛪 199

FIG. 4. Partial nucleotide sequence of OG-2 genomic DNA and predicted amino acid sequence. The nucleotide sequence and the predicted amino acid sequence of the OG-2 homeodomain are indicated by dark-shaded areas. Light shading represents a possible site for binding to an SH3 domain. Consensus sequences for intron ends are underlined. Sites for splicing are indicated by black arrowheads. The termination codon is underlined and is represented by an asterisk. Putative amino acid residues that may be phosphorylated are shown in white on black backgrounds. The numbers correspond to possible enzymes catalyzing phosphorylation (14, 20, 21): 1, tyrosine protein kinase; 2, protein kinase C; 3, cGMP-dependent protein kinase; 4, Ca<sup>2+</sup>/calmodulin-dependent protein kinase; 5, casein kinase I or II; 6, histone H1 kinase; 7, mitogen-activated protein kinases.

is rich in Pro (23%) and Ser or Thr residues (19%) and contains a putative SH3 domain binding site. Potential sites for phosphorylation catalyzed by protein kinases also are shown (14, 20, 21).

Expression of Uncx-4.1, OG-9, OG-12, and OG-2 Genes in Mouse Embryos and Adult Mouse Tissues. Northern blot analyses of  $poly(A)^+$  RNA from mouse embryos 7, 11, 15, and 17 days post coitum and from different adult mouse tissues are shown in Fig. 5 A and B, respectively. As shown in Fig. 5A, a 1.1-kb species of Uncx-4.1 RNA is present in  $poly(A)^+$  RNA from 7- through 17-day mouse embryos that increases in abundance by the 15th day of embryonic development. Two additional species of *Uncx-4.1* RNA were found in  $poly(A)^+$  RNA from 11-, 15-, and 17-day mouse embryos, a major band of RNA 2.2 kb long and a minor band of 4.5 kb. Two major bands of *OG-9* RNA, approximately 1.8 and 2.8 kb long, and several minor bands were found in  $poly(A)^+$  RNA from 7-through 17-day embryos. A diffuse band of *OG-9* RNA (0.5 kb) also was detected in  $poly(A)^+$  RNA from 15- and 17-day embryos. One band of *OG-12* RNA, about 5.0 kb long, is



FIG. 5. Northern blot hybridization of poly(A)<sup>+</sup> RNA from mouse embryos (A) or adult mouse tissues (B). Each lane contains 2  $\mu$ g of poly(A)<sup>+</sup> RNA from mouse embryos 7, 11, 15 or 17 days *post coitum* (A) or from adult mouse heart (H), brain (B), spleen (S), lung (L), liver (Li), skeletal muscle (SM), kidney (K), or testis (T) (B). Blots were hybridized with antisense RNA probes (*Uncx-4.1, OG-2, or OG-12*) or a denatured double-stranded DNA probe (*OG-9*) labeled with <sup>32</sup>P. Probes correspond to the following nucleotide sequences, which are downstream of homeobox sequences: *Uncx-4.1*, nt 622–742, Fig. 1; *OG-9*, nt 400–660, Fig. 2; *OG-12*, nt 1081–1492, Fig. 3; *OG-2*, nt 866-1123, Fig. 4.

SP	ECIE	S	% HOM-	REF.
		10 20 30 40 50 60	OLOGY	
Uncx-4	m	RRTRTNFTGWQLEELEKAFNESHYPDVFMREALALRLDLVESRVQVWFQNRRAKWRKKE	100	*
unc-4	се	<b>S</b> S <b>E</b> A <b>MLR</b> -	88-97	18
		a di katalah di katalah di katalah di katalah di katalah 🖬 di katalah 🖬 di katalah 🖬 di katalah 👘 di katalah di kat		
OG-9	m	HRRKRTTFSVGQLVELERVFAARPYPDISTREHLAQVTHLPEAKIQVWFQNRRAKRIKDR	100	*
Phox2	m	QITSAKETHYELKIE-TRVFR-QE	63-73	23
prd	d	QCA <b>S</b> DA-ERTQYER-N- <b>TR</b> S <b>R</b> LR- <b>QH</b>	63-73	24
OG-12	m	QRRSRTNFTLEQLNELERLFDETHYPDAFMREELSQRLGLSEARVQVWFQNRRAKCRKQE	100	*
Phox2	m	ITSAKV-AIYTALKIE-TF	72-82	23
pax-3	m	TAEA-ERIYTAAK-TSRWA	72-82	25
		Y		
0G-2	m	RKKTRTLYRSDQLEELERIFQEDHYPDSDKRHEISQMVGVTPQRIMVWFQNRRAKWRKVE	100	*
al	d	QRRYTFT-FKA-SRTVFT-E-LA-KI-L-EAQQ-	55-67	26
OG-12	m	QRRSNFTLENL-D-TAFM-E-LRL-LSEA-VQCQ-	53-78	*
repo	d	KTFTAYA-ERAPVFA-E-LAIKLNLSES-VQH-	53-70	27

FIG. 6. Similarity of Uncx-4.1, OG-9, OG-12, and OG-2 homeodomains to known homeodomains. Each predicted homeodomain amino acid sequence is aligned with the most closely related known homeodomains. Species abbreviations are as follows: m, mouse; ce, *C. elegans*; and d, *Drosophila*. A dash represents an identical amino acid residue; conservative amino acid substitutions (according to figure 84 in ref. 22) are shown in boldface type; conservative amino acid replacement families are as follows: G, A, P, S, T; V, M, I, L; F, Y, W; D, E, N, Q; K, R, H; and C. Each pair of values separated by a dash in the percent homology column corresponds to the following: percent of identical amino acid residues; and the percent of identical amino acid residues plus conservative amino acid replacements. Asterisks in the reference column correspond to this report.

present in poly(A)<sup>+</sup> RNA from 7-day mouse embryos. The most abundant species of OG-12 RNA detected in poly(A)<sup>+</sup> RNA from 11-day embryos is 4 kb long; trace bands of RNA, about 3.4, 5.0, and 8.0 kb, also were found. Poly(A)<sup>+</sup> RNA from 15- and 17-day embryos contains 3.4-, 4.0-, 6.0-, and 8.0-kb bands of OG-12 RNA. One major band of OG-2 RNA, 6 kb long, and trace bands, 2.6, 4, and 7 kb long, were found in poly(A)<sup>+</sup> RNA from 15- and 17-day embryos.

Northern blot analyses of  $poly(A)^+$  RNA from adult mouse tissues are shown in Fig. 5B.  $Poly(A)^+$  RNA from brain contains five major bands of Uncx-4.1 RNA, approximately 0.7, 1.0, 1.9, 6.0, and 8.0 kb long, and a minor band of 4 kb. The 0.7and 1.0-kb species of Uncx-4.1 RNA also were found in kidney  $poly(A)^+$  RNA; whereas, a 0.7- and 1.3-kb species of Uncx 4.1 RNA were found in  $poly(A)^+$  RNA from testis.  $Poly(A)^+$  RNA from spleen, lung, and liver contain prominent 0.7-kb bands of Uncx-4.1 RNA; the abundance of this species of RNA is low in RNA from heart and striated muscle.

One major band of OG-9 RNA, 1.4 kb long, was detected in poly(A)<sup>+</sup> RNA from somatic muscle and a trace band, approximately 2.8 kb long, was found in poly(A)<sup>+</sup> RNA from brain; however, little or no OG-9 RNA was detected in other tissues tested.

OG-12 RNA was detected in all tissues tested; however, the size and number of bands of OG-12 RNA differed, depending upon the tissue. Poly(A)<sup>+</sup> RNA from striated muscle contained the most abundant band of OG-12 RNA (about 2 kb long).

Two prominent bands of OG-2 RNA, 1.9 and 2.6 kb long, and a trace 4.4-kb band of RNA were found only in poly(A)<sup>+</sup> RNA from somatic muscle. A trace band of OG-2 RNA, 2.6 kb long, was found in poly(A)<sup>+</sup> RNA from heart, and trace bands, approximately 3.5, 4.0, and 6.0 kb long, were found in poly(A)<sup>+</sup> RNA from liver and kidney. However, little or no OG-2 RNA was found in poly(A)<sup>+</sup> RNA preparations from other tissues tested.

These results show that multiple bands of Uncx-4.1, OG-9, OG-12, and OG-2 RNA are present in  $poly(A)^+$  RNA from mouse embryos and adults, that species of Uncx-4.1, OG-9, OG-12, and OG-2 poly(A)<sup>+</sup> RNA change during development, and that each gene is expressed in the adult mouse in a tissue-specific pattern.

Homeodomain Homologies. In Fig. 6, the predicted amino acid sequences of the homeodomains encoded by Uncx-4.1,

OG-9, OG-12, and OG-2 genes are compared with the most closely related homeodomains described thus far. The homeodomain of Uncx-4.1 is 88% identical and 97% similar (identical amino acid residues plus conservative amino acid substitutions) to the unc-4 homeodomain of C. elegans. unc-4 homeodomain protein probably regulates the expression of one or more genes that encode proteins that are required in the nematode to form proper synaptic connections between interneurons and some members of a class of ventral cord motoneurons (18, 28). An intron is present between codons for homeodomain amino acid residues 46 and 47 in both mouse Uncx-4.1 and C. elegans unc-4 genes; however, another intron in the unc-4 gene between codons for homeodomain amino acid residues 9 and 10 is not present in the mouse Uncx-4.1 gene.

The OG-9 homeodomain is distantly related to the mouse Phox2 (23) and *Drosophila* prd (24) homeodomains (both 63% homology) while the OG-12 homeodomain is most closely related to the mouse Phox2 (23) and pax-3 (25) homeodomains (both 72% homology). However, Gln residues are found at position 50 of the OG-9 and OG-12 homeodomains; whereas, Ser-50 is found in the paired family of homeodomains, i.e., *Drosophila* prd, gsb-p and gsb-d, and vertebrate Pax homeodomains (for review, see ref. 29). Amino acid residue 50 plays a role in DNA base recognition (30), and amino acid substitutions at this position have been shown to alter the nucleotide sequence specificity of prd (31) and bicoid (32).

The homeodomain of OG-2 is a relatively diverged homeodomain that exhibits only 53-55% homology to the homeodomains of *Drosophila al* (26), OG-12, and repo (27), a glial-specific homeodomain protein. The OG-2 homeodomain is the only homeodomain that has been found thus far that contains a Met residue at position 46. Most homeodomains contain Lys-46, or less frequently, Arg-46. A single nucleotide substitution would be sufficient to change a Lys codon, AAG, to a Met codon, ATG.

Each novel gene contains an intron within the homeobox between codons for the 46th and 47th amino acid residues of the homeodomain. Another intron is present in the OG-2 gene, between homeodomain amino acid residues 9 and 10. Other homeobox genes from *C. elegans*, *Drosophila*, and mouse have been reported that also have an intron between the 46th and 47th and/or the 9th and 10th amino acid residues of the homeodomains (for review, see ref. 33).

Further work is needed to localize Uncx-4.1, OG-9, OG-12, and OG-2 RNAs at the tissue and cellular levels, both in embryos and adult mice.

We thank Dr. Konrad Huppi, National Cancer Institute, National Institutes of Health, for the gift of a mouse genomic DNA library, Derek Tang and Vicky Guo for excellent technical work, and Yoshinobu Hara, Dervla Mellerick-Dressler, and Alan Peterkofsky for comments on the manuscript.

- Gehring, W. J., Affolter, M. & Bürglin, T. (1994) Annu. Rev. Biochem. 63, 487-526.
- Manak, J. R. & Scott, M. P. (1994) Development (Cambridge, U.K.), Suppl., 61–77.
- Stein, S., Fritsch, R., Lemaire, L. & Kessel, M. (1996) Mech. Dev. 55, 91-108.
- Holland, P. W. H., Garcia-Fernandez, J., Williams, N. A. & Sidow, A. (1994) Development (Cambridge, U.K.), Suppl., 125– 133.
- Sawadogo, M. & Van Dyke, M. V. (1991) Nucleic Acids Res. 19, 674–675.
- Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1990) Current Protocols in Molecular Biology (Wiley, New York).
- Zhao, G.-Q., Zhou, X., Eberspaecher, H., Solursh, M. & de Crombrugghe, B. (1993) Proc. Natl. Acad. Sci. USA 90, 8633– 8637.
- Tan, D.-P., Ferrante, J., Nazarali, A., Shao, X., Kozak, C. A., Guo, V. & Nirenberg, M. (1992) Proc. Natl. Acad. Sci. USA 89, 6280-6284.
- Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M. & Boncinelli, E. (1992) *EMBO J.* 11, 2541–2550.
- Simeone, A., D'Apice, M. R., Nigro, V., Casanova, J., Graziani, F., Acampora, D. & Avantaggiato, V. (1994) *Neuron* 13, 83–101.
- Kern, M. J., Witte, D. P., Valerius, M. T., Aronow, B. J. & Potter, S. S. (1992) Nucleic Acids Res. 20, 5189–5195.
- Cserjesi, P., Lilly, B., Bryson, L., Wang, Y., Sassoon, D. A. & Olson, N. E. (1992) Development (Cambridge, U.K.) 115, 1087– 1101.
- 13. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) Annu. Rev. Biochem. 55, 1119–1150.
- 14. Pearson, R. B., Kemp, B. E. (1991) Methods Enzymol. 200, 62-81.

- Ruibao, R., Mayer, B. J., Cicchetti, P. & Baltimore, D. (1993) Science 259, 1157–1161.
- Gout, I., Dhand, R., Hiles, I. D., Fry, M. J., Panayoton, G., Das, P., Truong, O., Totty, N. F., Hsuan, J., Booker, G. W., Campbell, I. D. & Waterfield, M. D. (1993) *Cell* **75**, 25–36.
- Barfod, E. T., Zheng, Y., Kuang, W.-J., Hart, M. J., Evans, T., Cerione, R. A. & Ashkenazi, A. (1993) J. Biol. Chem. 268, 26059-26062.
- Miller, D. M., Shen, M. M., Schamu, C. E., Bürglin, T. R., Ruvkun, G., Dubois. M. L., Ghee, M. & Wilson, L. (1992) *Nature* (London) 355, 841–845.
- Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990) Trends Biochem. Sci. 15, 430-435.
- Clark-Lewis, I., Sanghera, J. S. & Pelech, S. (1991) J. Biol. Chem. 266, 15180–15184.
- Gonzales, F. A., Raden, D. L. & Davis, R. J. (1991) J. Biol. Chem. 266, 22159–22163.
- Dayhoff, M. O., Schwartz, R. M. & Orcutt, B. C. (1978) in Atlas of Protein Sequence and Structure, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Silver Spring, MD) 5, 345–352.
- Valarche', I., Tissier-Seta, J.-P., Hirsch, M.-R., Martinez, S., Goridis, C. & Brunet, J.-F. (1993) Development (Cambridge, U.K.) 119, 881–896.
- Frigerio, G., Burri, M., Bopp, D., Baumgartner, S. & Noll, M. (1986) Cell 47, 735–746.
- Goulding, M. D., Chalepakis, G., Deutsch, U., Erselius, J. & Gruss, P. (1991) EMBO J. 10, 1135–1147.
- 26. Schneitz, K., Spielmann, P. & Noll, M. (1993) Genes Dev. 7, 114-129.
- Xiong, W.-C., Okano, H., Patel, N. H., Blendy, J. A. & Montell, C. (1994) Genes Dev 8, 981–994.
- White, J. G., Southgate, E., Thomson, J. N. (1992) Nature (London) 355, 838-841.
- 29. Hill, R. E. & Hanson, I. M. (1992) Curr. Opin. Cell Biol. 4, 967–972.
- Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G. & Wüthrich, K. (1994) Cell 78, 211–223.
- Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. & Despian, C. (1989) Cell 59, 553–562.
- 32. Hanes, S. & Brent, R. (1989) Cell 57, 1275-1283.
- 33. Bürglin, T. R. (1994) in *Guidebook to the Homeobox Genes*, ed. Deboule, D. (Oxford Univ. Press, New York), pp. 27-71.