

Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes

Beat W. Schäfer*, Thomas Czerny¹, Michele Bernasconi, Michele Genini and Meinrad Busslinger¹

University of Zürich, Department of Pediatrics, Division of Clinical Chemistry, Steinwiesstrasse 75, 8032 Zürich, Switzerland and ¹Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, 1030 Vienna, Austria

Received August 25, 1994; Revised and Accepted October 11, 1994

EMBL accession no. Z35141

ABSTRACT

The myogenic basic helix-loop-helix proteins are essential components of the regulatory network controlling vertebrate myogenesis. However, determined myoblasts appear in the limb buds which do not initially express any member of this transcription factor family. In a search for potential novel regulators of myogenesis, a human PAX-7 cDNA was isolated from primary myoblasts. Analysis of the DNA-binding properties of the Pax-7 paired domain revealed that it binds DNA in a sequence-specific manner indistinguishable from that of the paralogous Pax-3 protein. Each of the two proteins also binds to palindromic homeodomain-binding sites by cooperative dimerization. Both Pax-3 and Pax-7, which are known to partially overlap in their expression during development, can also efficiently form heterodimers on these sites and stimulate reporter gene transcription in transient transfection experiments which, in the case of Pax-7, is dependent on the transactivation function encoded by the C-terminal sequences. Thus, the formation of heterodimers might have important consequences for target gene recognition and regulation during development. PAX-7 was found to be weakly expressed in normal human myoblasts, while PAX-3 could not be detected in these cells at all. However, transcripts for either PAX-3 and/or PAX-7 were expressed at elevated levels in tumorigenic rhabdomyosarcoma cell lines. Hence, overexpression of these PAX genes may be involved in the genesis of myogenic tumors.

INTRODUCTION

Myogenesis is a multistep process during which pluripotent stem cells are committed to the myogenic lineage and subsequently are induced to express many muscle specific genes. Expression of these genes is, in part, regulated by a cis-regulatory DNA motif called the E box (CANNTG), which can be bound by the transcription factors myoD (myf3), myogenin (myf4), myf5 and

MRF4 (myf6) of the myogenic basic helix-loop-helix (bHLH) protein family (for review, see ref. 1,2). Recent gene targeting experiments in mice demonstrated that mice lacking either both myoD and myf5, or myogenin alone display severe skeletal muscle deficiencies (3–5) which underscores the importance of this family of transcription factors for myogenesis. In addition, these experiments placed myoD and myf5 upstream of myogenin in a possible transcriptional hierarchy (6). The myogenic bHLH factors have furthermore the potential to convert several non-muscle cell types to the muscle phenotype upon ectopic expression (7). However, expression of these factors does not necessarily lead to the activation of the myogenic program in all cells as has been shown for the cell lines HeLa or HepG2 (7,8). In these cells, the expression of muscle specific genes can, however, be induced by cell fusion with myoblasts. In addition, there is substantial evidence for the existence of committed myogenic cells in limb buds of vertebrate embryos that do not express any of the myogenic bHLH regulators (9,10). Such observations led to a search for additional regulators involved in the establishment of the myogenic lineage and muscle specific gene expression. Indeed, a family of four gene products called myocyte enhancer factor 2 (MEF2) (11) has been described that contain the MADS domain (12) and that are able to bind an A/T rich DNA motif found in several muscle specific genes.

Homeodomain proteins which represent another important family of transcription factors are known to play key roles in the establishment of body axes and the regulation of developmental programs. So far little is known about homeobox containing genes expressed during myogenesis. One gene, called MHox, has been reported to bind to an essential site in the enhancer of the muscle creatine kinase gene (13). The expression pattern of MHox suggests, that the gene product may be involved in the specification of somites (14). In addition, two members of the Pax gene family, Pax-3 and Pax-7, comprising both a paired domain as well as a paired-type homeodomain, are expressed in the dermomyotomal part of the somites, from which muscle cells derive in the developing mouse embryo (15,16). Both genes are also transcribed in the dorsal region of the neural

*To whom correspondence should be addressed

tube as well as in specific brain regions. In somites, transcripts begin to accumulate between day 8.5 (Pax-3) and day 9 (Pax-7). Later, during somite maturation, Pax-3 expression becomes confined to the ventrolateral part of the dermomyotome (17). Whereas the expression of Pax-3 and Pax-7 is overlapping in the somitic region, only Pax-3 expression persists in prospective myoblasts migrating from the somitic dermomyotome into the limb buds (18). Notochord transplantation experiments indicated that signals from the notochord determine the expression pattern of Pax-3 and Pax-7 in somites and alter the subsequent differentiation of cell types, including muscle cells, arising from the somitic mesoderm (17,19). These experiments suggest that Pax genes might play an important role in the development of limb muscles.

The Pax genes constitute a family of nine genes in the mammalian genome (for review see ref. 20,21). They have been subdivided into six classes based on sequence homology (22). Recent evidence suggests that Pax proteins can act as transcription factors, since specific DNA binding as well as transactivation has been shown for a number of Pax proteins (reviewed in 23). Three products of Pax genes have been associated with mouse developmental mutants, namely Pax-1 with *undulated* (24), Pax-3 with *Splotch* (25) and Pax-6 with *Small eye* (26). Furthermore, mutations in the human PAX-3 and PAX-6 genes cause Waardenburg's syndrome (27-29) and aniridia (30,31), respectively. In addition, the PAX-3 gene has been shown to be altered by a specific translocation, t(2;13)(q35;q14), in alveolar rhabdomyosarcoma, leading to a chimeric gene product between PAX-3 and a novel member of the fork head family of transcription factors (32). Together these data suggest that Pax genes are important regulators of developmental processes.

To study the role of PAX-7 in the establishment of the myogenic lineage, we have isolated and characterized a human PAX-7 cDNA. Here we show that PAX-7 is expressed at low levels in primary human myoblasts and at elevated levels in cell lines derived from rhabdomyosarcomas. Analysis of the DNA binding and transactivation properties of the PAX-7 protein demonstrated that PAX-7 is a sequence-specific transcription factor expressed during human myogenesis.

MATERIALS AND METHODS

Cell lines and primary cultures

Human rhabdomyosarcoma cell lines A673 and RD were obtained from American Type Culture Collection (Rockville, MD), and the cell lines Rh1, Rh18 and Rh30 were a generous gift of Dr P.Houghton. Primary human myoblasts were isolated from autopsy material (gastrocnemius) of a 6 month old boy as described (33). To obtain pure myoblast cultures, cells were cloned in 96 well microtiter plates subsequent to isolation and tested for cell fusion by switching from growth media (GM: Ham's F-10, 15% fetal calf serum) to fusion media (FM: DME, 2% horse serum, 2.5×10^{-6} M dexamethasone, 10^{-6} M insulin). The murine plasmacytoma cell line J558L was obtained from Dr U.Chen.

Isolation and sequencing of a human PAX-7 cDNA

RNA was isolated from primary human myoblasts (clone B6M/7) according to Chirgwin *et al.* (34). Three micrograms of total RNA were reverse transcribed with a mixture of random hexamers and oligo(dT) using Moloney murine leukemia virus reverse transcriptase (Superscript, Life Technologies). 1/20th of

this reaction was used as template for a polymerase chain reaction (PCR) with the degenerated primers HoxF and HoxG. PCR was performed for 30 cycles at 94°C for 1 min, 50°C for 1 min and 72°C for 90 s. PCR products of the correct size were subcloned into the *EcoRI/BamHI* site of pBluescript SK⁺ and sequenced using the T7 sequencing kit (Pharmacia) according to the supplier's instruction. For rapid amplification of cDNA 5' ends (5' RACE; 35), RNA was reverse transcribed with random hexamers as described above. A polyA tail was added with terminal transferase and two subsequent rounds of amplification carried out, using an annealing temperature of 58°C and primers M3-1REV/M3-1REVI as sequence specific primers. A single band of ~900 bp was obtained, gel purified (GeneClean, Bio101) and subcloned into the *BamHI/HindIII* sites of pBluescript SK⁺. The 3' RACE was performed with cDNA obtained by reverse transcription with the oligonucleotide RACEtot (35). In order to enhance the yield of PCR products, 5 mM MeHgOH and 30 mM β -mercaptoethanol were added to the reaction. After amplification with gene specific primers M3-1FOR/M3-1FORI, a number of bands were obtained, gel purified and subcloned as described above. Sequencing identified B2-5 as the longest cDNA clone (681 bp). The longest 5' and 3' RACE products (III-9 and B2-5) were combined taking advantage of a common *StuI* site in primer M3-1FORI yielding cDNA clone D8 which was sequenced on both strands. To verify the sequence of the cloned cDNA, overlapping RT-PCR products covering the entire open reading frame were generated and directly sequenced. We found a discrepancy at two positions (position 1046 C to T and position 1310 G to A) between the direct sequencing and the cDNA clone D8. In Figure 1, the bases determined by direct sequencing are shown, assuming that clone D8 contains two base exchanges due to Taq polymerase errors. PCR cloning of the murine Pax-3 cDNA has been described (36).

Oligonucleotides

Oligonucleotides and their sequences are as follows: HoxF, CCGAATTCCG(A/C)CG(A/C)(G/A)(C/G)(C/T)CG(C/G)ACC-(G/A)C(C/G)TT(C/T)AC; HoxG, CGTGGATCC(G/C)CG-(G/C)CGGTT(C/A)(T/C)(C/A)GAACCA; M3-1FOR, GGC-CGAGCAGCTGGAGGAG; M3-1FORI, CACGAATTCC-TGGAGAAGGCCTTTGAGAGG; M3-1REV, GACCTGC-ACACGCGCCTCT; M3-1REVI, CGTGGATCCCCATTG-ATGAAGACCCCTC. All oligonucleotides used for EMSA were previously described (36,37) except for the H2A-17C oligonucleotide (TGTGACGCAGCGGTGCGTGACGACTT).

RNase protection and Northern blot analysis

PAX-3 and PAX-7 transcripts were detected in 0.5 μ g polyA⁺ RNA (rhabdomyosarcoma cells) or 10 μ g total RNA (primary myoblasts) by RNase protection analysis (RPA II Kit, Ambion) as described by the manufacturer. The human PAX-7 probe was obtained by cloning a blunt-ended 302 bp *Asp718-BstXI* fragment from clone D8 into the *SalI* site of pSP64. For the generation of the human PAX-3 probe a blunt ended 152 bp *Asp718-HincII* fragment was cloned into the *SalI* site of pSP64. Both plasmids were linearized with *EcoRI* in order to synthesize α -³²P-dUTP labeled antisense RNA (MAXIscript, Ambion). The internal reference used as probe was a 267 bp *XbaI-MscI* glyceraldehyde-phosphate-dehydrogenase (GAPDH) fragment in pSP65, which was linearized with *BstXI* resulting in a 90 bp protected fragment. Northern blot analysis was carried out according to standard procedures (38). For PAX-7, the entire

insert of clone D8 was used. For the myogenic factors, the following 3' end fragments were gel purified and labeled with ^{32}P -dATP by random priming (Prime-a-Gene, Promega) in order to avoid cross-hybridization with other myogenic factors in the conserved basic helix-loop-helix region: a 800 bp *EcoRI* fragment of *myf3*, a 890 bp *EcoRI*-*PstI* fragment of *myf4*, a 650 bp *EcoRI*-*SphI* fragment of *myf5* and a 600 bp *EcoRI*-*XhoI* fragment of *myf6*. RNA molecular standards (Boehringer) were used to estimate the sizes of the transcripts.

EMSA and transactivation assays

The different cDNAs used in this study were cloned in the sense orientation downstream of the cytomegalovirus (CMV) enhancer/promoter region and the SP6 promoter of the expression vector pKW10 (39). For the construction of PAX-7 Δ an internal *PstI* site was used to delete the C-terminal sequences. For EMSA either whole cell extracts from transiently transfected COP-8 cells or *in vitro* translated protein was used as described (36). For the transactivation assays, the reporter and transactivator plasmids (3 and 1 μg , respectively) were cotransfected transiently into J558L cells by electroporation. After three freezing/thawing cycles and a final centrifugation step, the supernatant was used to determine the luciferase activity. The reporter plucTKCD19 was constructed by inserting multimerized CD19-2(A-ins) oligonucleotides (36) upstream of the TK promoter of plucTK. The plasmid lucTK was constructed by inserting a TK promoter fragment from -109 to +52 of the herpes simplex virus upstream of a luciferase reporter gene in the pSP64 vector.

RESULTS

Isolation of a human PAX-7 cDNA

To date, little is known about the expression and the role of homeobox-containing genes during myogenesis in mammals. In *Drosophila* it was shown that the expression of one homeobox protein, S59, is restricted to a subset of myogenic precursor cells (40). In order to identify homeobox proteins expressed during mammalian myogenesis, we have used degenerated oligonucleotides in a polymerase chain reaction (PCR) to amplify homeobox sequences from human primary muscle cells. These experiments yielded several new homeobox sequences, one of which represented a new member of the POU family of homeobox proteins (41).

In the same experiment, a cDNA clone of 157 bp was isolated (clone F-G-3), which showed a high degree of identity to the mouse Pax-7 homeobox. To obtain a full-length cDNA, several rapid amplification of cDNA ends (RACE) experiments were performed using primers derived from the sequence of clone F-G-3 and RNA isolated from primary human myoblasts (see Fig. 1B). In two independent 5' RACE experiments yielded the same 5' end sequences, leading us to conclude that we have reached the 5' end of the transcript. In contrast, various fragments were obtained from the 3' end. The PAX-7 cDNA, depicted in Figure 1A, was assembled from the longest cDNA clones and contains an open reading frame of 467 amino acids which has an overall homology of 92% compared with the known region of mouse Pax-7 protein. The 3' end of the cDNA lacks, however, a stop codon. Comparison with the mouse Pax-3 sequence suggests that only six amino acids may be missing from the full-length coding region (Fig. 2). The first AUG which is in good agreement with the Kozak consensus sequence (42) was chosen as the translation start site. In addition, sequence comparison with the mouse and

human PAX-3 proteins shows that the equivalent methionine has been assigned as the translation start site. The paired domain and the paired-type homeodomain show an exceptionally high degree of identity between PAX-3 and PAX-7 at the amino acid level (85.5 and 96.7%, respectively), whereas the C-terminal halves of both proteins are only 35.1% identical. Six additional nucleotides leading to the insertion of the two amino acids glycine and leucine were found in the C-terminal part of the paired domain in our cDNA clone when compared with the partial mouse Pax-7 cDNA sequence (15) and the partial human PAX-7 sequence derived from genomic DNA (43). This insertion is due to the use of an alternative 3' splice site upstream of exon 3, and analysis of PCR products spanning this splice junction confirmed the equivalent use of both splice sites in human myoblasts (data not shown). The C-terminal half of PAX-7 is very rich in serine, proline, alanine and glycine which represent 48.5% of all amino acids in this region, suggesting the presence of a possible transactivation domain (see below). Together these data demonstrate that we have cloned, from human myoblasts, a cDNA coding for almost full-length PAX-7, thus allowing the molecular characterization of the corresponding protein.

DNA-binding properties of the PAX-7 protein

The paired domain and the paired-type homeodomain of PAX-7 are almost identical to those of PAX-3. We would therefore expect that the DNA-binding properties of the two proteins are similar, but distinct from those of other Pax proteins like BSAP (PAX-5). To test this hypothesis, we synthesized all three Pax proteins by *in vitro* translation and then quantitated the relative protein concentrations by PhosphorImager analysis (Fig. 3A). Adjusted protein amounts were used for electrophoretic mobility shift assays (EMSA) to determine the relative DNA-binding affinities of the three Pax proteins. A high affinity binding site for BSAP [CD19-2(A-ins)] was used as a probe which is known to be recognized by the paired domains of all main subfamilies of Pax proteins (36). The results in Figure 3B show that BSAP binds to this site at least with a 10-fold higher affinity than both Pax-3 and PAX-7. No difference in binding efficiencies was observed between Pax-3 and PAX-7. To test the specificity of DNA-binding displayed by PAX-7, extracts of COP-8 cells overexpressing the three Pax proteins were analyzed by EMSA with a panel of paired domain recognition sequences. As previously shown (36), PAX-5 bound to all of these sequences, albeit with different efficiencies. In contrast, the more divergent paired domains of Pax-3 and PAX-7 interacted only with a subset of the recognition sequences (Fig. 3C). All of these sites show a perfect match in the 3' consensus motif and belong to the class II sequences defined by Czerny *et al.* (36). Interestingly, no difference in binding specificity was observed between Pax-3 and PAX-7. We conclude therefore, that Pax-3 and PAX-7 recognize similar, if not identical, DNA sequences and that optimal binding sites for Pax-3 and PAX-7 must differ from those of BSAP (PAX-5).

It has recently been demonstrated that *Drosophila* paired-type homeodomains are able to mediate cooperative dimerization on palindromic DNA sequences (37). To test whether the vertebrate Pax-3 and PAX-7 proteins are able to display the same cooperative dimerization, the full-length proteins including the homeodomain and paired domain were subjected to EMSA. As control, a paired domain binding site (H2A-17C) was used, which has previously been shown to be recognized only by monomeric Pax proteins (39). Binding of Pax-3 and PAX-7 to the palindromic

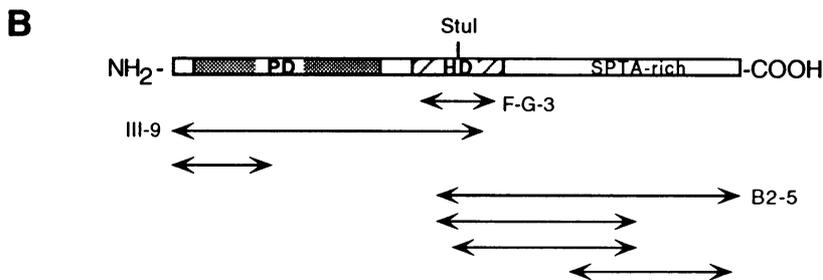
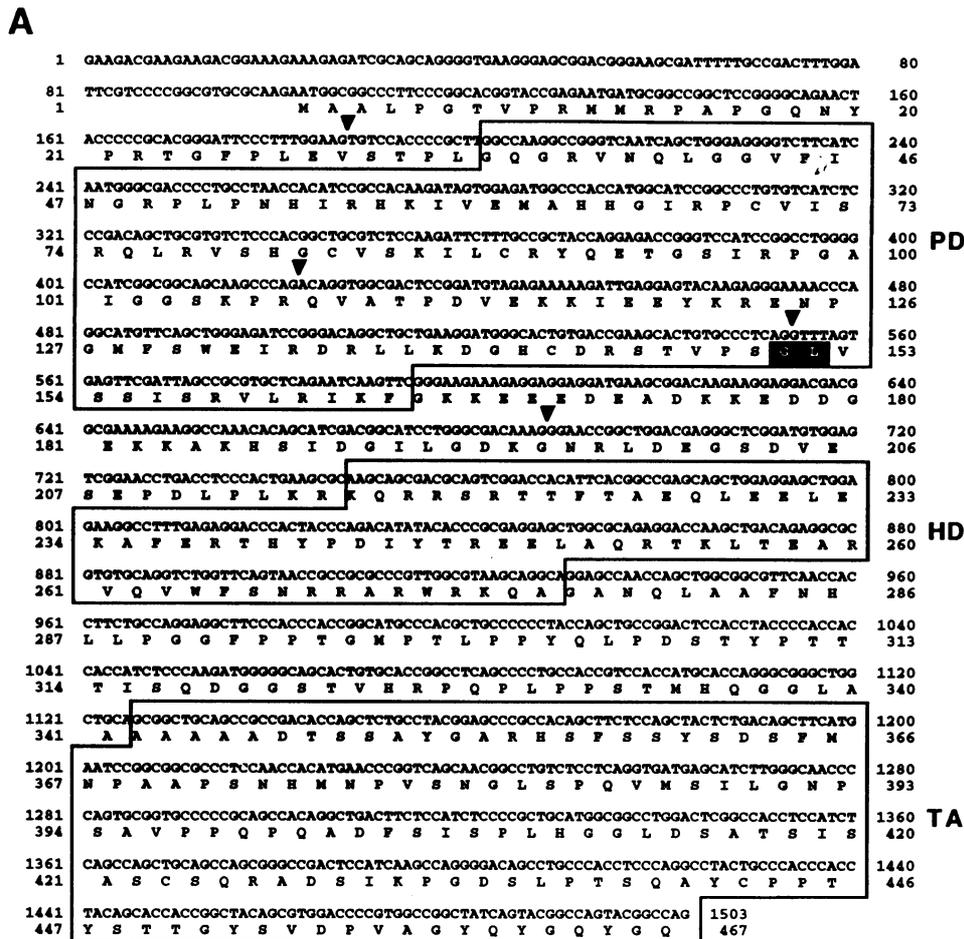


Figure 1. Sequence and structural organization of the human Pax-7 cDNA. (A) cDNA and deduced amino acid sequence for PAX-7. The conserved paired box (PD), the paired-type homeobox (HD) as well as the region containing the transactivation domain (TA) are boxed. Arrowheads indicate the positions of known introns (43). The beginning of the fourth exon was shifted by six bases to accommodate the two additional amino acids (highlighted by block overlay) found in the cDNA. The sequence was verified and corrected at two positions (1046 and 1310) by direct sequencing of overlapping PCR products. (B) Schematic representation of the PAX-7 cDNA cloning by RACE. cDNA products which were obtained in independent experiments are shown. The two longest cDNAs (III-9 and B2-5) were used for compilation of the sequence and for reconstruction of the entire reading frame in clone D8 by fusion at the common *StuI* site. Filled and hatched boxes represent the paired domain and homeodomain, respectively.

P2 site, which is recognized by the paired-type homeodomain (37), revealed an additional complex of lower electrophoretic mobility, which results from dimer formation (Fig. 4A). This experiment indicates that vertebrate Pax-3 and PAX-7 proteins are able to cooperatively dimerize on palindromic recognition sequences most probably via the homeodomain, as has been shown for the *Drosophila* Paired protein (37).

Since Pax-3 and Pax-7 show an overlapping expression profile during mammalian development, we investigated whether they are capable of forming heterodimers by binding to the P2 site. For this purpose, we constructed a PAX-7 deletion mutant (PAX-7Δ) by eliminating the last 126 C-terminal residues, thus leaving both DNA binding domains intact. This truncated PAX-7 cDNA clone was either expressed on its own or in combination

```

mouse Pax7                               1 ----- 61
human PAX7  1 MAALPGTVPRMRPAPGQNYPRTFPLEVSTPLGGQQRVNLGGVFINGRPLPNHIRHKIVEMAHHGIRPCVISRQLRVSHGCVSKILCR 89
human PAX3  1 -TT-A-A-----G-----S----- 89
mouse Pax3  1 -TT-A-A-----G-----S----- 89

62 -----DDE-.G---E----- 157
90 YQETGSIRPGAIGGSKPRQVATPDVEKKIEYKRENPMFSWEIRDRLLDGHCDRSTVPSGLVSSISRVLRIKFGKKEEDE. .ADKKEDDGEKKAKHS 186
90 -----K-T-----K-----AV--N-----I-S--G---EADLER-EAEES-- 187
90 -----K-T-----K-----AV--N-----I-S--G---EADLER-EAEES-- 187

158 -----G-----F----- 255
187 IDGILGDKGN. RLDEGSDVESEPDPLKRRQRRTTFTAEQLELEKAFERTHYPDIYTREELAQRTKLTEARVQVWFSNRRARWRKQAGANQLAAFN 285
188 --I-----A-----T---SETS-QP-S.-P-AVS-PS-----V--STI...PSNP-S---CLPST-G---T---VP-SG---P---T 287
188 --SERASAPQS-----ID-----R-----A-----M--- 287
188 --SERASAPQS-----ID-----R-----A-----M--- 287

256 -----T----- 305
286 HLLPGGFPTGMPTLPPYQLPDSTYPTT. ISQ. .DGGSTVHRPQLPPSTMHQGLAAAAAADTSSAY. .GARHSFSSYSDSFMNPAAPSNHMNP. 376
288 --I-----A-----T---SETS-QP-S.-P-AVS-PS-----V--STI...PSNP-S---CLPST-G---T---VP-SG---P---T 383
288 --I-----A-----T---SEHS-QP-S.-P-AVS-PS-----V--STI...PSNP-S---CLPST-G---T---VP-SG---P---T 383

378 VSNGLSPQVMSILGNPSAVPPQPADFSISPLHGGLDSATSISASCSQRADSIKPGDSLPTSQSYCPPTYSTTGYSVDPVAGYQYQYQGQ 467
384 IGN-----GL-T-HGG--H---T-YAL---T---EPT-TV-----L-HM-SL----- 459
384 IGN-----GL-T-HGG--H---T-YAL---T---EPT-TV-----LEHM-NV-----P-----A---M---T-----SKPWTF* 479

```

Figure 2. Sequence comparison of human and mouse Pax-7 and Pax-3 amino acid sequences. The entire sequence of the human PAX-7 protein is shown, while only those amino acids differing from human PAX-7 are indicated for the other proteins. Gaps were introduced in the sequences for optimal alignment and are shown by dots. The paired domain as well as the paired-type homeodomain are drawn in bold. The asterisk denotes the stop codon found in the mouse Pax-3 cDNA.

with full length Pax-3 in transiently transfected COP-8 cells. As shown by the EMSA analysis in Figure 4B, a complex of intermediate mobility is formed which corresponds to a heterodimer of Pax-3 and PAX-7. Hence, Pax-3 and PAX-7 can efficiently heterodimerize on palindromic recognition sequences.

PAX-7 can act as a transactivator

As Pax-3 and PAX-7 display similar DNA binding properties, we wanted to test whether their transactivation potential would also be alike. To this end, Pax-3 and PAX-7 expression constructs were transfected into the plasmacytoma line J558L, which does not express any endogenous Pax proteins (data not shown). As a reporter gene we used a construct containing six copies of the CD19-2 (A-ins) site upstream of a TK promoter linked to a luciferase gene. Cotransfection of this reporter construct with an expression vector for Pax-3 consistently resulted in a 6-fold increase in luciferase activity, when compared to the expression vector pKW10 alone (Fig. 5). PAX-7 exhibited in the same assay a 3-fold induction of luciferase activity above basal level. The transcriptional stimulation was completely lost, if the PAX-7 deletion mutant (PAX-7 Δ) lacking the C-terminal sequences was used instead of the full-length cDNA. These experiments suggest that PAX-7 like Pax-3 can act as transcriptional activator *in vivo*. Furthermore, the transactivation domain resides within a region of 126 amino acids of the PAX-7 C-terminus.

PAX-7 is expressed in human myogenic cells

In situ hybridization experiments have previously revealed expression of the mouse Pax-7 gene in the ventricular zone of the neural tube and the dermomyotome of the somites during embryogenesis (15). Since we have isolated the PAX-7 cDNA from primary human myoblasts, and since expression of PAX-3 has previously been reported in human rhabdomyosarcoma cells, we compared the expression pattern of PAX-7 during normal human myogenesis and in tumorigenic rhabdomyosarcoma cells with that of PAX-3. To this end, we performed RNase protection experiments using N-terminal sequences of PAX-3 and PAX-7 as probes. Primary human myoblast cells, derived from satellite cells, were cultivated in media containing high levels of mitogens (GM). To induce muscle differentiation, the cells were switched

to low mitogen medium (FM) for different periods. As shown in Figure 6A, PAX-7 specific message could be found in primary human myoblasts at constant low levels throughout the differentiation period examined. In contrast to PAX-7, we found no expression of PAX-3 in these primary neonatal myoblasts. The expression pattern seen in normal cells was then compared to two human rhabdomyosarcoma cell lines. Transcripts for both PAX-3 and PAX-7 were found at high levels in A673 and RD cells. These results were corroborated by RT-PCR assays which showed essentially the same expression pattern (data not shown). We therefore conclude that the rhabdomyosarcoma cell lines tested express both PAX-3 and PAX-7 at elevated levels, whereas primary myoblasts show only low levels of PAX-7 expression and no expression of PAX-3.

The expression of Pax-7 in the somites during mouse embryogenesis as well as the fact that Pax genes can act as transcriptional regulators make it plausible that they are involved in the network regulating myogenesis. We have therefore compared the expression pattern of PAX-3 and PAX-7 to the expression of the myogenic bHLH proteins (myf3–myf6). Primary myoblasts were found to contain myf3 (myoD) mRNA at similar levels throughout differentiation, myf4 (myogenin) levels increased upon differentiation, whereas myf5 levels were dramatically reduced (Fig. 6B). No myf6 transcripts could be detected in any of the myoblast cultures analyzed. The rhabdomyosarcoma cell line RD expressed myf3 and myf4, whereas no transcripts of the myogenic bHLH genes were found in A673 cells. To determine the size of the human PAX-7 mRNA, we performed Northern blot experiments with the two rhabdomyosarcoma cell lines expressing high levels of PAX-7 mRNA. A transcript of 7.2 kb length could be detected in both A673 and RD cells (data not shown), indicating that the human PAX-7 mRNA is larger than the 4.9kb mouse Pax-7 transcript (15).

DISCUSSION

Pax genes are important regulators of developmental processes. Here, we isolated a cDNA coding for the human PAX-7 protein. Characterization of this protein demonstrated that it can act as

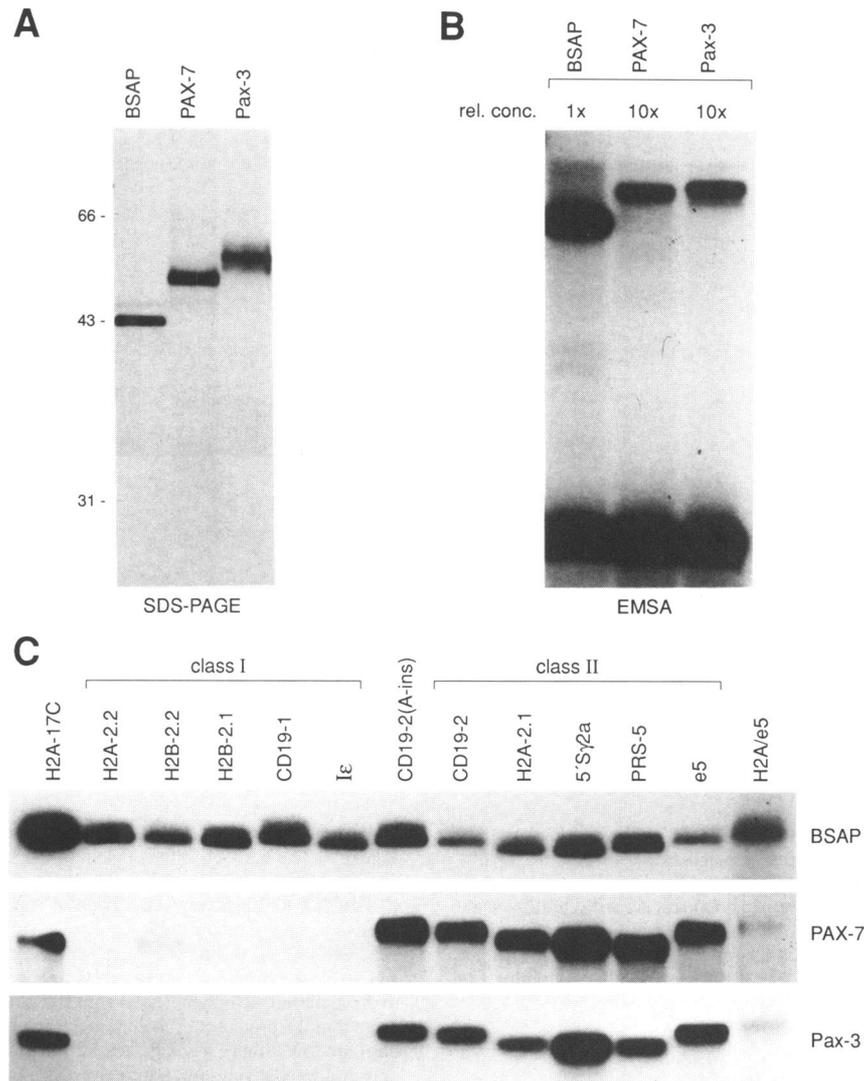


Figure 3. DNA binding of the PAX-7 paired domain. (A) BSAP (PAX-5), PAX-7 and Pax-3 were synthesized *in vitro* and analyzed by SDS-PAGE followed by autoradiography. The positions of the marker proteins are indicated at the left. (B) The concentrations of the different *in vitro* translated peptides were adjusted according to PhosphoImager quantification and binding to the CD19-2(A-ins) oligonucleotide was analyzed by EMSA. (C) Binding of different Pax proteins to a panel of BSAP recognition sequences. The Pax proteins were expressed by transient transfection in COP-8 cells, and whole cell extracts were analyzed by EMSA with the labelled oligonucleotides corresponding to known BSAP recognition sequences (36). The definition of class I and class II binding sites has been described by Czerny *et al.* (36).

a transcription factor through binding to specific DNA sequences. PAX-3 and PAX-7, which are closely related to each other and constitute a subclass of Pax proteins (44), both recognize DNA in a similar, if not identical, manner which is clearly different from PAX-5 (BSAP) representing a different subclass. Furthermore, Pax-3 and PAX-7 show transcriptional activity with the same reporter construct in transient assays. This might have important consequences for the regulation of target genes *in vivo*. Interestingly, Pax-3 and Pax-7 have a partially overlapping expression pattern during both mouse and chicken development. Both genes are expressed in the ventricular zone of the neural tube and in the dermomyotome of the developing somite. However, the presumptive precursor myoblasts migrating from the somites into the limb buds appear to express only Pax-3 and

not Pax-7 (17,19). Interestingly, these cells are absent in homozygous *Splotch* embryos lacking a functional Pax-3 protein. As a result, limb muscles in these mice are lost (17,19,45). It would be interesting to see, whether ectopic expression of Pax-7 in the migratory cells of a *Splotch* mouse could rescue the defective phenotype.

Analysis of the sequence specificity of Pax-3 and PAX-7 revealed a strict preference of the paired domain for class II binding sites, which contain only the 3' half site of the consensus recognition sequence for BSAP (PAX-5). We showed previously that this interaction critically depends on the N-terminal subregion of the paired domain, but not on C-terminal sequences (36). As the C-terminal sequences in the paired domains of Pax proteins are highly conserved between *Drosophila* and vertebrates, it is

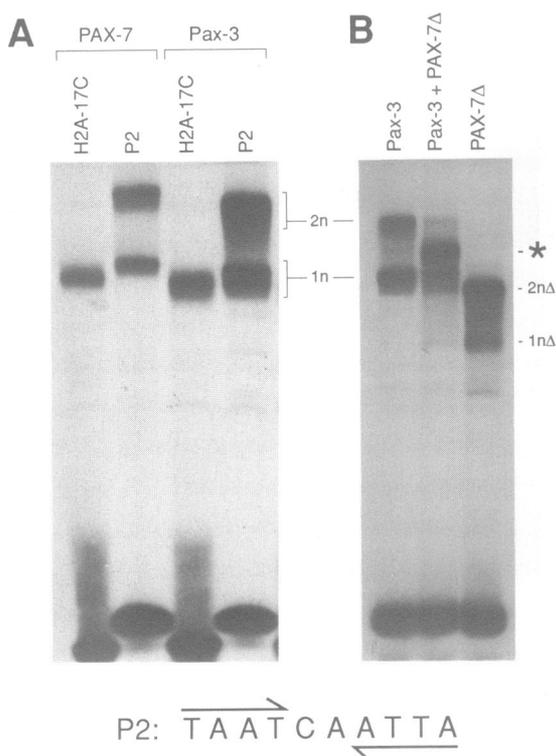


Figure 4. Cooperative dimerization of Pax-3 and PAX-7. The indicated proteins were overexpressed in COP-8 cells, and whole cell extracts were analyzed by EMSA with the H2A-17C and P2 oligonucleotides (A) and only the P2 oligonucleotide (B). Monomeric and dimeric complexes are indicated as 1n and 2n, respectively. An asterisk marks the heterodimeric complex formed between PAX-7Δ (amino acids 1–341) and full-length Pax-3. The sequence of the palindromic homeodomain-binding site P2 (37) is shown in the bottom part of the figure.

likely that this region of Pax-3 and PAX-7 also functions in DNA sequence recognition, albeit with a different sequence specificity than PAX-5. This would imply the existence of a different set of class I binding sites for this subfamily. The PAX-7 cDNA, which we isolated from myoblasts, contains a small insertion within the C-terminal region of the paired domain that has not been observed so far in any other Pax protein. This insertion of two amino acids results from the preferential use of an alternative 3' splice site in myoblasts. Similar alternative splice products giving rise to small in-frame insertions within the paired domain have been observed for Pax-6 and Pax-8 (46,47). In both cases these insertions result in a dramatic reduction of the DNA-binding activity of the paired domain (Kozmik *et al.*, in preparation). In contrast, PAX-7 containing the two amino acid insertion does not exhibit any differences in DNA binding, when compared to Pax-3 lacking this insertion. As argued above, the PAX-7 binding sites identified in our analysis are not recognized by the C-terminal half of the PAX-7 paired domain. The two amino acid insertion in the PAX-7 paired domain may, however, have severe effects on the recognition of class I binding sites, if they exist at all for PAX-7.

Our DNA-binding analysis have furthermore demonstrated that PAX-7 and Pax-3 are able to cooperatively dimerize on

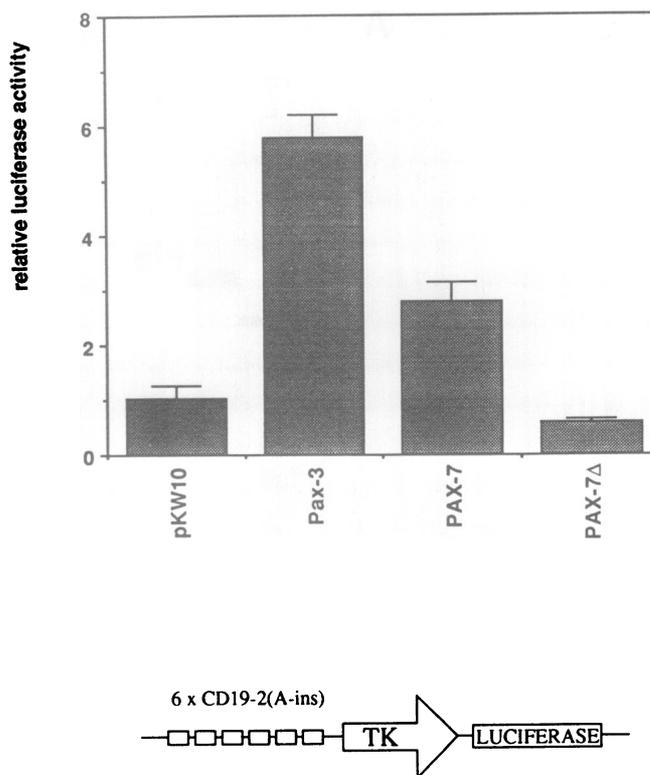


Figure 5. The transactivation domain of PAX-7 resides in the C-terminus. The reporter gene lucTKCD19 (3 μg), the indicated expression plasmids (1 μg) and CMV-CAT (0.2 μg) were cotransfected into the plasma cell line J558L. The relative luciferase activity was standardized to the CAT activity of the reference gene CMV-CAT. All transfection experiments were repeated at least five times and the results of separate experiments were combined by normalizing the values to the luciferase activity obtained with Pax-3. Note that the CD19-2(A-ins) sites present in the promoter of the reporter gene lucTKCD19 (shown in the bottom part) are low affinity binding sites for Pax-3 and PAX-7 (Fig. 3A) which may explain the relatively low transactivation by these proteins.

palindromic homeodomain binding sites as previously shown for the *Drosophila* Paired protein (37). The potential for this interaction is therefore highly conserved in evolution, although other homeodomain-containing Pax proteins such as Pax-6 do not show this behavior (Czerny *et al.*, in preparation). Furthermore, we provide the first evidence that Pax-3 and PAX-7 can form heterodimers on these sites. In particular, heterodimer formation between these two proteins appears to be favoured over homodimer formation (Fig. 4B). This finding has important implications for target gene activation by Pax-3 and Pax-7 in cells expressing both proteins (see above).

We have demonstrated that PAX-3 and PAX-7 are expressed in two rhabdomyosarcoma cell lines RD and A673 at higher levels than in normal myoblasts. Interestingly, the cell line A673 lacks transcripts coding for the myogenic bHLH proteins. Previously, expression of myf3 (myoD) has been suggested as a marker for rhabdomyosarcoma cells (48). The expression of PAX-3 and PAX-7 in A673 cells in the absence of myf3 expression suggests, that these Pax genes might be useful additional markers for tumors of small round cell histology and may be able to identify less

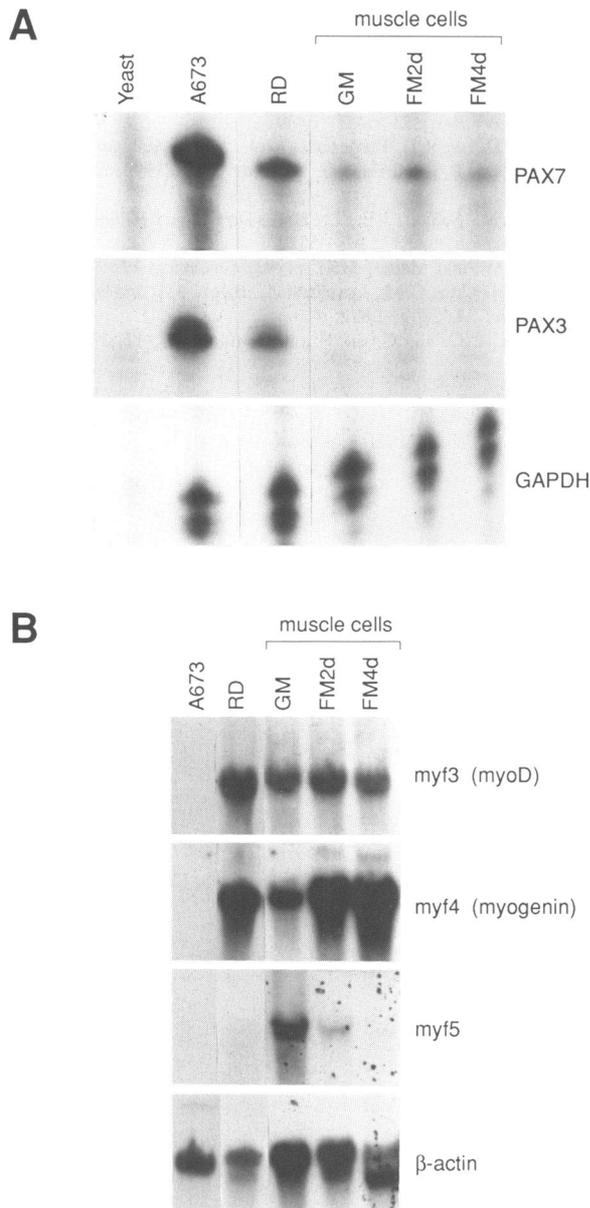


Figure 6. Expression of PAX-7 and PAX-3 genes in human myogenic cells in comparison to the expression of the myogenic bHLH transcripts. (A) RNase protection analysis of PAX-7, PAX-3 and glyceraldehyde phosphate dehydrogenase (GAPDH) mRNA were performed with 0.5 μ g poly(A)⁺ RNA isolated from the human myogenic cell lines indicated. RNA probes are described in detail in Materials and Methods. The autoradiograph of the GAPDH experiment was exposed for only 4 h, compared to 2 days for those of PAX-7 and 4 days for PAX-3. mb-GM are primary human myoblasts cultured in growth medium; mb-FM2d and mb-FM4d are the same cells grown in fusion media for 2 and 4 days, respectively. (B) Northern blot analysis of myogenic basic helix-loop-helix transcripts with total RNA (10 μ g) of the indicated cells. In order to avoid cross-hybridization, only probes containing sequences outside the common helix-loop-helix region were used (see Materials and Methods). No signals could be detected in any of the cells with the myf6 probe. The β -actin probe used is also detecting α -actin mRNA which serves as a marker for differentiation. All autoradiographs were exposed for 18 h, except for that of β -actin which was exposed for 4 h.

differentiated cells of myogenic origin. We have screened three additional rhabdomyosarcoma cell lines (Rh1, Rh18 and Rh30), which all have been reported to express myf3 (49), and found

high expression levels of PAX-3 and/or PAX-7, thus supporting the notion that these Pax genes are expressed in both embryonal (RD, Rh1) and alveolar (Rh18, Rh30) rhabdomyosarcomas (data not shown).

A specific chromosome translocation, t(2;13), resulting in the fusion of PAX-3 to a fork head domain gene (FKHR) has been implicated in the genesis of alveolar rhabdomyosarcoma (32,50,51). Recently, the PAX-7 gene has been localized to human chromosome 1p36 (21,52), and subsequently two cases of alveolar rhabdomyosarcoma have been identified with a translocation t(1;13)(p36;q14) resulting in the generation of a similar fusion protein between PAX-7 and FKHR (53). However, in these experiments the expression levels of the fusion transcripts have not been compared to those of PAX-3 and PAX-7 in normal myoblasts. Here we show that PAX-3 and/or PAX-7 are expressed at higher levels in rhabdomyosarcoma cells. Interestingly, ectopic expression of Pax-3 itself in mouse or rat fibroblasts can lead to transformation (54). It is therefore tempting to speculate, that the overexpression of the PAX-3 or PAX-7 protein *per se* could be implicated in the genesis of rhabdomyosarcomas. In the case of embryonal rhabdomyosarcoma alterations in the transcriptional machinery could lead to overexpression of the PAX genes, whereas in alveolar rhabdomyosarcoma the generation of a fusion protein could lead to the deregulated expression of PAX-3 and PAX-7, for example through removal of potential negative regulatory elements or the insertion of an additional enhancer originating from the fork head domain (FKHR) gene. We are currently testing this model by overexpressing Pax-7 and Pax-3 in myoblasts to determine, if overexpression leads to transformation and a concomitant block of differentiation, as has been shown for other oncogenes like ras, jun or src (42). Alternatively, the elevated expression levels of PAX3/7 found in rhabdomyosarcoma cells might reflect a partial reversion to an earlier embryonal phenotype in these tumors. The isolation of a PAX-7 cDNA will now enable use to directly investigate the role of this transcription factor in normal and malignant myogenesis.

ACKNOWLEDGEMENTS

We are grateful to Z.Kozmik for providing the PAX-3 and PAX-7 RNase protection probes and to G.Christofori for critical reading to the manuscript. This work was, in part, supported by a grant from the Austrian Industrial Research Promotion Fund and by grant 3100-37378.93 from the Swiss National Science Foundation (to B.W.S.).

REFERENCES

- Weintraub, H., Davis, R., Tapscott, S., Thayer, M., Krause, M., Benzeira, R., Blackwell, T.K., Turner, D., Rupp, R., Hollenberg, S. and *et al* (1991) *Science*, **251**, 761–766.
- Olson, E.N. and Klein, W.H. (1994) *Genes Dev.*, **8**, 1–8.
- Rudnicki, M.A., Schnegelsberg, P.N.J., Stead, R.H., Braun, T., Arnold, H.H. and Jaenisch, R. (1993) *Cell*, **75**, 1351–1359.
- Nabeshima, Y., Hanaoka, K., Hayasaka, M., Esumi, E., Li, S., Nonaka, I. and Nabeshima, Y. (1993) *Nature*, **364**, 532–535.
- Hastie, P., Bradley, A., Morris, J.H., Edmondson, D.G., Venuti, J.M., Olson, E.N. and Klein, W.H. (1993) *Nature*, **364**, 501–506.
- Weintraub, H. (1993) *Cell*, **75**, 1241–1244.
- Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B. and Miller, A.D. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5434–5438.

8. Schäfer, B.W., Blakely, B.T., Darlington, G.J. and Blau, H.M. (1990) *Nature*, **344**, 454–458.
9. Sassoon, D., Lyons, G., Wright, W.E., Lin, V., Lassar, A., Weintraub, H. and Buckingham, M. (1989) *Nature*, **341**, 303–307.
10. Tajbakhsh, S. and Buckingham, M.E. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 747–751.
11. Yu, Y.T., Breitbart, R.E., Smoot, L.B., Lee, Y.S., Mahdavi, V. and Nadalginard, B. (1992) *Genes Dev.*, **6**, 1783–1798.
12. Pollock, R. and Treisman, R. (1991) *Genes Dev.*, **5**, 2327–2341.
13. Cserjesi, P., Lilly, B., Bryson, L., Wang, Y.Q., Sassoon, D.A. and Olson, E.N. (1992) *Development*, **115**, 1087–1101.
14. Kuratani, S., Martin, J.F., Wawersik, S., Lilly, B., Eichele, G. and Olson, E.N. (1994) *Dev. Biol.*, **161**, 357–369.
15. Jostes, B., Walther, C. and Gruss, P. (1991) *Mech. Dev.*, **33**, 27–38.
16. Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R. and Gruss, P. (1991) *EMBO J.*, **10**, 1135–1147.
17. Bober, E., Franz, T., Arnold, H.H., Gruss, P. and Tremblay, P. (1994) *Development*, **120**, 603–312.
18. Williams, B.A. and Ordahl, C.P. (1994) *Development*, **120**, 785–796.
19. Goulding, M., Lumsden, A. and Paquette, A.J. (1994) *Development*, **120**, 957–971.
20. Gruss, P. and Walther, C. (1992) *Cell*, **69**, 719–722.
21. Stapleton, P., Weith, A., Urbanek, P., Kozmik, Z. and Busslinger, M. (1993) *Nature Genetics*, **3**, 292–298.
22. Walther, C., Guenet, J.L., Simon, D., Deutsch, U., Jostes, B., Goulding, M.D., Plachov, D., Balling, R. and Gruss, P. (1991) *Genomics*, **11**, 424–434.
23. Strachan, T. and Read, A.P. (1994) *Curr. Op. Genet. Dev.*, **4**, 427–438.
24. Balling, R., Deutsch, U. and Gruss, P. (1988) *Cell*, **55**, 531–5.
25. Epstein, D.J., Vekemans, M. and Gros, P. (1991) *Cell*, **67**, 767–774.
26. Hill, R.E., Favor, J., Hogan, B.L.M., Ton, C.C.T., Saunders, G.F., Hanson, I.M., Prosser, J., Jordan, T., Hastie, N.D. and van Heyningen, V. (1991) *Nature*, **354**, 522–525.
27. Baldwin, C.T., Hoth, C.F., Amos, J.A., da-Silva, E.O. and Milunsky, A. (1992) *Nature*, **355**, 637–638.
28. Morell, R., Friedman, T.B. and Asher, J.H. (1993) *Human Mol. Genet.*, **2**, 1487–1488.
29. Tassabehji, M., Read, A.P., Newton, V.E., Patton, M., Gruss, P., Harris, R. and Strachan, T. (1993) *Nature Genetics*, **3**, 26–30.
30. Glaser, T., Walton, D.S. and Maas, R.L. (1992) *Nature Genetics*, **2**, 232–239.
31. Ton, C.C.T., Hirvonen, H., Miwa, H., Weil, M.M., Monaghan, P., Jordan, T., van Heyningen, V., Hastie, N.D., Meijers-Heijboer, H., Drechsler, M., Royer-Pokora, B., Collins, F., Swaroop, A., Strong, L.C. and Saunders, G.F. (1991) *Cell*, **67**, 1059–1074.
32. Barr, F.G., Galili, N., Holick, J., Biegel, J.A., Rovera, G. and Emanuel, B.S. (1993) *Nature Genetics*, **3**, 113–117.
33. Blau, H.M. and Webster, C. (1981) *Proc. Natl Acad. Sci. USA*, **78**, 5623–5627.
34. Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) *Biochemistry*, **18**, 5294–5299.
35. Frohman, M.A. (1990) *Amplifications*, **5**, 11–15.
36. Czerny, T., Schaffner, G. and Busslinger, M. (1993) *Genes Dev.*, **7**, 2048–2061.
37. Wilson, D., Sheng, G., Lecuit, T., Dostatni, N. and Desplan, C. (1993) *Genes Dev.*, **7**, 2120–2134.
38. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed. Cold Spring Harbor University Press, Cold Spring Harbor.
39. Adams, B., Dörfler, P., Aguzzi, A., Kozmik, Z., Urbánek, P., Maurer-Fogy, I. and Busslinger, M. (1992) *Genes Dev.*, **6**, 1589–1607.
40. Dohrman, C., Azpiazu, N. and Frasch, M. (1990) *Genes Dev.*, **4**, 2098–2111.
41. Wey, E., Lyons, G.E. and Schäfer, B.W. (1994) *Eur. J. Biochem.*, **220**, 753–762.
42. Kozak, M. (1991) *J. Cell Biol.*, **115**, 887–903.
43. Burri, M., Tromvoukis, Y., Bopp, D., Frigerio, G. and Noll, M. (1989) *EMBO J.*, **8**, 1183–1190.
44. Deutsch, U. and Gruss, P. (1991) *Semin. Dev. Biol.*, **2**, 413–424.
45. Franz, T., Kothary, R., Surani, M., Halata, Z. and Grim, M. (1993) *Anatomy and Embryology*, **187**, 153–160.
46. Walther, C. and Gruss, P. (1991) *Development*, **113**, 1435–1449.
47. Kozmik, Z., Kurzbauer, R., Dörfler, P. and Busslinger, M. (1993) *Mol. Cell. Biol.*, **13**, 6024–6035.
48. Scrabble, H., Witte, D., Shimada, H., Seemayer, T., Sheng, W.W., Soukup, S., Koufos, A., Houghton, P., Lampkin, B. and Cavenee, W. (1989) *Genes Chromosom Cancer*, **1**, 23–35.
49. Dias, P., Parham, D.M., Shapiro, D.N., Tapscott, S.J. and Houghton, P.J. (1992) *Cancer Res.*, **52**, 6431–6439.
50. Galili, N., Davis, R.J., Fredericks, W.J., Mukhopadhyay, S., Rauscher, F.J., Emanuel, B.S., Rovera, G. and Barr, F.G. (1993) *Nature Genetics*, **5**, 230–235.
51. Shapiro, D.N., Sublett, J.E., Li, B.T., Downing, J.R. and Naeve, C.W. (1993) *Cancer Res.*, **53**, 5108–5112.
52. Schäfer, B.W. and Mattei, M.G. (1993) *Genomics*, **17**, 249–251.
53. Davis, R.J., D'Cruz, C.M., Lovell, M.A., Biegel, J.A. and Barr, F.G. (1994) *Cancer Res.*, **54**, 2869–2872.
54. Maulbecker, C.C. and Gruss, P. (1993) *EMBO J.*, **12**, 2361–2367.