THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 272, No. 50, Issue of December 12, pp. 31777–31784, 1997

Printed in U.S.A.

A Transcription Activator with Restricted Tissue Distribution Regulates Cell-specific Expression of α1(XI) Collagen*

(Received for publication, August 5, 1997)

Akihiro Kinoshita‡, Patricia Greenwel‡, Shizuko Tanaka‡, Maurizio Di Liberto§, Hidekatus Yoshioka¶, and Francesco Ramirez‡‖

From the ‡Brookdale Center for Developmental and Molecular Biology, Mt. Sinai School of Medicine, New York, New York 10029, the §Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, and the ¶Department of Molecular Biology and Biochemistry, Okayama University Medical School, Okayama 565, Japan

Different regulatory programs are likely to control expression of the α1(XI) collagen (*COL11A1*) gene in cartilaginous and non-cartilaginous tissues and in coordination with different collagen genes. Here, we report the identification of a *cis*-acting element that is required for constitutive and tissue-specific activity of the proximal *COL11A1* promoter. The element binds an apparently novel activator whose expression is restricted mostly, but not exclusively, to cells of mesenchymal origin. Transient transfection experiments using wild-type and mutant constructs demonstrated the critical contribution of a 45-base pair upstream element (FP9) to promoter activity. The same functional tests and DNA binding assays narrowed down the critical portion of FP9 to a 20-base pair sequence, which consists of an imperfect palindrome with strong homology to the GATA consensus motif. Despite being able to bind GATA proteins *in vitro*, FP9 is actually recognized by a distinct ~100-kDa polypeptide (FP9C) probably belonging to the zinc-finger family of transcription factors. FP9C binding was mostly identified in nuclei from cells of mesenchymal origin, including those actively engaged in *COL11A1* transcription. A positive correlation was also established between the level of FP9C binding and the degree of cell differentiation *in vitro*. Thus, FP9C represents an unusual example of tissue-specific and differentiation-related transcription factor with overlapping expression in hard and soft connective tissues.

Proper expression of collagen genes during embryogenesis and in the adult organism are required for the correct assembly and the physiological maintenance of the extracellular matrix (ECM)[1] (1). Conversely, deregulated production of collagen molecules is the hallmark of a variety of connective tissue disorders (2). During the past few years, substantial effort has been directed toward understanding the transcription of collagen genes, including the identification of factors that confer spatiotemporal specificity. By and large, most of the progress has been made with the genes coding for the subunits of types I and II collagen. This body of work has identified *cis*-acting elements and *trans*-acting factors that restrict expression of these collagens to distinct mesenchymal cell lineages. Several reports have demonstrated that transcription of the α1(I) collagen gene in osteoblasts is under the control of an upstream element that contains multiple binding sites for nuclear proteins (4–7). This modularly arranged *cis*-acting element is recognized by both ubiquitous and osteoblast-specific factors (6, 7). There is evidence suggesting that the latter includes an apparently new member of the homeodomain proteins (7). Similarly, an 18-bp sequence within the first intron of the α1(II) collagen gene has been shown to control promoter expression in cartilaginous tissues by binding a transcriptional complex that includes the SOX9 activator (8, 9). Finally, expression of the proximal promoter of α2(I) collagen in dermis, tendon, and the fibrous layers of many internal organs has been reported to be significantly augmented by a far-upstream enhancer that contains fibroblast-specific DNase-hypersensitive sites (10).

We are interested in the control of the genes coding for the subunits of the so-called minor fibrillar collagens (11). The two members of this subgroup of collagens, types V and XI, play a critical role in regulating the formation of the types I and II fibrillar networks in cartilaginous and non-cartilaginous tissues, respectively (12, 13). They can also give rise to cross-type trimers consisting of α2(V) and α1(XI) chains in vascular muscles and in the developing bone, among other tissues (11). Although the precise function of this cross-type molecule remains a mystery, its tissue distribution implies that the α2(V) and α1(XI) collagen (*COL5A2* and *COL11A1*) genes are more broadly expressed than those coding for the subunits of the parental trimers, as well as the co-expressed types I and II collagen genes. This in turn endows *COL11A1* with the distinction of being the sole collagen gene to be transcribed in both cartilaginous and non-cartilaginous tissues. We have previously reported the characterization of the proximal upstream sequence of *COL11A1* (14). DNA sequencing has revealed that the promoter contains GC-rich boxes in place of the TATA motif. DNA transfections have established that *cis*-acting elements located between nucleotides −541 and −199 drive constitutive transcription from the proximal promoter in α1(XI) collagen-producing cells. The latter tests have employed the rhabdomyosarcoma A-204 line and vascular smooth muscle cells (smc), as the *COL11A1*-positive cells, and the fibrosarcoma HT-1080 line, as the *COL11A1*-negative cell. Finally, DNase I footprinting assays have mapped nine areas of interaction with nuclear proteins (FP1 to FP9) within the proximal

* This work was supported by Grant R37-38648 from the NIAMS, National Institutes of Health. This is article 244 from the Brookdale Center for Developmental and Molecular Biology. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact. .

‖ To whom correspondence should be addressed: Brookdale Center for Developmental and Molecular Biology, Mount Sinai School of Medicine, One Gustave Levy Place, New York, NY 10029. Tel.: 212-241-7984; E-mail: ramirez@msvax.mssm.edu.

[1] The abbreviations used are: ECM, extracellular matrix; bp, base pair(s); CAT, chloramphenicol acetyltransferase; *COL5A2*, the gene coding for the α2 chain of type V collagen; *COL11A1*, the gene coding for the α1 chain of type XI collagen; EMSA, electrophoretic mobility shift assay; HUVEC, human umbilical vein endothelial cells; MELC, mouse erythroleukemia cells; OPA, 1,10-phenanthroline; smc, smooth muscle cell(s).

promoter (Fig. 1). Based on sequence homologies, most of them were thought to represent binding sites for ubiquitous activators, like AP2 and Sp1 (14). An interesting correlation was also noted with the proximal promoter of the coordinately expressed *COL5A2* gene (15). The two collagens were in fact found to share an upstream element, called FPB in *COL5A2* and FP7 in *COL11A1*, that binds the same ubiquitously expressed *trans*-acting factor (14).

The present study was designed to further characterize the *cis*-acting elements of the −541 to −199 segment and implicitly, the *trans*-acting factors involved in *COL11A1* regulation. The results indicate that constitutive and cell type-specific transcription of the −541*COL11A1* promoter is under the control of FP9, an element farther upstream of FP7. More importantly, the experimental evidence suggests that the transcription factor binding to FP9 is an apparently new member of the zinc-finger family of nuclear proteins. The factor (FP9C) is mostly expressed in cells of mesenchymal origin and always in those actively transcribing *COL11A1*. Interestingly, we also found a positive correlation between the level of FP9C binding and the stage of differentiation of osteoblastic and skeletal muscle cell lines.

MATERIALS AND METHODS

Cells and Other Reagents—Human cell lines included primary embryonic dermal fibroblasts (CF 37), adult keratinocytes, neonatal melanocytes, umbilical vein endothelial cells (HUVEC), and the rhabdomyosarcoma A-204 and T-lymphocyte Jurkat lines. Keratinocytes, melanocytes and HUVEC were purchased from Cascade Biologics, Inc. (Portland, OR). Rat cells included primary vascular smooth muscle cells (smc) and hepatocytes, the neurogenic PC12 line, the osteosarcoma lines ROS 17/2.8 and ROS 25, and the chondrosarcoma line RCS. Mouse cells included primary costal chondrocytes, NIH3T3 fibroblasts, the multipotential mesodermal line C3H10T1/2, C2C12 myoblasts, the C3H10T1/2-derived P2 myogenic line, MC3T3-E1 osteoblast-like calvarial cells, and erythroleukemia cells (MELC). Consensus recognition sequences for and antibodies against various transcription factors were purchased from Santa Cruz Biotechnology (Santa Cruz, CA).

Cell Cultures, RNA Analysis, and DNA Transfection Experiments—Most cells were grown and maintained in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum. MC3T3-E1 and Jurkat cells were cultured in 10% fetal calf serum supplemented with α-minimal essential medium and RPMI medium, respectively; melanocytes and keratinocytes were cultured in Medium 154 supplemented with HMGS and HKGS, respectively; HUVEC were grown in Medium 200 supplemented with LSGS. C2C12 myoblasts and the myogenic P2 line were induced to differentiate by placing confluent cells in 2% horse serum (16). Short term cultures of mouse costal chondrocytes were prepared as described (17). RNA was purified from cultured cells and subjected to Northern hybridization to probes for *COL11A1* and glyceraldehyde-3-phosphate dehydrogenase according to standard protocols (18). Progressive 5′ deletion of the *COL11A1* promoter and internal substitutions of the 45-bp and 20-bp FP9 with unrelated sequences were engineered using specific oligonucleotide primers and the polymerase chain reaction technique and were all verified by DNA sequencing (18). The mutant sequences were subcloned 5′ of the chloramphenicol acetyltransferase (CAT) reporter gene in the vector pBLCAT3 (19).

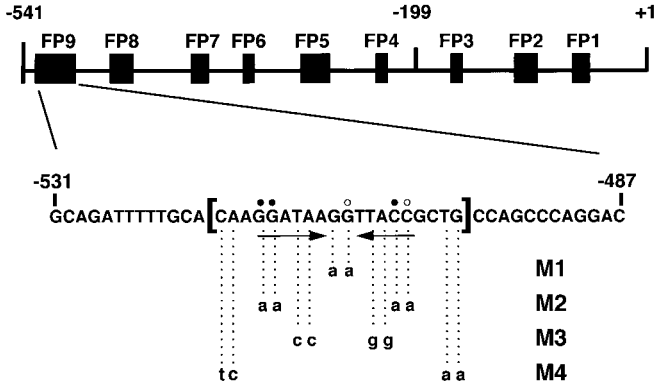

FIG. 1. ***COL11A1* proximal promoter.** *Top*, schematic representation of the promoter showing the relative positions of the footprinted areas including those six that lay between nucleotides −541 and −199 (14). The nucleotide composition of the 45-bp long FP9 is shown below with the 20-bp core highlighted by the *brackets*; the inverted GATA-like repeats are *underlined*; and the major and minor contact points of the nuclear factors are indicated by *closed* and *open circles*, respectively (see Fig. 3). Below the sequence are the nucleotide substitutions in the 20-bp core used in the competition assays shown in Fig. 3*A*.

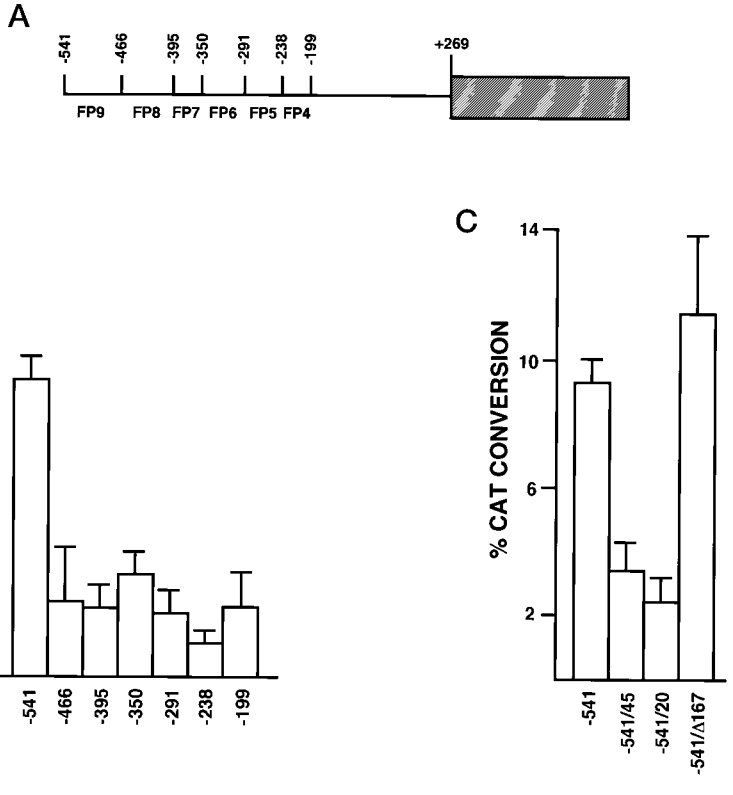

FIG. 2. **Functional analysis of the 342-bp promoter segment in A-204 cells.** *A*, schematic map of the proximal promoter, which includes the nucleotide position of the 5′ ends of the deletion mutants, the location of the footprinted areas, and the CAT reporter gene represented by the *gray rectangle*. *B* and *C* show the functional assays of the 5′ deletion mutants (−541 to −199) and of the constructs with nucleotide substitutions (−541/45 and −541/20) or internally deleted (−541/Δ167). Histograms indicate the percentage of CAT activity normalized for the co-transfected control and expressed relatively to pSVCAT activity. The data represent an average of three to five independent tests ± S.D.

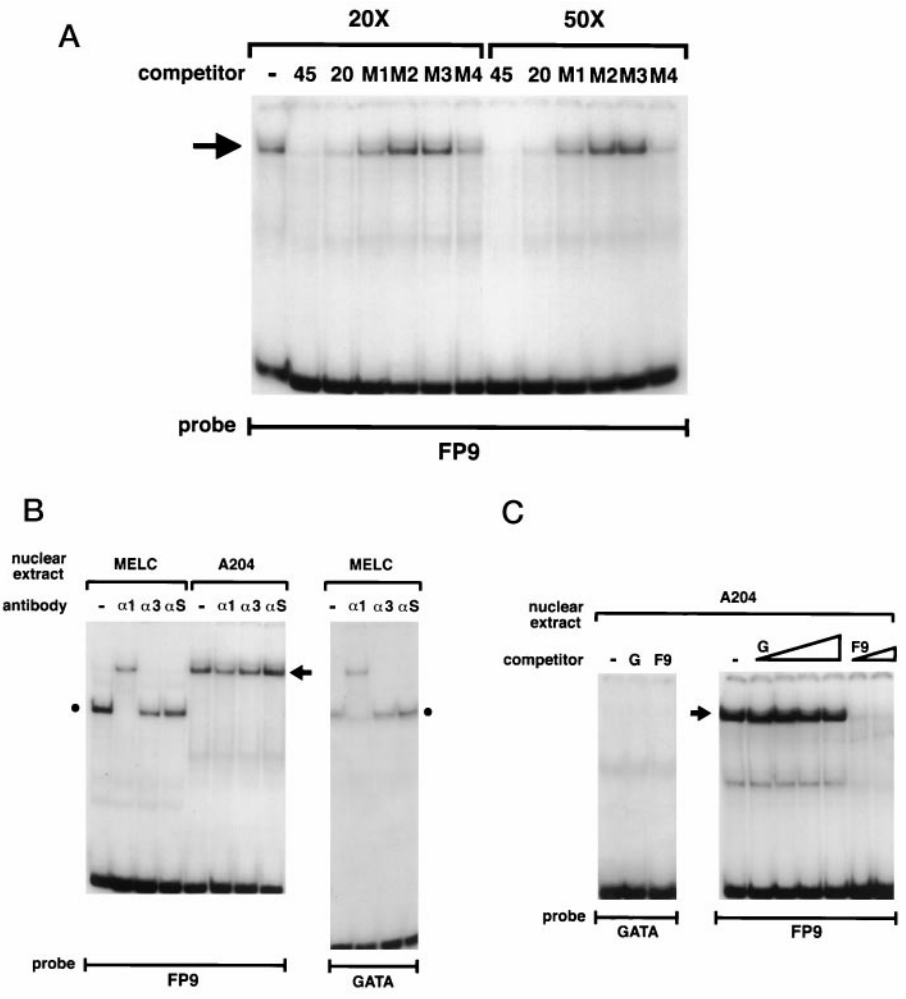

FIG. 3. **Binding specificity of A-204 nuclear proteins to FP9.** *A*, EMSA was performed with the 45-bp probe and binding was competed with 20- and 50-fold molar excess of the same unlabeled probe (*45*) or the wild-type (*20*) and mutant versions (*M1–M4*) of the 20-bp core. The composition of the 45-bp and 20-bp oligonucleotides are shown in Fig. 1. *B*, EMSAs were performed using the indicated probes and nuclear extracts without (−) and with preincubation with antibodies against GATA-1 (*α1*), GATA-3 (*α3*), or Sp1 (*αS*). *C*, EMSAs were performed using the indicated probes and A-204 nuclear extract without (−) and with unlabeled competitors corresponding to the GATA (*G*) and FP9 (*F9*) sequences. In the EMSA on the *left*, competitors were added at 100-fold excess; in the EMSA on the *right*, the GATA competitor was added at 20-, 50-, 100-, and 300-fold excess and the FP9 competitor at 20- and 50-fold excess. In all panels, the *arrow* and the *closed circle* identify FP9C and GATA-1, respectively.

Purified plasmid DNA was introduced into the cells by the calcium phosphate method according to the published protocol along with the normalizing standard pSVLUC, a plasmid containing the luciferase reporter gene under the transcriptional control of the SV40 promoter (20). About 48 h after transfection, cells were harvested and assayed for CAT and luciferase activities. Values were normalized for the co-transfected plasmid and expressed relatively to parallel transfections of the pSVCAT vector. Cell transfections were performed multiple times and the resulting data were evaluated using Student's *t* test.

In Vitro DNA Binding Assays—Crude nuclear extracts were purified according to our published protocol (15). Unless otherwise specified, 5 μg of nuclear proteins were used in the electrophoretic mobility shift assay (EMSA) together with approximately 10,000 dpm of labeled oligonucleotide probe as described previously (15). Unlabeled competitors were added in the amounts indicated in the figure legends, or otherwise at 100-fold molar excess. In the zinc-dependence assay, 0.5 μl of 1,10-phenanthroline (OPA) resuspended in ethanol at the concentration of 1 M, 0.2 M, or 0.1 M was added to 47.5 μl of buffer solution containing nuclear proteins and dI-dC and was then incubated on ice for 5.5 h. This was followed by addition of the labeled probe and incubation on ice for an additional 30′ prior to EMSA. Conditions for DNase I footprinting and DNA methylation interference assays were essentially identical to the published protocols (15).

Gel Filtration Chromatography—To estimate the molecular weight of denatured FP9C, nuclear extracts from A-204 cells (500 μg) were dissolved in 500 μl of denaturing buffer containing 6 M guanidine HCl, 50 mM Tris-HCl, pH 8.0, 0.1 mM dithiothreitol, 0.1 mM of phenylmethylsulfonyl fluoride and left on ice for 30 min. After clearing it by centrifugation, the solution was applied to a Sephacryl S-300 column (50 cm × 1 cm, inner diameter) equilibrated in the same denaturing buffer. Fractions (0.75 ml) were collected and analyzed by EMSA after microdialysis of 100-μl aliquots against 20% glycerol, 20 mM Hepes, pH 7.9, 0.1 mM dithiothreitol, 0.1 mM EDTA, and 0.1 mM phenylmethylsulfonyl fluoride. Marker proteins ranging from 14 to 100 kDa were fractionated through the same column under identical conditions and analyzed by SDS-polyacrylamide gel electrophoresis.

RESULTS

A previous study has delineated the shortest segment of the proximal *COL11A1* promoter capable of directing constitutive and cell type-specific expression of the reporter CAT gene in transient transfection experiments (14). The segment extends from nucleotide −541 to nucleotide −199 with respect to the major start site of transcription of this tissue-specific TATA-less gene. Within it, DNase I footprinting assays identified six distinct areas of protection from nuclease digestion by nuclear proteins, which were termed FP9 to FP4 (Fig. 1). In the present study, the same combination of DNA transfection and DNA-binding assays was used to further characterize the 342-bp upstream segment of *COL11A1*.

FP9 Is Required for −541COL11A1 Promoter Activity—To dissect the functional properties of the 342-bp segment, the activity of mutant constructs harboring progressive 5′ promoter deletions, from FP9 to FP4, were compared after transfection into A-204 cells. The analysis showed that all mutant plasmids drive expression of the CAT gene to levels comparable to the basal activity of the −199*COL11A1* promoter (Fig. 2*A*). The results therefore suggested that the 45-bp-long FP9 may be the principle contributor to transcription from the −541 promoter. This postulate was corroborated by the finding that the −541 promoter construct containing an unrelated sequence in place of the 45-bp FP9 (construct −541/45) displayed the same activity as the basal −199*COL11A1* plasmid (Fig. 2*C*). To

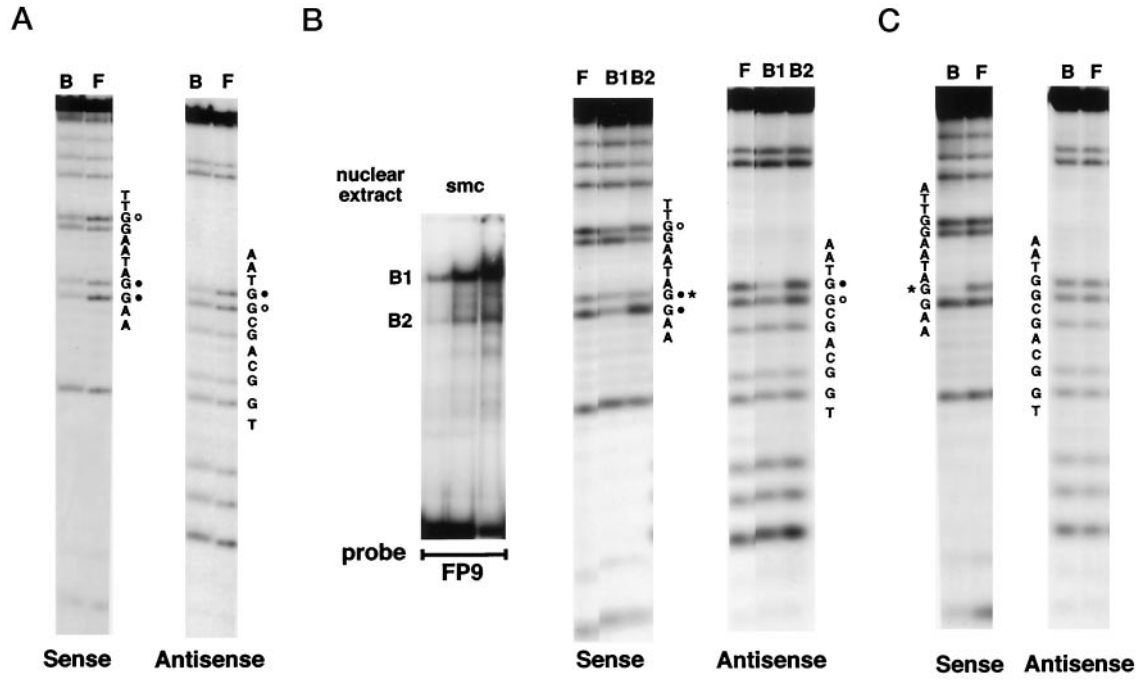

FIG. 4. **Methylation interference analyses.** *A*, methylation interference analysis of the single complex formed between the FP9 probe and the A-204 nuclear extract. Major and minor methylated G are indicated by *closed* and *open circles*, respectively. *B*, methylation interference analysis of the B1 and B2 complexes formed between the FP9 probe and smc nuclear extract (see EMSA on the *left*). The amount of smc nuclear extract used in the EMSA was 5, 10, and 20 μg. Contact points are indicated by the *open* and *closed circles* in B1 and by the *asterisk* in B2. The pattern of B1 should be compared with that in *A*, whereas the pattern of B2 should be compared with that in *C*. *C*, methylation inference analysis of the single complex formed between the GATA consensus sequence and the MELC. The contact point is indicated by the *asterisk*. In all panels, the results of the sense and antisense strand probes are shown along with the sequence around the contact points. *Lane F*, free DNA; *lane B*, bound DNA fragment.

establish whether other elements in the 342-bp segment might be required to support FP9 activity and indirectly to confirm the above results, another mutant construct was tested by transient transfection. The mutation was engineered using *Bst*XI restriction sites conveniently located at nucleotides −482 and −316 and resulted in the internal loss of the segment spanning from FP8 to FP6 (Fig. 2*A*). The mutant plasmid (construct −541/Δ167) exhibited slightly higher activity than the wild-type construct (Fig. 2*C*). Although a similar test was not performed for the sequence encompassing FP4 and FP5, the results strongly suggested that FP9 is the major contributor to constitutive and cell type-specific transcription from the proximal *COL11A1* promoter. Incidentally, wild-type and mutant constructs transfected into the HT-1080 line were transcriptionally inactive (data not shown).

The FP9 Binding Site Contains a GATA Palindrome—The EMSA was initially employed to assess the complexity of the FP9 binding pattern, as well as its specificity. To this end, nuclear proteins purified from A-204 cells were incubated with the radiolabeled 45-bp-long FP9 probe without and with increasing amount of the same unlabeled sequence. The results revealed the formation of a single predominant complex which migrates toward the top of the gel and is specifically competed by the same 45-bp-long sequence (Fig. 3*A*). Occasionally, we observed a lesser intense and faster moving complex (see Fig. 3*C*) whose identity was not investigated in the present study. Next, the methylation interference assay was performed to further characterize the *trans*-acting factor that binds to FP9. This delineated the main area of DNA-protein interaction by identifying the G residues that are contacted by the nuclear protein. Three major contact points and two minor ones were located within the 20-bp core of FP9 (Fig. 4*A*). This sequence is characterized by an imperfect palindrome with a 2-bp interruption and strong homology to the binding site for GATA proteins (Fig. 1) (21). The EMSA reiterated the importance of the 20-bp core by documenting the ability of this sequence to compete for binding to FP9 almost as effectively as the 45-bp long oligonucleotide (Fig. 3*A*). It also documented the specificity of the interaction, in that both the 45-bp and 20-bp oligonucleotides competed formation of the complex with comparable efficiency (Fig. 3*A*). Furthermore, it showed that the competitiveness of the core sequence was somewhat diminished by substitutions outside of the GATA motifs (*M1* and *M4*, Fig. 3*A*) and was completely abolished by those introduced within them (*M2* and *M3*, Fig. 3*A*). Altogether, the results of the *in vitro* binding tests implicated the imperfect GATA palindrome as the major recognition site of the FP9 complex (FP9C). Consistent with this conclusion, substitution of the 20-bp core of FP9 with an unrelated sequence rendered the −541 promoter construct (−541/20) as inactive as the plasmid without the entire FP9 or the construct with the mutant version of the 45 bp (Fis. 2, *B* and *C*).

FP9C Is a Novel Zinc-finger Protein—The GATA represents a family of six related zinc-finger transcription factors, which bind to the same (A/T)GATA(AG/GC) consensus sequence and are differently expressed in various tissues, including vascular muscles (21). To test whether or not the imperfect FP9 palindrome could be recognized *in vitro* by GATA proteins, binding of MELC nuclear proteins to FP9 and the GATA consensus sequence were compared. The erythroleukemia cells were chosen for the test because they produce large amounts of GATA-1, the prototype of the GATA family (22). The assay documented the ability of FP9 to bind *in vitro* the same protein of the MELC nuclear extract as the GATA consensus sequence (Fig. 3*B*). Aside from the similar migration, identity between the two complexes was supported by the finding that they were both supershifted by anti-GATA-1 antisera but not by anti-GATA-3 or anti-Sp1 antibodies (Fig. 3*B*). Based on the above results, we tested whether A-204 nuclei contain a GATA-like binding ac-

tivity and found that the GATA consensus sequence failed to yield a retarded band in the EMSA (Fig. 3*C*). To confirm the absence of GATA-like proteins in this α1(XI) collagen-producing tumor line, two additional EMSAs were performed. The first revealed that the A-204 nuclear factors that binds to the FP9 sequence is not recognized by the anti-GATA antibodies or by the control anti-Sp1 antisera (Fig. 3*B*). In the second EMSA, formation of the complex between A-204 nuclear proteins and the FP9 sequence was challenged with molar excess of the GATA consensus sequence or the 45-bp oligonucleotide. As expected, 20-fold excess of the latter abrogated FP9C binding; by contrast, competition with the GATA consensus sequence had no significant effect on FP9C formation even when the oligonucleotide was added in vast molar excess (Fig. 3*C*). Hence, a transcription factor other than GATA binds to the imperfect GATA palindrome of FP9 in the *COL11A1*-positive rhabdomyosarcoma cells.

OPA is a potent chelator of zinc and an effective inhibitor of DNA binding by zinc-finger proteins (23). Consistent with this property, formation of GATA-1 between the MELC extract and the GATA consensus sequence was gradually blocked by increasing amounts of OPA (Fig. 5*A*). The same result was obtained with the FP9 probe and the A-204 nuclear extract, thus implying that the factor recognizing the dyad of symmetry in the *COL11A1* promoter is most certainly a zinc-finger protein. Assuming a globular conformation for both proteins, the slower migration of FP9C compared with GATA-1 could be interpreted as suggesting that the former complex may be larger than the average GATA protein (~50 kDa) (21). To test this hypothesis, we estimated the molecular mass of the denatured protein by gel filtration chromatography. To this end, A-204 nuclear extract was denatured before being applied to the column equilibrated in the same denaturing buffer and the resulting fractions were renatured prior to be tested by the EMSA. Maximum FP9C binding activity was observed around a molecular range (80–120 kDa) significantly greater than the one expected for a GATA protein (Fig. 5*B*). A similar estimate was obtained for the native protein using glycerol gradient sedimentation analysis (data not shown). In sum, the above experiments suggested that an apparently novel ~100-kDa zinc-finger protein binds to a critical upstream regulator of the proximal *COL11A1* promoter.

FP9C Expression Is Cell Type-restricted—The absence of FP9C binding activity in MELC nuclei raised the possibility that the *COL11A1* activator may only be present in a restricted number of cells. To test this hypothesis, we undertook a systematic EMSA screen of nuclear extracts from normal and transformed cell lines. In addition to the rhabdomyosarcoma A-204, substantial amount of FP9C was observed in cells of mesenchymal origins, such as chondrocytes and smc, and in keratinocytes and vascular endothelial cells (Fig. 6*A*). By contrast, little or no FP9C binding activity was seen with nuclear extracts from fibroblasts, T-lymphocytes, neurogenic cells, hepatocytes, and melanocytes (Fig. 6*A*). Binding of the ubiquitous NF-1 protein to its cognate site was nearly the same in fibroblasts compared to A-204 nuclear extracts, thus indirectly validating the significant difference in FP9C content between the cell lines (Fig. 6*B*). Finally, Northern blot hybridizations established a positive correlation between FP9C binding activity and *COL11A1* gene expression (Fig. 6*C*). Altogether, the results strongly suggested that FP9C may be a regulator of cell type-specific transcription, in addition to controlling *COL11A1* gene expression.

Aside from FP9C, smc nuclear extracts yielded a second and less evident complex whose relative migration appeared similar to that of GATA-1 (Fig. 6*A*). This raised the possibility that

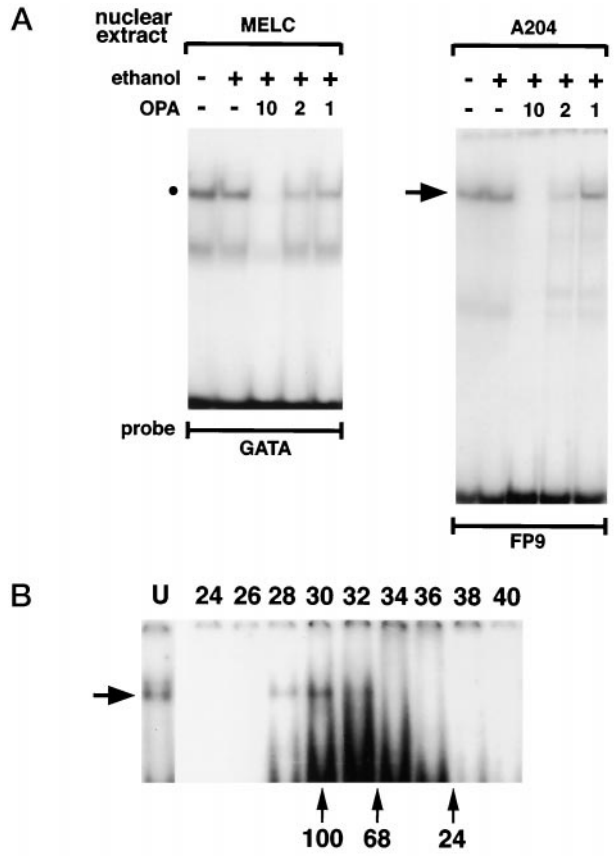

FIG. 5. **Characterization of FP9C.** *A*, EMSAs were performed with the indicated probes and nuclear extracts, and under increasing concentrations of OPA (in mM) for 5.5 h. Aside from the sample without OPA (−), another control included addition of ethanol to the binding reaction at the same final concentration as in the OPA-treated samples. *B*, gel filtration chromatography of denatured A-204 nuclear proteins. Aliquots (10 μl) from fractions were examined for FP9C binding activity by EMSA following renaturation. The migration of size markers (in kDa) is indicated above each autoradiography. In all panels, the *arrow* and the *closed circle* identify FP9C and GATA-1, respectively.

the second complex could either correspond to an smc-specific GATA protein or represent a product of FP9C degradation. A methylation interference test was therefore performed on each of the smc retarded bands to discriminate between these alternatives. As expected, the slower migrating EMSA band of the smc sample (B1) yielded the same contact points as the single band of the A-204 sample (Fig. 4, compare *A* and *B*). By contrast, the faster migrating band of the smc sample (B2) yielded a pattern different from FP9C but identical to GATA-1 (Fig. 4, compare *B* and *C*). Furthermore, competition with GATA consensus sequence eliminated the faster but not the slower migrating complex of the smc sample (data not shown). Recent expression data support the idea that the smc complex bound by FP9 is likely to correspond to GATA-6 (24). Aside from establishing the contact points of GATA-6, the methylation interference assay reiterated the specificity of FP9C binding and the identity of the FP9C protein in different cell lines. The competition experiments described in the previous section strongly suggested that FP9C binds to the 45-bp probe with significantly higher affinity than GATA-1 (Fig. 3*C*). Consistent with this postulate, addition of increasing amounts of smc nuclear extract to the FP9 binding reaction translated into substantially higher intensification of the B1 complex relatively to B2 (Fig. 4*B*). This last result was interpreted as indicating that GATA-6 binding *in vitro* may have no physiological relevance to *COL11A1* gene regulation.

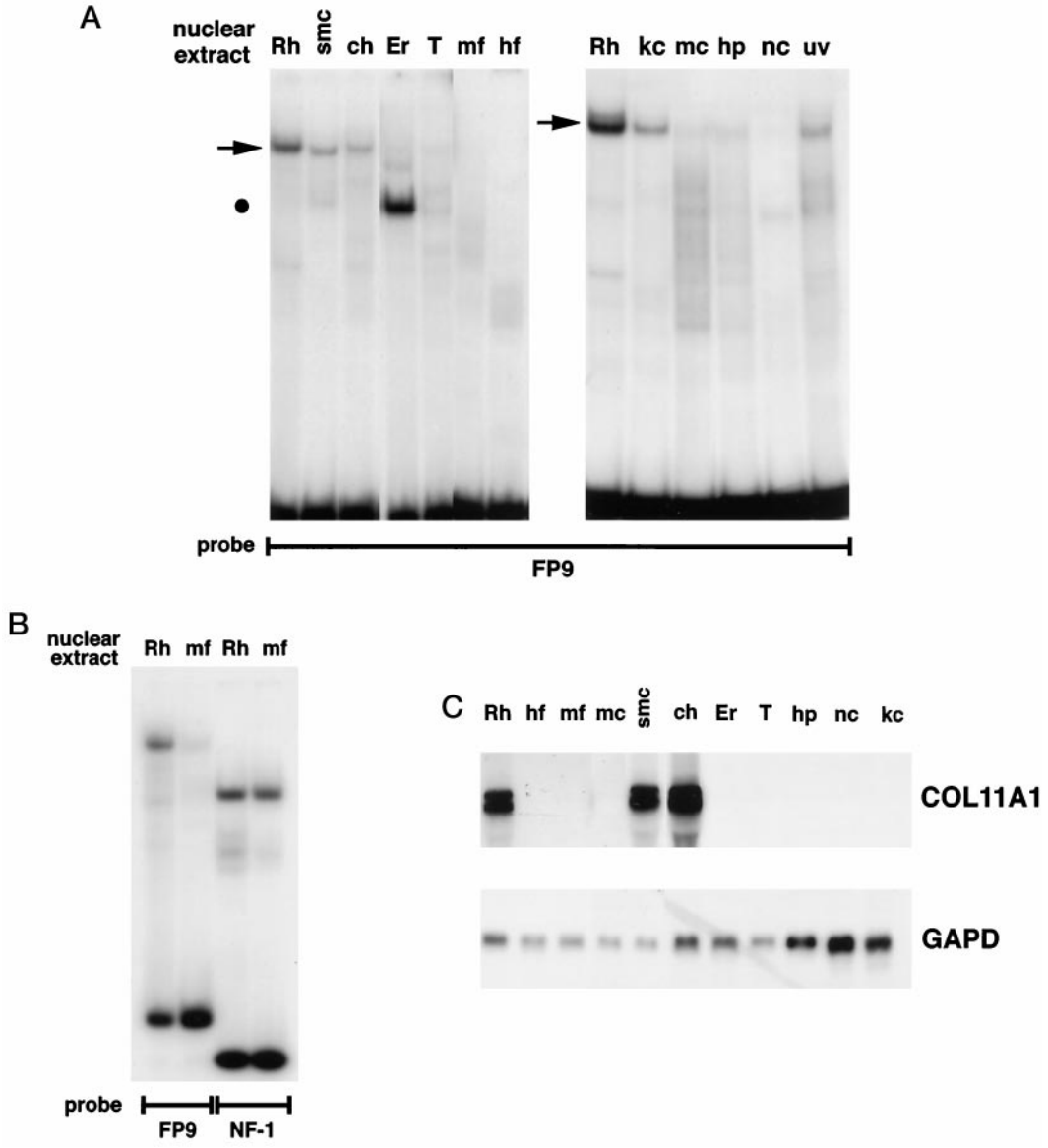

FIG. 6. **Screening cells for FP9C binding activity and *COL11A1* expression.** *A*, EMSAs were performed with the same amount of FP9 probe (10,000 dpm) and the same amount of nuclear extracts (5 μg) purified from A-204 (*Rh*), smc, chondrocytes (*ch*), MELC (*Er*), lymphocytes (*T*), NIH3T3 (*mf*) and CF37 (*hf*) fibroblasts, keratinocytes (*kc*), melanocytes (*mc*), hepatocytes (*hp*), and neurogenic cells (*nc*). The *closed circle* denotes the complex in common between smc and MELC extracts (see Fig. 3*B*), whereas the *arrow* points to FP9C. *B*, comparison of the binding intensity of F9PC and NF-1 proteins in A-204 *versus* NIH3T3 nuclear extracts. Equal amounts (5 μg) of nuclear extracts and radiolabeled probes (10,000 dpm) were added in each binding assay. *C*, Northern analysis of RNA from the above cells hybridized to the *COL11A1* and glyceraldehyde-3-phosphate dehydrogenase probes.

The screen for FP9C binding activity included myogenic cell lines that can be readily differentiated *in vitro* and osteoblastic representatives of different stages of differentiation. Together, they provided the experimental means to examine the possible relationship between FP9C expression and differentiation of skeletal muscle and bone cells. In the osteoblastic model, the transition from early to late stages of differentiation is represented by MC3T3-E1, ROS 25, and ROS 17/2.8 cells, in that order (25). EMSA revealed that FP9C binding increases gradually with differentiation (Fig. 7*A*). Comparable changes in FP9C expression were observed in the myogenic differentiation model (Fig. 7*A*). This is exemplified by undifferentiated and *in vitro* differentiated C2C12 myoblasts, and by undifferentiated C3H10T1/2 cells and the clonally derived P2 line differentiated *in vitro* (16). As a control, the same EMSAs were repeated using the binding site of a transcription factor that remains constant during skeletal muscle and bone cell differentiation (Fig. 7*B*). An additional control included the Northern analysis of *COL11A1* gene expression in the same panel of cell lines (Fig. 7*C*). Within the experimental limitations of the two models, the results were therefore consistent with the notion that FP9C may be implicated in terminal differentiation of cell lineages that share a common embryological origin.

DISCUSSION

The list of nuclear factors that control transcription in specific cell types is growing at an increasing pace, thanks to the implementation of new screening strategies and the continued progress in the study of tissue-specific gene regulation. Although more tedious and time-consuming, the latter approach has the virtue of identifying cell type-specific factors through the cognate *cis*-acting elements and thus, of readily relating them with the expression and function of the target genes. Examples of nuclear factors identified in this way include the erythroid GATA-1 protein and the B lymphocyte activator Oct-2, among several others (22, 26). Very recently, the study of

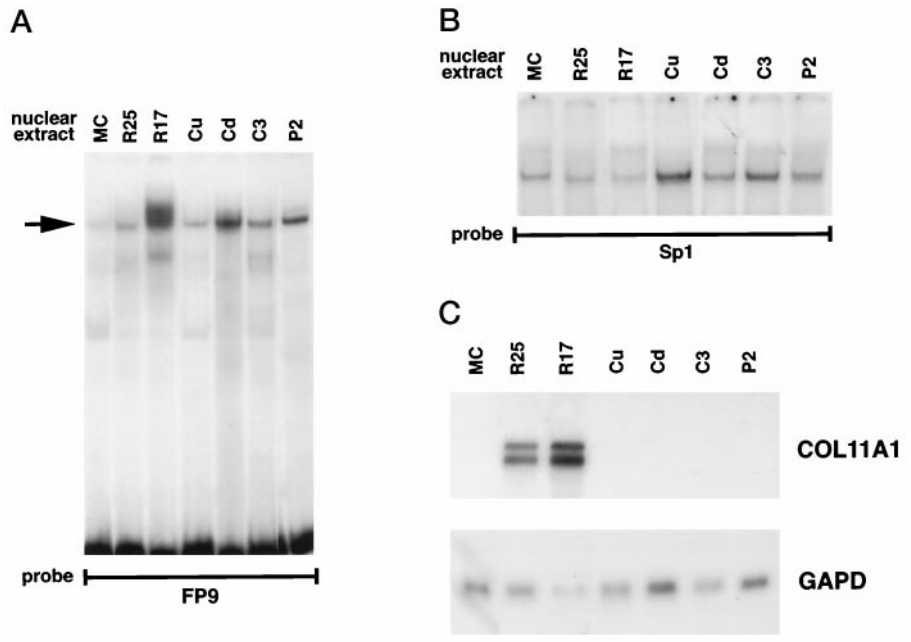

FIG. 7. **FP9C binding and cell differentiation.** *A* and *B*, EMSAs were performed using the same amount of labeled probe (10,000 dpm) and the same amount of nuclear proteins (5 μg) purified from MC3T3-E1 (*MC*), Ros 25 (*R25*), and Ros 17/2.8 (*R17*) cells; from undifferentiated (*Cu*) and *in vitro* differentiated (*Cd*) C2C12 myoblasts; and from C3H1OT1/2 (*C3*) and *in vitro* differentiated P2 cells (*P2*). The probes were the FP9 sequence (*A*) and the high affinity Sp1 binding site (*B*); the *arrow* in *A* points to FP9C. *C*, Northern analysis of RNA from the above cells hybridized to the *COL11A1* and glyceraldehyde-3-phosphate dehydrogenase probes.

the osteocalcin gene has led to the demonstration that Osf2/Cbfa1, the mammalian homologue of the *Drosophila* Runt protein, is a major determinant of osteoblasts differentiation and a regulator of several ECM-coding genes in bone (27–29). However, the Osf2/Cbfa1 case represents the exception rather than the rule among ECM-coding genes. Indeed, the genetic determinants responsible for orchestrating the assembly and remodeling of tissue-specific ECMs remain virtually unknown. By analogy to other systems, it is safe to predict that the diversification of matrix-coding gene expression in distinct mesenchymal cell lineages is probably achieved through combinatorial interactions of ubiquitous and cell type-specific factors. It is also reasonable to argue that expression of ECM-coding genes requires significantly more complicated networks of interactions than other tissue-specific gene families, both in terms of *trans*-acting factors and *cis*-acting elements. The collagen genes are a case in point. Current evidence excludes the existence of shared regulatory mechanisms involving common *cis*-acting elements. Although fibroblast-specific elements have been identified in the coordinately regulated type I collagen genes, they differ in composition and interact with distinct nuclear proteins (4, 5, 30–34). A SOX9 binding activity has recently been reported to be necessary for cartilage-specific production of type II collagen (8, 9). By contrast, there is no evidence of SOX9 participation in the exquisite specificity of α2(XI) collagen gene expression in restricted domains of the developing limbs (35). Relevant to the present study, there is also no apparent similarity between the *cis*-acting elements of the α2(XI) collagen gene and of the one coding for the α1(XI) partner. We believe the results described in this report are consistent with having identified a cell type-specific regulator of collagen gene expression.

Although solely based on transient transfection assays, FP9 adheres to the two experimental criteria that are normally used to define a tissue-specific element. First, FP9 is required for transcription from the proximal −541*COL11A1* promoter. This was demonstrated in functional assays that utilized progressively shorter promoter segments, internal deletions, and sequence substitutions of and within FP9. Particularly convincing was the finding that transcription from the −541 promoter depends on the 20-bp core sequence of FP9, the element that is principally involved in the binding of the cognate factor. To our surprise, however, the effects of more subtle mutations within the 20-bp core were not as drastic on transcription as they were on *in vitro* binding (data not shown). One of the problems with the functional evaluation of the promoter is the intrinsic weakness of the *COL11A1* promoter. Work with transgenic mice is currently searching for upstream and downstream elements that may enhance promoter expression. The problem of promoter strength notwithstanding, FP9 conforms to probably the most important criterion for tissue specificity. It in fact binds a transcription factor, FP9C, which is produced in significant amounts only by a restricted number of cells. Albeit far from being exhaustive, the screen nevertheless suggests that FP9C is predominantly, but not exclusively, a mesenchymal gene product. Relevant to the main scope of the study, FP9C is consistently found in cells actively engaged in the synthesis of α1(XI) collagen. As already mentioned, the binding specificity of FP9C is dictated by a sequence that consists of two GATA-like motifs arranged in a palindromic configuration interrupted by the GG dinucleotide. Like the remaining of the −541 promoter, this element is remarkably conserved in the mouse gene (data not shown). In point of fact, there are only two nucleotide substitutions in the 20-bp core sequence of the two promoters. The substitutions result in a GG → AA transition in the mouse gene at the 2-bp interruption of the GATA palindrome; this is the very same mutation (M1) that was found inconsequential for FP9C binding (Fig. 3*A*). FP9 can bind GATA-1 *in vitro* nearly as efficiently as the GATA consensus sequence. However, failure of the GATA consensus sequence to interfere with formation of FP9C clearly demonstrated that the 20-bp core is bound by the natural complex with very strong affinity, certainly higher than GATA-1. Along this line, incremental addition of smc nuclear extract augmented binding of FP9C significantly more than GATA-6. The sensitivity of FP9C to OPA treatment is widely regarded as a strong indication that the factor probably belongs to the zinc-finger family of proteins. Additional analyses suggested (but did not prove) that FP9C consists of a single polypeptide of about 100 kDa. Altogether the size of the nuclear factor, the composition of the cognate site and the functional effects of the FP9 mutations indicate that FP9C is probably a novel zinc-finger activator.

By extrapolating from the *in vitro* data, we suggest that organs actively involved in FP9C expression may include the

vascular system, skeletal muscle, bone, cartilage, and probably skin. In this respect, FP9C joins the list of transcription factors with restricted tissue distribution. One of such examples is *MHox*, whose expression in skeletal muscles and chondrocytes resembles that of FP9C (36). Relevant to ECM-coding genes, FP9C represents the first regulator that is expressed in both hard and soft connective tissue. In at least two of the available systems, myoblasts and osteoblasts, FP9C levels seem to increase in concert with cell differentiation. Within the limitation of the models employed, we propose that FP9C is not only a tissue-specific regulator but also a determinant of cell differentiation. The cloning of the FP9C gene will provide the means to eventually test this hypothesis and elucidate the mechanisms underlying FP9C regulation of *COL11A1* transcription. More generally, it may also shed new light on the pathways that govern cell commitment and differentiation in embryologically related but phenotypically distinct lineages.

Acknowledgments—We thank Drs. B. de Crombrugghe, M. Goldfarb, G. Karsenty, J. H. Kimura, V. Lefebvre, D. Sassoon, and M. Takigawa for helpful suggestions and invaluable reagents. We also thank Karen Frith for typing the manuscript and Cindy Else for excellent technical assistance.

REFERENCES

1. Hay, E. D. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., ed) 2nd Ed., pp. 419–462, Plenum Press, New York
2. Raghow, R. (1994) *FASEB J.* **9,** 823–831
3. Slack, J. L., Liska, D., and Bornstein, P. (1993) *Am. J. Med. Genet.* **45,** 140–151
4. Pavlin, D., Lichtler, A. C., Bedalov, A., Kream, B. E., Harrison, J. R., Thomas, H. F., Gronowicz, G. A., Clark, S. H., Woody, C. O., and Rowe, D. W. (1992) *J. Cell Biol.* **116,** 227–236
5. Rossert, J. A., Eberspaecher, H., and de Crombrugghe, B. (1995) *J. Cell Biol.* **123,** 1421–1432
6. Rossert, J. A., Chen, S. S., Eberspaecher, H., Smith, C. N., and de Crombrugghe, B. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93,** 1027–1031
7. Dodig, M., Kronenberg, M. S., Bedalov, A., Kream, B. E., Gronowicz, G., Clark, S. H., Mack, K., Liu, Y.-H., Maxon, R., Pan, Z., Upholt, W. B., Rowe, D. W., and Lichtler, A. C. (1996) *J. Biol. Chem.* **271,** 16422–16429
8. Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N., and de Crombrugghe, B. (1997) *Mol. Cell. Biol.* **17,** 2336–2346
9. Bell, D. M., Leung, K. K. H., Wheatley, S. C., Ng, L. J., Zhou, S., Ling, K. W., Sham, M. H., Koopman, P., Tam, P. L., and Cheah, K. S. E. (1997) *Nat. Genet.* **16,** 174–178
10. Bou-Gharios, G., Garrett, L. A., Rossert, J., Niederreither, K., Eberspaecher, H., Smith, C., Black, C., and de Crombrugghe, B. (1996) *J. Cell Biol.* **134,** 1333–1344
11. Linsenmayer, T. F. (1991) in *Cell Biology of Extracellular Matrix* (Hay, E. D., ed) 2nd Ed., pp. 7–44, Plenum Press, New York
12. Andrikopoulos, K., Liu, X., Keene, R. D., Jaenisch, R., and Ramirez, F. (1995) *Nat. Genet.* **3,** 31–36
13. Li, Y., Lacerda, D. A., Warman, M. L., Beier, D. R., Yoshioka, H., Ninomiya, Y., Oxford, J. T., Morris, N. P., Andrikopoulos, K., Ramirez, F., Wardell, B. B., Lifferth, G. D., Teuscher, C., Woodward, S. R., Taylor, B. A., Seegmiller, R. E., and Olsen, B. R. (1995) *Cell* **80,** 423–430
14. Yoshioka, H., Greenwel, P., Inoguchi, K., Truter, S., Inagaki, Y., Ninomiya, Y., and Ramirez, F. (1995) *J. Biol. Chem.* **270,** 418–424
15. Truter, S., Di Liberto, M., Inagaki, Y., and Ramirez, F. (1992) *J. Biol. Chem.* **267,** 25389–25395
16. Relaix, F., Weng, X., Marazzi, G., Yang, E., Copeland, N., Jenkins, N., Spence, S. E., and Sassoon, D. (1996) *Dev. Biol.* **177,** 383–386
17. Lefebvre, V., Zhou, G., Mukhopadhyay, K., Smith, C. N., Zhang, Z., Eberspaecher, H., Zhou, X., Sinha, S., Maity, S. N., and de Crombrugghe, B. (1996) *Mol. Cell. Biol.* **16,** 4512–4523
18. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1983) *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
19. Lackow, B., and Schutz, G. (1987) *Nucleic Acids Res.* **15,** 5490
20. Boast, S., Su, M. W., Ramirez, F., Sanchez, M., and Avvedimento, E. V. (1990) *J. Biol. Chem.* **265,** 13351–13356
21. Simon, M. C. (1995) *Nat. Genet.* **11,** 9–11
22. Tsai, F. Y., Martin, D. K., Zon, L. I., D'Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) *Nature* **339,** 446–451
23. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51,** 1079–1090
24. Morrisey, E. E., Ip, H. S., Lu, M. M., and Parmacek, M. S. (1996) *Dev. Biol.* **177,** 309–322
25. Majeska, R. J., Rodan, S. B., and Rodan, G. A. (1980) *Endocrinology* **107,** 1421–1432
26. Singh, H., Sen, R., Baltimore, D., and Sharp, P. A. (1986) *Nature* **319,** 154–158
27. Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L., and Karsenty, G. (1997) *Cell* **89,** 747–754
28. Komori, T., Yagi, H., Nomura, S., Yamaguchi, A., Sasaki, K., Deguchi, K., Shimizu, Y., Bronson, R. T., Gao, Y. H., Inada, M., Sato, M., Okamoto, R., Kitamura, Y., Yoshiki, S., and Kishimoto, T. (1997) *Cell* **89,** 755–764
29. Otto, F., Thornell, A. P., Crompton, T., Denzel, A., Gilmour, K. C., Rosewell, I. R., Stamp, G. W. H., Beddington, R. S. P., Mundlos, S., Olsen, B. R., Selby, P. B., and Owen, M. J. (1997) *Cell* **89,** 765–771
30. Karsenty, G., Golumbeck, P., and de Crombrugghe, B. (1988) *J. Biol. Chem.* **263,** 13909–13915
31. Goldberg, H., Helaakoski, T., Garrett, L. A., Karsenty, G., Pellegrino, A., Lozano, G., Maity, S., and de Crombrugghe, B. (1992) *J. Biol. Chem.* **267,** 19622–19630
32. Inagaki, Y., Truter, S., and Ramirez, F. (1994) *J. Biol. Chem.* **269,** 14828–14834
33. Inagaki, Y., Truter, S., Tanaka, S., Di Liberto, M. and Ramirez, F. (1995) *J. Biol. Chem.* **270,** 3353–3358
34. Tamaki, T., Ohnishi, K., Hartl, C., LeRoy, E. C., and Trojanowska, M. (1995) *J. Biol. Chem.* **270,** 4299–4304
35. Tsumaki, N., Kimura, T., Matsui, Y., Nakata, K., and Ochi, T. (1996) *J. Cell Biol.* **134,** 1573–1582
36. Cserjesi, P., Lilly, B., Brynson, L., Wang, Y., Sassoon, D. A., and Olson, E. N. (1992) *Development* **115,** 1087–1101