Nucleic Acids Research, 1996, Vol. 24, No. 19 **3661–3669**

SP1-binding elements, within the common metaxin-thrombospondin 3 intergenic region, participate in the regulation of the metaxin gene

Malcolm Collins and Paul Bornstein*

Department of Biochemistry, Box 357350, University of Washington, Seattle, WA 98195, USA

Received August 1, 1996; Accepted August 13, 1996

DDBJ/EMBL/GenBank accession no. U66257

ABSTRACT

Metaxin (*Mtx*) is an essential nuclear gene which is expressed ubiquitously in mice and encodes a mitochondrial protein. The gene is located upstream and is transcribed divergently from the thrombospondin 3 (*Thbs3*) gene; 1352 nucleotides separate the putative translation start sites. Although the *Mtx* and *Thbs3* genes share a common intergenic region, transient transfection experiments in rat chrondrosarcoma cells and in NIH-3T3 fibroblasts demonstrated that the elements required for expression of the *Mtx* gene are situated within a short proximal promoter and have no major effect on the transcription of *Thbs3*. The metaxin –377 bp promoter contains four clustered GC boxes between nucleotides –146 and –58 and an inverted GT box between nucleotides –152 and –161, but does not contain TATA or CCAAT boxes. Like many genes regulated by a TATA-less promoter, the transcription start site of metaxin is heterogeneous. The major start site is only 13 bp upstream from the putative translation start site. Electrophoretic mobility shift, competition and supershift assays showed that the ubiquitous transcription factor, Sp1, and, to a lesser extent, the Sp1-related protein, Sp3, bind to four of these Sp1-binding motifs. Co-transfection of metaxin promoter–luciferase constructs and an Sp1 expression vector into Schneider *Drosophila* cells, which do not synthesize Sp1, demonstrated that the metaxin gene is activated by Sp1. Deletion of the four upstream Sp1-binding elements, on the other hand, demonstrated that these motifs are superfluous in context of the larger *Mtx* promoter. Thus, despite the potential for common regulatory mechanisms, the available evidence indicates that the *Mtx* minimal promoter does not significantly affect *Thbs3* gene expression.

INTRODUCTION

The mouse metaxin (*Mtx*) gene has been mapped to chromosome 3E3-F1 and is closely linked to the episialin (*Muc1*), glucocerebrosidase (*Gba*) and thrombospondin 3 (*Thbs3*) genes (1,2). The transcriptional start site of Muc1, which encodes a tumor-associated polymorphic epithelial mucin (3,4), is located 2.3 kb downstream from the polyadenylation signal of *Thbs3* (5), while the *Gba* polyadenylation signal is situated only 6 kb upstream from the 5′-end of *Thbs3* (2). Mutations within the *GBA* gene are responsible for Gaucher disease, the most common lysosomal storage disorder in humans (reviewed in 6), while thrombospondin 3 is a secreted extracellular matrix glycoprotein of unknown function (7). The fourth transcription unit in this region, the metaxin gene, is situated within the 6 kb region separating the glucocerebrosidase and thrombospondin 3 genes. Thus, *Mtx* and *Gba* are transcribed convergently and *Mtx* and *Thbs3* are transcribed divergently (2). The major polyadenylation sites of *Mtx* and *Gba* are only 431 bp apart, while the putative translation start sites of the *Mtx* and *Thbs3* genes are separated by 1.4 kb of DNA. It is, therefore, possible that the primary transcripts of the *Mtx* and *Gba* genes overlap and that *Mtx* and *Thbs3* are regulated by common promoter elements.

In an attempt to establish a mouse model for the mild form of Gaucher disease, an A to G substitution was introduced into exon 9 of the mouse *Gba* gene by homologous recombination in embryonic stem cells. The selectable marker, phosphoglycerate kinase-neomycin, was inserted downstream from the *Gba* gene, within, as it turned out, the terminal exon of metaxin (2). Homozygous mice that contain the disrupted *Mtx* gene die early during gestation (2). These findings strongly suggested that the metaxin gene was required for embryonic development and that it has an essential function. Unlike the *Thbs3* gene, which has a more limited distribution (5,8,9), the metaxin gene is expressed ubiquitously in murine tissues with the highest levels detected in the kidney (2). The 1.3 kb metaxin mRNA encodes a protein of 317 amino acids. The 35 kDa protein has been localized to the mitochondria and cytosol (10), and preliminary data suggest that metaxin is part of the mitochondrial import machinery (L.C. Armstrong, T. Komiya, K. Mihara and P. Bornstein, manuscript in preparation). Furthermore, metaxin has significant sequence similarity with the yeast mitochondrial import protein, Mas37p, (10) and a mitochondrial localization sequence has also been identified within the C-terminus of the protein (L.C. Armstrong, T. Komiya, K. Mihara, and P. Bornstein, manuscript in preparation). This finding strengthens the hypothesis that metaxin is involved in mitochondrial protein import.

* To whom correspondence should be addressed

An increasing number of eukaryotic genes, that are in a head-to-head orientation and are transcribed in opposite directions, have been characterized. These gene pairs include, among others, the *COL4A1* and *COL4A2* genes (11,12), the human dihydrofolate reductase and mismatch protein 1 genes (13), the chicken glycinamide ribonucleotide transformylase and 5-amino-imidazole ribonucleotide carboxylase genes for *de novo* purine nucleotide synthesis (14), and the human histidyl-tRNA synthetase (HRS) and HRS homologous genes (15). The transcription start sites of these divergent genes are separated by fairly short intergenic regions, ranging from 114 to 229 bp. The transcriptionally divergent yeast histone H2A and H2B genes (16), *Gal1* and *Gal10* (17), and Drosophila yolk protein genes, *yp1* and *yp2* (18), on the other hand, are separated by much larger intergenic regions of between 606 and 1225 bp in length. The majority of these genes have related functions and are coordinately regulated by bi-directional promoters. However, there are examples of divergently transcribed genes which are regulated by a bi-directional promoter and encode proteins that have no apparent relationship, like the human dihydrofolate reductase and mismatch protein 1 genes (13).

Since the divergently transcribed metaxin and thrombospondin 3 genes could share common *cis*-acting elements, the minimal promoter region in metaxin was identified and partially characterized in this study. Promoter elements in metaxin were mapped to the 377 bases upstream from its translation start site and were shown to contain four functional Sp1-binding motifs. Three of the distal functional Sp1-binding elements were nevertheless shown to be superfluous when deleted from the –1.7 kb metaxin promoter, since additional upstream elements appeared to compensate for their absence. The influence of this minimal metaxin promoter on thrombospondin 3 gene expression was also studied, and no major effects were detected.

MATERIALS AND METHODS

DNA sequence analysis

A 2.3 kb *Hin*dIII–*Bam*HI fragment, containing the mouse metaxin-thrombospondin 3 intergenic region, the first exons of both genes and the 5′-portions of their first introns (1), was digested with the appropriate restriction enzymes (see Figs 3 and 7) and subcloned into Bluescript. Double-stranded plasmid DNA was sequenced using the dideoxy chain-terminating method and the Taq DyeDeoxy™ Terminator Cycle Sequencing Kit (Applied Biosystems), as described by the manufacturer. The samples were analyzed with an Applied Biosystems model 373A automated sequencer and the resulting sequences were analyzed with the GENEPRO (Riverside Scientific, Seattle) or the Genetics Computer Group (GCG) programs.

5′-Rapid amplification of cDNA ends (5′-RACE)

Poly A+ RNA was extracted from 4-week-old female BALB/c mouse liver and kidney by established methods, and the transcription start site of the metaxin gene was determined using 5′-rapid amplification of cDNA ends (5′-RACE). Briefly, cDNA was synthesized from 2 µg mouse liver or kidney poly A+ RNA using an oligonucleotide primer (5′-GGAAAGGGCATAGC-CTCTGCATATC-3′) corresponding to sequences within exon 5 of the metaxin gene (2). The cDNA was purified and ligated to the anchor oligonucleotide using the AmpliFINDER™ RACE Kit (Clontech), as described by the manufacturer. The anchor-ligated metaxin cDNA or mouse liver 5′-RACE-Ready™ cDNA (Clontech) was PCR amplified using 20 pmol of anchor primer, 20 pmol of primer (5′-GATGGGTGATCTTGTCTGGCAC-3′) corresponding to sequences within exon 3, and the Stoffel fragment of AmpliTaq DNA polymerase (Perkin Elmer) in a final reaction volume of 100 µl containing 10 mM Tris–HCl (pH 8.3); 10 mM KCl; 2.5 mM $MgCl_2$ and 0.2 mM each dATP, dCTP, dGTP and dTTP. Two µl of the primary PCR reaction mixture was re-amplified as above with 20 pmol each of the anchor primer and a nested metaxin primer (5′-CTAGGATCCAGCACGGCCAGACTATCCAG-3′) corresponding to sequences within exon 1. Thirty five cycles, denaturing at 95°C for 1 min, annealing at 55°C for 1 min and, extending at 72°C for 2 min, with a final extension time of 7 min were used for both the primary and secondary PCR reactions. The secondary PCR products were cloned into pUC19 and 96 positive clones were analyzed by restriction digestion. Selected clones were sequenced as described above.

Cell culture and transient transfection assays

Rat chondrosarcoma cells (RCS) were a gift from Dr J. Kimura and were originally obtained from a rat Swarm chrondrosarcoma as described in Choi *et al.* (19) and Mukhopadhyay *et al.* (20). RCS cells were cultured as monolayers in high glucose Dulbecco's modified Eagle's medium (DMEM), supplemented with 20% heat-inactivated fetal calf serum, 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin and 0.25 µg/ml amphotericin B. NIH-3T3 cells (ATCC CRL-1658) were cultured in low glucose DMEM containing the above supplements, except that 10% heat-inactivated fetal calf serum was added to the medium. Schneider Drosophila Line 2 (SL2) cells (ATCC CRL-1963) were cultured in Schneider Drosophila Medium (Gibco) containing 10% heat-inactivated fetal bovine serum (Sigma, F-3018), 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin and 0.25 µg/ml amphotericin B at 25°C.

The *Mtx* and *Thsb3* promoter–luciferase reporter gene constructs were prepared by digesting the 2.3 kb *Hin*dIII–*Bam*HI fragment of clone 11 (1) with the appropriate restriction enzymes (see Figs 3, 6 and 7) and ligating the fragments into the pGL2-Basic vector (Promega). *Rsa*I and *Cel*II sites situated immediately upstream from the *Mtx* and *Thbs3* translation start sites, respectively, were used to define the 3′-ends of the promoters.

Sub-confluent NIH-3T3 and RCS cells were transiently transfected with 10 µg of the promoter–luciferase plasmid constructs in 60 mm Petri dishes using the calcium phosphate–DNA precipitation method (21). To control for variations in transfection efficiency, the cells were co-transfected with 2 µg of a β-galactosidase construct in which the gene is driven by the SV40 promoter and enhancer (Promega). SL2 cells were transfected with 5 µg of the luciferase construct in 60 mm dishes using the calcium phosphate–DNA precipitation method (22). These cells were co-transfected with 4 µg of the plasmid δF-gal containing the β-galactosidase gene driven by the Drosophila hsp70 core promoter and the transposable element F enhancer-like sequences to control for transfection efficiency and with 1 µg of either the human Sp1 expression vector, pPacSp1, or the parent vector pPac0. Plasmids pPacSp1 and pPacO were a gift from Dr R. Tjian (23), while plasmid δF-gal was a gift from Dr P. De Nocera (24). Cell lysates were prepared from the transfected eukaryotic and insect cells using the freeze–thaw method (21). The luciferase activity in the extracts were measured using the

Luciferase Assay System (Promega) as described by the manufacturer, while the β-galactosidase activity was assayed using *O*-nitophenyl-β-D-galactopyranoside as a substrate (21), or the Galacto-Light Chemiluminescent Reporter Assay (Tropix) as described by the manufacturer.

Electrophoretic mobility shift assays (EMSA)

The various DNA fragments used as probes in EMSA (Fig. 4A) were prepared from the mouse metaxin proximal promoter, radiolabeled with the Klenow fragment of DNA polymerase. Nuclear proteins were isolated from NIH-3T3 and RCS cells using the method of Lee and Green (25) and their concentrations were determined by the Bradford method (26) using chymotrypsin as a standard. Crude nuclear extract (4 µg) was incubated in 20 mM HEPES (pH 7.9), 50 mM KCl, 0.5 mM dithiothreitol, 0.2 mM EDTA, 1 mM $MgCl_2$, 4% Ficoll 400 and 4 µg poly dI-dC.poly dI-dC in a final volume of 20 µl for 10 min at room temperature, prior to the addition of 10^4 c.p.m. ^{32}P-labeled probe. The incubation was continued for an additional 30 min at 4°C and the DNA–protein complexes were analyzed on 5% non-denaturing polyacrylamide gels in 0.5× TBE at 4°C. The gels were dried and exposed to X-ray film for 16 h.

In the competition assays, double-stranded oligonucleotides were prepared by denaturing complementary single-stranded oligonucleotides (synthesized by IDT®) at 90°C for 5 min and allowing them to anneal at 37°C for 60 min prior to cooling to room temperature. The unlabeled double-stranded GC box consensus (5′-ATTCGATCGGGGGGGGGGCGAGC-3′) or GC box mutant (5′-ATTCGATCGGTTGGGGGCGAGC-3′) oligonucleotides were incubated with the nuclear proteins prior to the addition of the ^{32}P-labeled probe.

In the supershift assays, 1.0 µl anti-Sp1 (PEP2), Sp2 (K-20), Sp3 (D-20) or Sp4 (V-20) Trans Cruz™ affinity-purified rabbit polyclonal IgG (Santa Cruz Biotechnology) was added to the reaction mixtures after the probe and incubated at 4°C for 60 min. To test for the specificity of the antibodies, 1.0 µl rabbit IgG was incubated with the appropriate blocking peptide prior to its addition to the reaction mixture.

RESULTS

Characterization of the metaxin–thrombospondin 3 intergenic region

In a previous study we identified a novel murine gene, metaxin, which is located within the 6 kb region separating the glucocerebrosidase (*Gba*) and thrombospondin 3 (*Thbs3*) genes on chromosome 3E3-F1 (2). Metaxin is situated upstream from the thrombospondin 3 gene, in a head-to-head orientation, so that the putative translation start sites of the two genes are separated by 1352 base pairs of intergenic sequence (Fig. 1). Analysis of the intergenic region revealed that it is G+C rich (61%), that it contains 86 CpG dinucleotides, and that both the metaxin and thrombospondin 3 promoters lack a CCAAT and a TATA box. Comparison of the mouse and human (27) intergenic regions reveals an identity of 59%. However, there is a sequence of 184 bp situated between nucleotides –592 and –409 of the mouse intergenic region which is 88% identical between the two species (Fig. 1). The sequence of the mouse intergenic region has been deposited in the DDBJ/EMBL/GenBank Data Libraries under Accession no. U66257.

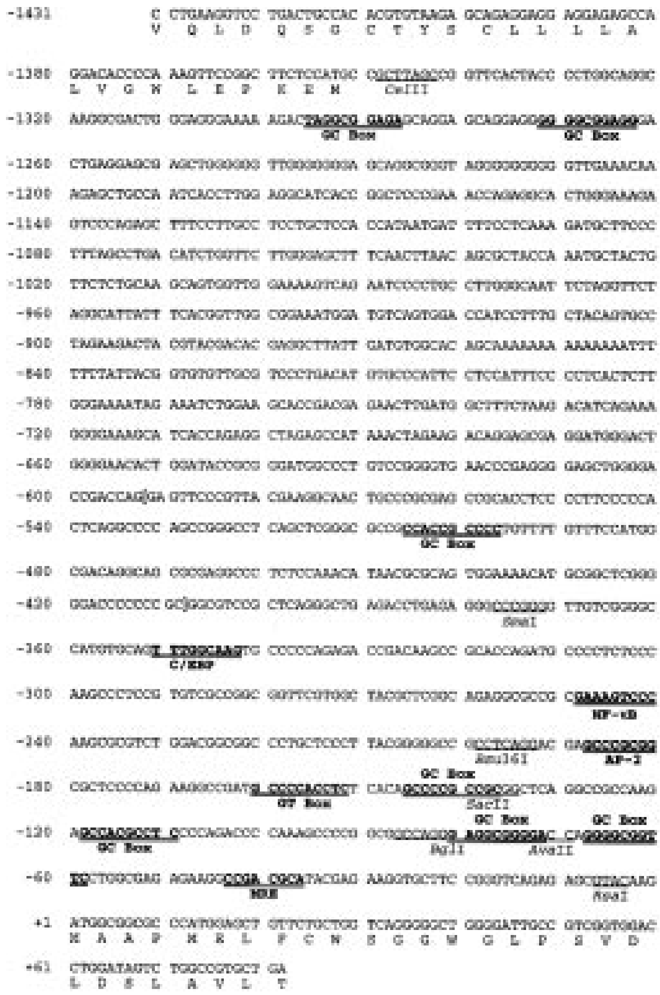

Figure 1. Nucleotide sequence of the mouse thrombospondin 3–metaxin intergenic region and the first exons of the two genes. The reverse complement sequence of the *Thbs3* first exon is shown. The deduced amino acid sequences of the *Mtx* (downstream) and *Thbs3* (upstream) first exons are indicated. Selected restriction enzyme cleavage sites are underlined and the putative translation start site of *Mtx* is defined as +1. A sequence which is 88% identical in the mouse and human intergenic regions, is indicated by square brackets. The major transcription factor binding motifs (in bold with double underlining), in the –377 bp proximal promoter of the metaxin gene, and additional GC boxes upstream from the –377 bp promoter, are also indicated. These sequences were identified by searching databases of transcription factors.

Identification of the transcription start site of the metaxin gene

The transcription start site of the metaxin gene was determined using 5′-rapid amplification of cDNA ends (5′-RACE) as described in Materials and Methods. The transcription start site was shown to be heterogeneous (Fig. 2) with the major start site being only 13 bp upstream from the putative translation start site (seven out of 17 clones sequenced). The longest and shortest 5′-UTR were determined to be 26 and 7 bp, respectively. As shown in Figure 2, four overlapping initiator (INR)-like elements were identified within the short untranslated region. All four sequences (5′-GTGCTTC-3′; 5′-TGCTTCC-3′; 5′-CCGGGTC-3′ and 5′-TCAGAGA-3′), contained 2-base mismatches from the

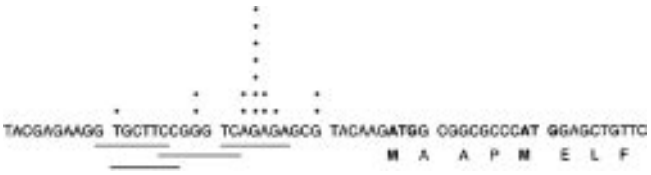

Figure 2. Transcription start sites of the metaxin gene. Transcription start sites were determined using 5′-rapid amplification of cDNA ends, as described in Materials and Methods. The identified start sites are indicated with asterisks and the deduced amino acid sequence from the putative translation start site (in bold) is also indicated. An in-frame downstream methionine residue and its codon are also shown in bold. The four INR-like sequences are underlined.

consensus sequence, 5′-YYANWYY-3′ (28). The majority of the transcription start sites were mapped to the downstream INR-like element, suggesting that this element was functionally the most important. RT–PCR of liver or kidney poly A+ RNA using 5′-primers which flanked the most upstream and downstream transcription start sites, or included some of these sites, confirmed the 5′-RACE results (data not shown).

Mapping the minimal metaxin promoter

To determine the minimal promoter elements required to stimulate transcription of the metaxin gene, several 5′-deletion fragments from the thrombospondin 3-metaxin intergenic region were cloned upstream from the luciferase reporter gene in the metaxin direction (Fig. 3). In addition to the intergenic region, the largest construct (–1661) also contained exon A and the 5′-region of intron A of the thrombospondin 3 gene. The various metaxin promoter–luciferase constructs were co-transfected, together with the β-galactosidase reporter gene, into a rat chondrosarcoma cell line (RCS) and into NIH-3T3 fibroblasts. Northern blot analysis and/or RNase protection assays showed that the metaxin and thrombospondin 3 genes were transcribed at relatively high levels in RCS cells, while only trace amounts of the metaxin message were detected in NIH-3T3 cells (data not shown). Similar luciferase activity was obtained when all of the 5′-deletions, from the *Bam*HI site, 1661 bases upstream from the metaxin translation start site, to the *Sma*I site at –377, were assayed in both cell lines (Fig. 3). There was, however, a slight decrease in luciferase activity when the –635 (*Hph*I) promoter construct was assayed in both RCS ($P = 0.04$) and NIH-3T3 ($P = 0.002$) cells, suggesting that factors which bind upstream from the *Sma*I site play a minor role in the regulation of the metaxin gene. Further analysis of the –377 and –138 bp proximal promoters showed that there was a significant difference in expression of luciferase when the activities of these constructs were compared with that of the –83 bp promoter construct. Thus, the major elements responsible for activating the metaxin gene are primarily situated within the proximal promoter between –138 and –83 bp upstream from the translation start site.

A comparison of luciferase activity in extracts prepared from cells transfected with the largest metaxin promoter construct, with those transfected with the luciferase gene driven by the strong SV40 promoter, showed that the relative activity of the metaxin promoter was 64.3 ± 5.9% (n = 4) and 38.9 ± 6.6% (n = 4) in RCS and NIH-3T3, cells respectively. These findings suggest that the metaxin promoter functions as a fairly strong promoter in both cell lines. However, the constructs that were tested did not appear to be dramatically less active in NIH-3T3 fibroblasts, which only produce trace amounts of metaxin mRNA, than in RCS cells. *Cis*-acting elements situated upstream or downstream from the –1661 bp promoter and/or post-transcriptional mechanisms may be responsible for down-regulating metaxin in NIH-3T3 fibroblasts.

Putative regulatory elements within the metaxin proximal promoter

Since the –377 bp proximal promoter contained all the major *cis*-acting elements required for directing expression of the metaxin gene in transfected cells, candidate regulatory elements were identified within the minimal promoter by searching the transcription factor databases, TFD sites (release 7.4) (29) and

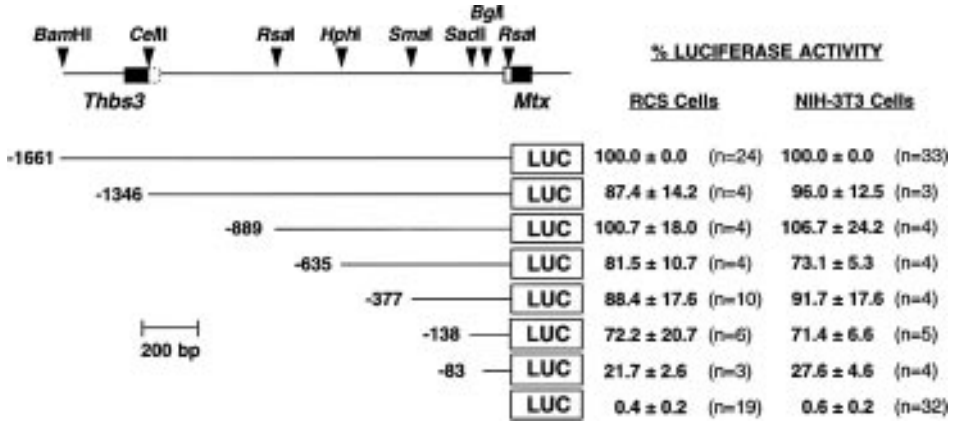

Figure 3. Deletion analysis of the mouse metaxin promoter in RCS and NIH-3T3 cells. A schematic diagram of the mouse thrombospondin 3–metaxin intergenic region and the 5′-portions of the two genes is shown. The translated sequences of metaxin exon 1 and thrombospondin 3 exon A (filled boxes), and the 5′-UTRs (open boxes), are also shown. The thrombospondin 5′-untranslated region is shown as a dashed box because the transcription start site has not yet been determined. The indicated restriction enzymes were used to clone several 5′-deletions of the intergenic region into pGL2-Basic. An *Rsa*I site situated within the short 5′-UTR of metaxin served as the 3′-boundary of the metaxin promoter. The metaxin promoter–luciferase constructs were co-transfected, with SV40-β-gal, into RCS and NIH-3T3 cells, as described in Materials and Methods. The percentage luciferase activity, expressed relative to the value for the –1661 bp construct, standard deviations for each promoter construct, and number of determinations (n) are indicated.

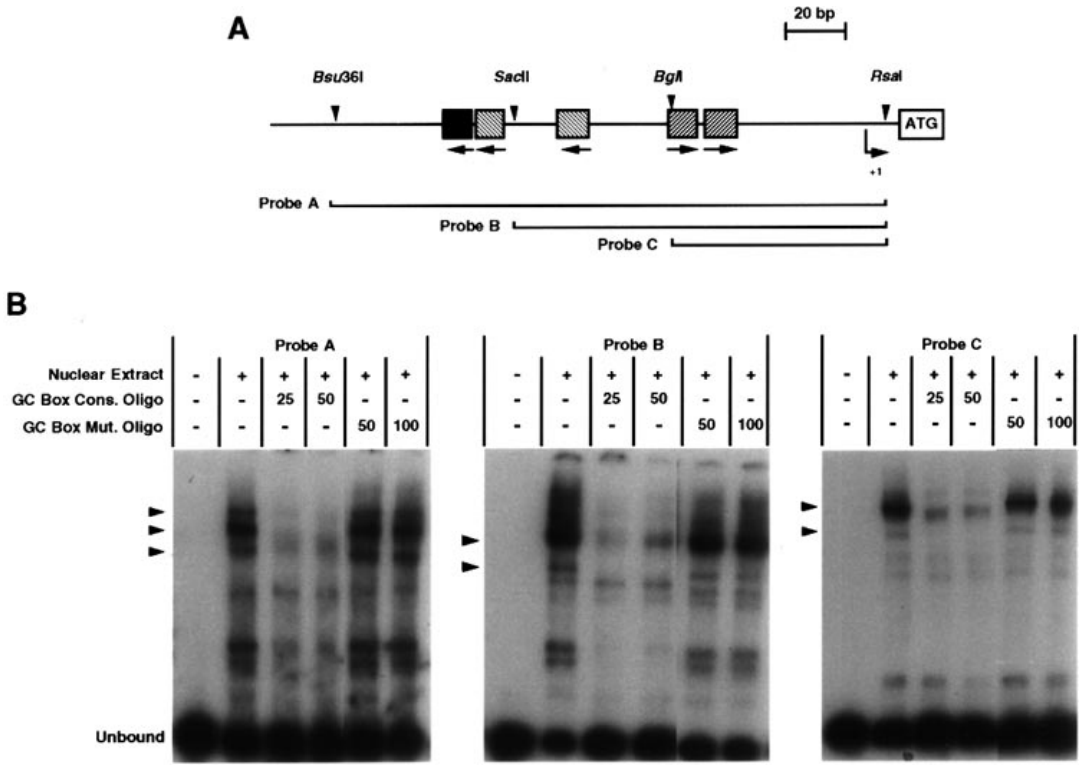

Figure 4. Analysis of DNA–protein interactions in the metaxin promoter by EMSA. (**A**) A schematic diagram of the proximal promoter (solid line) and the putative translation start site (open box) of the metaxin gene. The relative positions and orientations (arrows) of the four GC boxes (hatched boxes) and the GT box (filled box) are shown. DNA fragments A, B and C were generated by digesting the promoter with the indicated restriction enzymes and gel purified. (**B**) Increasing concentrations (shown as molar excesses) of double-stranded GC box consensus (GC Box Cons. Oligo) and mutated (GC Box Mut. Oligo) oligonucleotides were incubated with 4µg of RCS nuclear extracts for 10 min prior to the addition of ^{32}P-labeled probe A, B or C. The DNA–protein complexes were analyzed on 5% non-denaturing polyacrylamide gels. The positions of the complexes are indicated with arrowheads.

Transfac (release 2.2) (30). Four putative GC boxes, the proximal two in the positive orientation and the distal two inverted, were identified within the promoter between –146 and –58 nucleotides upstream from the translation start site (Fig. 1). The three downstream GC boxes all contained two base mismatches when compared with the Sp1 decanucleotide consensus sequence, 5′-KGGGCGGRRY-3′ (31). The fourth upstream GC box, on the other hand, contains only a single base mismatch, making it more similar to the Sp1 consensus sequence. A fifth potential Sp1-binding motif, an inverted GT box (32), was also identified between nucleotides –152 and –161. This element also contains two base mismatches from the published sequence. A potential inverted metal responsive element (MRE) was identified downstream from the Sp1 elements between nucleotides –54 and –46 (33). Furthermore, a C/EBP (–354 to –346), an inverted NF-κB (–252 to –244) and an AP-2 (–191 to –184) consensus sequence (34–36) were identified upstream from the five Sp1-binding motifs.

Binding of nuclear proteins to the metaxin proximal promoter *in vitro*

Since transient transfection experiments showed that the –138 metaxin promoter–luciferase construct, which contained three out of the four GC boxes, had over 70% of the activity of the larger promoter constructs (Fig. 3), we decided to test whether GC box-binding proteins were the main factors responsible for activating the metaxin gene. Accordingly, the 192 bp *Bsu*36I–*Rsa*I, 133 bp *Sac*II–*Rsa*I and 74 bp *Bgl*I–*Rsa*I fragments of the metaxin proximal promoter were radiolabeled, as described in Materials and Methods, and used as probes A, B and C, respectively, in electrophoretic mobility shift assays (EMSA) (Fig. 4). Probes A, B and C contained four, three and one intact GC boxes, respectively. Probe A also contained the GT box and the AP-2 consensus sequence. All three probes contained the MRE. Similar DNA–protein complexes were observed when RCS or NIH-3T3 nuclear extracts were assayed with these probes (data not shown). Two complexes, a major complex and a faster migrating minor complex, were identified (indicated by arrowheads) when RCS nuclear extracts were assayed with probes B or C (Fig. 4B). In addition to these two complexes, a second more slowly migrating minor complex was identified when nuclear extracts were assayed with probe A. The additional faster migrating minor complexes shown in the figures were not reproducible.

We next tested the ability of increasing molar excesses of unlabeled double-stranded GC box consensus and mutant oligonucleotides to compete with ^{32}P-labeled probes A, B or C for complex formation with RCS nuclear extracts in EMSA (Fig. 4B). Figure 4B shows that the GC box consensus sequence specifically inhibited the formation of the major and two minor complexes when RCS nuclear extracts were assayed with probe A. Similarly, the GC box consensus oligonucleotide inhibited the formation of both the major and minor complexes when nuclear extracts were assayed with probes B or C. No detectable

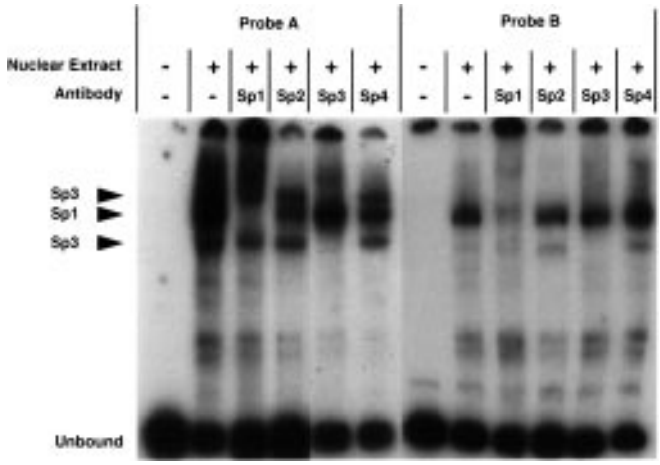

Figure 5. Supershift assays with anti-Sp1 and related antibodies. Four µg of RCS nuclear extract was incubated with anti-Sp1, Sp2, Sp3 or Sp4 and ^{32}P-labeled probe A or B. The DNA–protein complexes were electrophoresed, together with free probe, on 5% non-denaturing polyacrylamide gels at 4°C in 0.5× TBE. The position of the complexes formed by Sp1 and Sp3 are indicated by arrowheads.

competition was observed when the mutant GC box oligonucleotide was used in the assays. The GC box consensus oligonucleotide also specifically inhibited the formation of the major and minor complexes when NIH-3T3 nuclear extracts were assayed with probes A and B (data not shown). These findings suggest that *trans*-acting factors present in RCS and NIH-3T3 nuclear extracts bind to one or more of the GC boxes within the metaxin proximal promoter.

Identification of DNA–protein complexes with antibodies to Sp1 and Sp1-related proteins

To date, at least three factors, Sp1 and two related proteins Sp3 and Sp4, have been shown to bind GC boxes with similar specificities and affinities (37,38). All three factors are also able to bind to the GT box. A fourth member of the Sp transcription factor family, Sp2, is only able to interact with the GT box (38). To distinguish which member(s) of this transcription factor family binds to the metaxin promoter, anti-Sp1, Sp2, Sp3 and Sp4 polyclonal antibodies were used in supershift assays. As shown in Figure 5, anti-Sp1 supershifted the major complex when RCS nuclear extracts were assayed with probes A and B. Prebinding of the antibody with a control peptide containing the anti-Sp1 epitope prevented the major complex from supershifting in EMSAs (data not shown). The anti-Sp3 antibody, on the other hand, resulted in the supershift of the minor band(s) when nuclear extracts were assayed with probe A or B. However, there was no detectable shift in any of the complexes when anti-Sp2 or Sp4 was assayed, suggesting that mainly Sp1 and, to a lesser extent the transcriptional repressor, Sp3, are responsible for the formation of the observed DNA–protein complexes. Similar results were obtained when NIH-3T3 nuclear extracts were assayed with the antibodies (data not shown).

Sp1 activates the metaxin promoter

The functional studies in Figure 3 clearly suggest that more than one of the Sp1-binding motifs are involved in regulating the metaxin gene, and binding experiments with purified protein showed that four out of the five Sp1-binding motifs were capable of binding Sp1 *in vitro* (data not shown). Schneider *Drosophila* line 2 (SL2) cells, which do not contain endogenous Sp1, were therefore transfected with metaxin-promoter luciferase constructs in the presence or absence of an insect cell Sp1 expression vector. As shown in Figure 6A, there was a marked increase in luciferase activity when SL2 cells were co-transfected with metaxin–promoter constructs and the Sp1 expression vector, as compared with the activity in cells transfected with the metaxin constructs alone. The fold increase in activities from the –83 to the –138, from the –138 to the –377, and from the –377 to the –1661 promoter fragments were, however, not similar to that seen in RCS and NIH-3T3 cells. The increase in luciferase activity in the insect cells appeared to be proportional to the number of Sp1-binding elements present in the *Mtx* promoter constructs (Fig. 1). In addition, an insect-specific transcriptional activator(s) could activate the *Mtx* promoter constructs in insect cells, or the cells could lack a transcriptional repressor(s). An insect-specific DNA–protein complex was observed when SL2 nuclear extracts were assayed with probe A (data not shown). In any event, the data strongly suggest that Sp1 is a transcriptional activator of the metaxin gene.

To test whether the Sp1-binding elements were critical for *Mtx* promoter activity, internal deletion mutants of the –1661 (*Bam*HI) promoter were generated and assayed (Fig. 6B). As shown in Figure 6B, there was no significant decrease in luciferase activity when the –1661 (Δ*Bsu*36I–*Sac*II) promoter construct, in which the distal GC box and GT box had been deleted, was assayed in RCS and NIH-3T3 cells. There was however, only a slight decrease in luciferase activity when the –1661 (Δ*Bsu*36I–*Ava*II) promoter construct, in which the four upstream Sp1-binding motifs were deleted, was assayed in both RCS and NIH-3T3 cells, suggesting that factors which bind upstream from the Sp1-binding elements are able to activate the *Mtx* gene by compensating for the deletion of the GC and GT boxes. To test this hypothesis, –377 (Δ*Bsu*36I–*Sac*II) and –377 (Δ*Bsu*36I–*Ava*II) metaxin promoter constructs were assayed. Except for a C/EBP and an NF-κB consensus sequence (Fig. 1), all of the other potential upstream *cis*-acting elements were deleted in these constructs. As shown in Figure 6B, there was no significant decrease in luciferase activity when the –377 and the –377 (Δ*Bsu*36I–*Sac*II) *Mtx* promoter constructs were assayed in both RCS and NIH-3T3 cells, suggesting that the distal GT and GC boxes are not important for *Mtx* gene regulation. This finding is consistent with the data shown in Figure 3, where there was also no significant difference in activity between the –377 and the –138 *Mtx* promoter constructs, which contained five and three Sp1-binding elements, respectively. There was however a significant decrease in luciferase activity when the –377 (Δ*Bsu*36I–*Ava*II) promoter construct, containing only one proximal GC box and the C/EBP consensus sequence, was assayed in RCS ($P < 0.001$) and NIH-3T3 ($P = 0.013$) cells. The activity of the –83 bp promoter construct, which contains only one intact GC box, was about half that of the –377 (Δ*Bsu*36I–*Ava*II) construct. This slight increase in activity with the –377 (Δ*Bsu*36I–*Ava*II) promoter construct could be due to a cooperative interaction between Sp1 and C/EBP (39). Taken together these data suggest that the proximal three GC boxes are involved, but are not critical, in regulating the metaxin gene. Factors which bind upstream from the minimal *Mtx* promoter, but which normally may not have a major effect on *Mtx*

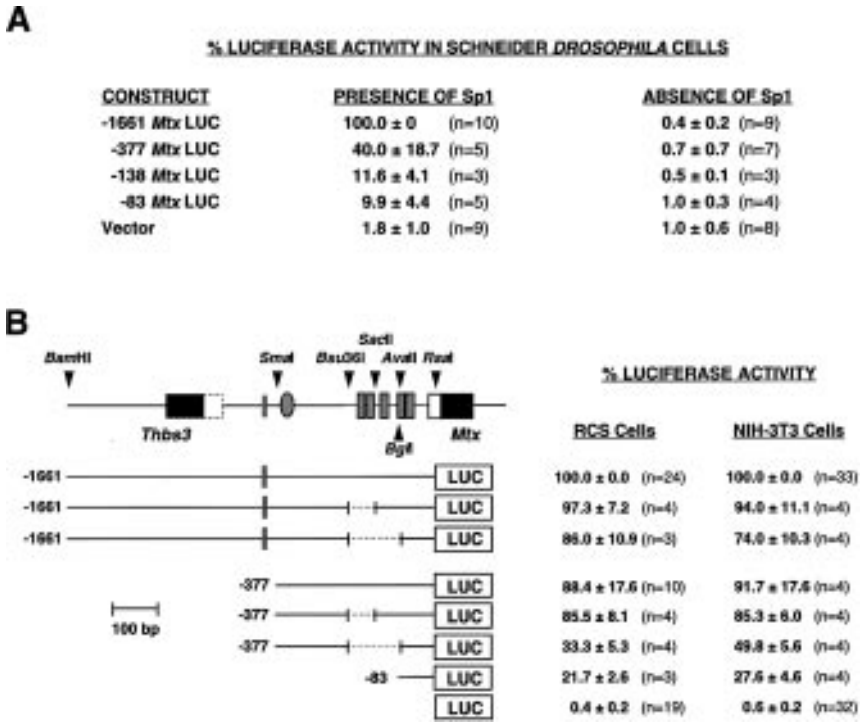

Figure 6. (**A**) Analysis of the metaxin promoter in SL2 cells. SL2 cells were co-transfected with the indicated metaxin promoter–luciferase constructs together with the β-galactosidase gene and either an Sp1 expression or control vector. The luciferase activity, expressed relative to the activity of the –1661 bp metaxin promoter in the presence of Sp1, and standard deviations for each promoter construct, were determined. The number of determinations (n) are also indicated. (**B**) Deletion of the Sp1-binding elements from the –1.7 kb metaxin promoter. A schematic diagram of the mouse thrombospondin 3–metaxin intergenic region and the 5′-portions of the two genes is shown. The translated sequences of metaxin exon 1 and thrombospondin 3 exon A (filled boxes), the 5′-UTRs (open boxes), potential Sp1-binding elements (hatched boxes), and a C/EBP consensus sequence (hatched ellipse) are also shown. The thrombospondin 5′-untranslated region is shown as a dashed box because the transcription start site has not yet been determined. The indicated restriction enzymes were used to clone several internal deletions of the –1661 bp and –377 bp metaxin promoter into pGL2-Basic. An *Rsa*I site situated within the short 5′-UTR of metaxin served as the 3′-boundary of the metaxin promoter. The metaxin promoter–luciferase constructs were co-transfected, with SV40–β-gal, into RCS and NIH-3T3 cells, as described in Materials and Methods. The percentage luciferase activity, expressed relative to the value for the –1661 bp construct, standard deviations for each promoter construct, and number of determinations (n) are indicated.

gene expression (Fig. 3), can compensate for deletions of any or all of the four upstream Sp1 motifs.

The metaxin minimal promoter does not regulate the thrombospondin 3 gene

In order to test whether the Sp1-binding elements within the minimal metaxin promoter were also able to regulate the thrombospondin 3 gene, various thrombospondin 3 promoter–luciferase reporter gene constructs were tested in RCS and NIH-3T3 cells (Fig. 7). The largest construct, –2034 *Thbs3* promoter–luciferase, contains the entire intergenic region, exon 1 of metaxin and the 5′-part of the metaxin first intron. The second construct, –1348 *Thbs3* promoter–luciferase, contains the entire metaxin thrombospondin 3 intergenic region, which includes the minimal metaxin promoter. In the –1150 bp *Thbs3* promoter–luciferase construct, the Sp1-binding elements in the metaxin minimal promoter were deleted. As shown in Figure 7, there was a minor, but significant, decrease in luciferase activity when the –1348 and –1150 *Thbs3* promoter–luciferase constructs were assayed in both RCS ($P = 0.035$) and NIH-3T3 ($P = 0.002$) cells, suggesting that the minimal metaxin promoter does not, at least in these cell lines, play a major role in regulating the expression of the thrombospondin 3 gene. Further deletions of the –1150 *Thbs3* promoter–luciferase construct, resulted in a gradual decrease in luciferase activity in both cell lines. These experiments also showed that the –2034 thrombospondin 3 promoter construct was only about a sixtieth or a tenth as active as the –1661 metaxin promoter in RCS and NIH-3T3 cells, respectively. It is possible that an as yet unidentified enhancer element(s) situated further upstream, within the body of the metaxin gene, or downstream within the body of the thrombospondin 3 gene, contributes to the expression of the *Thbs3* gene. A potential enhancer element for the human *THBS3* gene has recently been identified within the body of the human *MTX* gene (J. Silver, M. Collins, and P. Bornstein, unpublished data).

DISCUSSION

The mouse metaxin and thrombospondin 3 genes are arranged in a head-to-head orientation on chromosome 3E3-F1 so that their putative translation start sites are separated by 1352 nucleotides (2). Since these divergently transcribed genes are regulated by a common promoter region, the metaxin minimal promoter was characterized and its effect on *Thbs3* gene expression was studied. Transient transfection experiments in *Mtx*-expressing rat chondrosarcoma cells and NIH-3T3 fibroblasts demonstrated that the elements required for expression of the *Mtx* gene are situated within the short G+C rich proximal promoter region containing five clustered Sp1-binding motifs, but lacking a CCAAT or a TATA box. Electrophoretic mobility shift, competition and

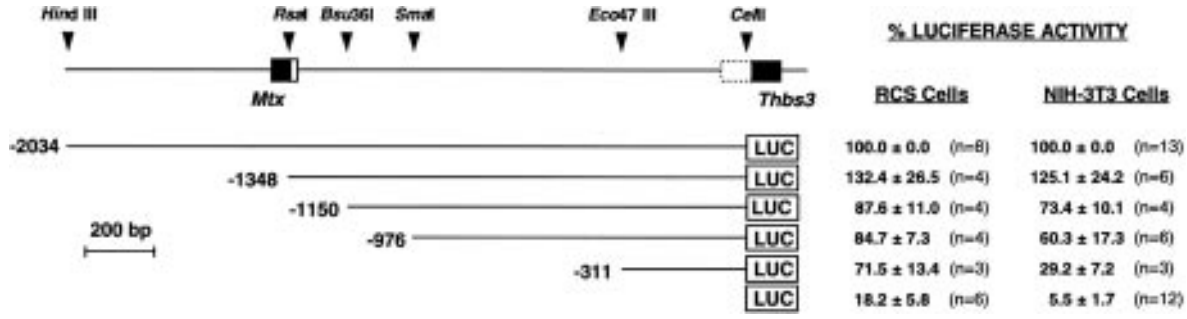

Figure 7. Deletion analysis of the mouse thrombospondin 3 promoter in RCS and NIH-3T3 cells. A schematic diagram of the mouse metaxin–thrombospondin 3 intergenic region and the 5′-portions of the two genes is shown. The translated sequences of metaxin exon 1 and thrombospondin 3 exon A (filled boxes), and the 5′-UTRs (open boxes) are also shown. The 5′-UTR of thrombospondin 3 is shown with a dashed box because the transcription start site is uncertain. A*Cel*II site situated immediately upstream from the thrombospondin 3 translation start served as the 3′-boundary of the promoter. The thrombospondin 3 promoter–luciferase constructs were co-transfected with SV40–β-gal into RCS and NIH-3T3 cells, as described in Materials and Methods. The percent luciferase activity, expressed relative to the value for the –2034 bp thrombospondin 3 promoter construct, standard deviations for each promoter construct, and number of determinations (n) are indicated.

supershift assays demonstrated that Sp1, and to a lesser extent the Sp1-related protein, Sp3, bind to these elements *in vitro*. Functional studies in SL2 insect cells, co-transfected with metaxin promoter constructs in the presence or absence of an Sp1 expression vector, confirmed that Sp1 activates the *Mtx* gene. The four upstream Sp1-binding motifs were also shown to be superfluous when deleted from the –1.7 kb promoter. Additional elements probably compensate for these deletions.

Sp1 is a ubiquitous transcription factor which is responsible for activating many cellular and viral genes (31). Like metaxin, many eukaryotic genes with TATA-less promoters contain multiple Sp1-binding motifs arranged in tandem and in close proximity within the proximal promoter (40). Sp1 can activate many of these promoters synergistically by Sp1–Sp1 interactions and these interactions are an important mechanism for modulating the expression of this class of genes (40,41). These genes are further regulated by Sp1-related factors, such as Sp3. Hagen *et al.* (32) have demonstrated that the ubiquitous transcription factor, Sp3, represses Sp1-mediated transcriptional activation in a linear dose-dependent manner.

Besides the GC and GT boxes, an inverted NF-κB and an AP-2 consensus sequence were identified upstream from the Sp1-binding motifs within the –377 bp metaxin minimal promoter, between nucleotides –252 and –244 and between –191 and –184, respectively. Both NF-κB and AP-2 are inducible transcription factors (35,36), suggesting that metaxin gene expression could be regulated in response to external stimuli. A highly specific cooperative interaction between NF-κB and Sp1 is responsible for the induction of the HIV-1 long terminal repeat expression (42,43). Since the NF-κB element is located upstream from the cluster of Sp1-binding elements within the metaxin proximal promoter, it is possible that a similar mechanism may serve to regulate the metaxin gene in response to external stimuli. A cooperative interaction between Sp1 and C/EBP has also been previously described (39). Since, the *Mtx* minimal promoter also contains a C/EBP element, it is possible that C/EBP may also play a role in regulating *Mtx* gene expression. Interestingly, the clustered Sp1-binding motifs as well as the NF-kB, AP-2 and C/EBP elements are all conserved within the human *MTX* proximal promoter (27), strengthening the hypothesis that these elements modulate metaxin gene expression.

The transcription start site of the metaxin gene was shown to be heterogenous. The major start site is only 13 bp upstream from the putative translation start site, and the longest and shortest 5′-UTR is 26 and 7 bp, respectively. The 5′-UTR of metaxin is therefore shorter than the average length of between 20 and 100 nucleotides for most eukaryotic mRNAs (44). Investigators have shown that the accuracy of translation from an initiation AUG codon can decrease when it is very close to the transcription start site. Initiation at a downstream AUG can also occur with increased efficiency under these circumstances, even when the upstream AUG is in a better sequence context (45,46). As shown in Figure 2, there is an in-frame AUG situated 12 bp downstream from the putative metaxin translation start site. This downstream AUG is also conserved within the human gene (27). In both species the upstream AUG is situated within an ideal sequence context, 5′-A^{-3}AG<u>A</u>$^{+1}$<u>UG</u>G^{+4}–3′, for translation initiation, while the downstream AUG is in a less favorable one, 5′-C^{-3}C<u>CA</u>$^{+1}$<u>UG</u>G^{+4}-3′ (47,48). Although the upstream AUG was initially defined as the translation start site for both the mouse and the human metaxin genes (2,27), it is possible that ribosomes use the downstream AUG as the initiation codon or, alternatively, two forms of metaxin can result from alternative initiation at the two in-frame AUGs. Strubin *et al.* (49) have demonstrated that two major forms of the human Ia antigen-associated invariant chain are produced by alternative initiation from two in-frame AUG codons, situated 8 and 56 bases from the transcription start site. Four overlapping initiator-like elements, each containing 2-base mismatches from the published consensus sequence, 5′-YYANWYY-3′ (28), were identified around the cap sites of the mouse metaxin gene. This observation could explain why metaxin transcription is initiated heterogeneously.

Previous studies have demonstrated that metaxin is a ubiquitously expressed gene which encodes a mitochondrial protein and is essential for embryogenesis (2,10). Unlike metaxin, thrombospondin 3 has a much limited tissue distribution. The *Thbs3* gene is expressed predominately in the lung, in the hippocampus of the brain, in cartilage, and in the gastrointestinal tract, with lower levels of expression in other extracellular matrix-producing tissues (5,8,9). Thrombospondin 3, like other members of the thrombospondin protein family, probably plays a role in modulating the function of the extracellular matrix. Although these two divergent genes encode proteins with apparently different

functions, it was still possible that common elements regulated *Mtx* and *Thbs3*. Thus, the divergent human dihydrofolate reductase and mismatch protein 1 genes encode proteins which have no apparent relationship, but are nevertheless regulated by a bi-directional promoter (13). The –2 kb *Thbs3* promoter was shown to be relatively weak in both rat chondrosarcoma (RCS) and NIH-3T3 cells. The promoter was about a sixtieth or a tenth as active as the metaxin promoter in RCS and NIH-3T3 cells, respectively. There was a minor, but significant, difference in promoter activity when the –1.3 kb and the –1.1 kb *Thbs3* promoter, in which the metaxin minimal promoter had been deleted, was assayed in both cell lines. These findings suggest that the minimal metaxin promoter only plays a minor role in modulating the expression of the thrombospondin 3 gene in RCS or NIH-3T3 cells. Since the thrombospondin 3 promoter was weak in both cell lines, it is possible that an as-yet-unidentified enhancer element(s) may contribute to the expression of the *Thbs3* gene in cells, such as RCS cells, in which this gene is expressed. If this is the case, then the effects of the *Mtx* minimal promoter on *Thbs3* gene expression would probably be even less significant.

ACKNOWLEDGEMENTS

This work was supported in part by grant DE 08229 from the National Institutes of Health. M.C. was supported by a Fogarty International Research Fellowship (1 F05 TW05218-01), and a Foundation for Research Development Postdoctoral Fellowship from South Africa. We thank Lynn Law for assistance with the insect cell expression system. We also thank members of our laboratory for helpful comments on the manuscript.

REFERENCES

1 Bornstein,P., Devarayalu,S., Edelhoff,S. and Disteche,C.M. (1993) *Genomics*, **15**, 607–613.
2 Bornstein,P., McKinney,C.E., LaMarca,M.E., Winfield,S., Shingu,T., Devarayalu,S., Vos,H.L. and Ginns,E.I. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 4547–4551.
3 Spicer,A.P., Parry,G., Patton,S. and Gendler,S.J. (1991) *J. Biol. Chem.*, **266**, 15099–15109.
4 Vos,H.L., de Vries,Y. and Hilkens,J. (1991) *Biochem. Biophys. Res. Commun.*, **181**, 121–130.
5 Vos,H.L., Devarayalu,S., de Vries,Y. and Bornstein,P. (1992) *J. Biol. Chem.*, **267**, 12192–12196.
6 Beutler,E. (1991) *N. Engl. J. Med.*, **325**, 1354–1360.
7 Bornstein,P. (1992) *FASEB*, **6**, 3290–3299.
8 Iruela-Arispe,M.L., Liska,D., Sage,E.H. and Bornstein,P. (1993) *Dev. Dynamics*, **197**, 40–56.
9 Qabar,A.N., LIn,Z., Wolf,F.W., O'Shea,K.S., Lawler,J. and Dixit,V.M. (1994) *J. Biol. Chem.*, **269**, 1262–1269.
10 Armstrong,L.C., Vos,H., Bergman,B. and Bornstein,P. (1995) *Mol. Biol. Cell*, **6**, 434a.
11 Burbelo,P.D., Martin,G.R. and Yamada,Y. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 9679–9682.
12 Poschl,E., Pollner,R. and Kuhn,K. (1988) *EMBO J.*, **7**, 2687–2695.
13 Fujii,H., Shinya,E. and Shimada,T. (1992) *FEBS Lett.*, **314**, 33–36.
14 Gavalas,A., Dixon,J.E., Brayton,K.A. and Zalkin,H. (1993) *Mol. Cell. Biol.*, **13**, 4784–4792.
15 O'Hanlon,T.P., Raben,N. and Miller,F.W. (1995) *Biochem. Biophys. Res. Commun.*, **210**, 556–566.
16 Osley,M.A., Gould,J., Kim,S., Kane,M. and Hereford,L. (1986) *Cell*, **45**, 537–544.
17 Johnston,M. and Davis,R.W. (1984) *Mol. Cell. Biol.*, **4**, 1440–1448.
18 Garabedian,M.J., Hung,M. and Wensink,P.C. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 1396–1400.
19 Choi,H.U., Meyer,K. and Swarm,R. (1971) *Proc. Natl. Acad. Sci. USA*, **68**, 877–879.
20 Mukhopadhyay,K., Lefebvre,V., Zhou,G., Garofalo,S., Kimura,J.H. and de Crombrugghe,B. (1995) *J. Biol. Chem.*, **270**, 27711–27719.
21 Rosenthal,N. (1987) *Meth. Enzymol.*, **152**, 704–720.
22 Di Nocera,P.P. and Dawid,I.B. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7095–7098.
23 Courey,A.J. and Tjian,R. (1988) *Cell*, **55**, 887–898.
24 Contursi,C., Minchiotti,G. and Di Nocera,P.P. (1995) *J. Biol. Chem.*, **270**, 26570–26576.
25 Lee,K.A.W. and Green,M.R. (1990) *Meth. Enzymol.*, **181**, 20–30.
26 Bradford,M.M. (1976) *Anal. Biochem.*, **72**, 248–254.
27 Long,G.L., Winfield,S., Adolph,K.W., Ginns,E.I. and Bornstein,P. (1996) *Genomics*, **33**, 177–184.
28 Javahery,R., Khachi,A., Lo,K., Zenzie-Gregory,B. and Smale,S.T. (1994) *Mol. Cell. Biol.*, **14**, 116–127.
29 Ghosh,D. (1992) *Nucleic Acids Res.*, **20**, 2091–2093.
30 Wingender,E. (1994) *J. Biotechnol.*, **35**, 273–280.
31 Kadonaga,J.T., Jones,K.A. and Tjian,R. (1986) *Trends Biochem. Sci.*, **11**, 20–23.
32 Hagen,G., Muller,S., Beato,M. and Suske,G. (1994) *EMBO J.*, **13**, 3843–3851.
33 Andersen,R.D., Taplitz,S.J., Wong,S., Bristol,G., Larkin,B. and Herschman,H.R. (1987) *Mol. Cell. Biol.*, **7**, 3574–3581.
34 Yuh,C. and Ting,L. (1991) *Mol. Cell. Biol.*, **11**, 5044–5052.
35 Lenardo,M.J. and Baltimore,D. (1989) *Cell*, **58**, 227–229.
36 Imagawa,M., Chiu,R. and Karin,M. (1987) *Cell*, **51**, 251–260.
37 Hagen,G., Muller,S., Beato,M. and Suske,G. (1992) *Nucleic Acids Res.*, **20**, 5519–5525.
38 Kingsley,C. and Winoto,A. (1992) *Mol. Cell. Biol.*, **12**, 4251–4261.
39 Lee,Y., Yano,M., Liu,S., Matsunaga,E., Johnson,P.F. and Gonzalez,F.J. (1994) *Mol. Cell. Biol.*, **14**, 1383–1394.
40 Anderson,G.M. and Freytag,S.O. (1991) *Mol. Cell. Biol.*, **11**, 1935–1943.
41 Courey,A.J., Holtzman,D.A., Jackson,S.P. and Tjian,R. (1989) *Cell*, **59**, 827–836.
42 Perkins,N.D., Edwards,N.L., Duckett,C.S., Agranoff,A.B., Schmid,R.M. and Nabel,G.J. (1993) *EMBO J.*, **12**, 3551–3558.
43 Majello,B., De Luca,P., Hagen,G., Suske,G. and Lania,L. (1994) *Nucleic Acids Res.*, **22**, 4914–4921.
44 Kozak,M. (1987) *Nucleic Acids Res.*, **15**, 8125–8148.
45 Sedman,S.A., Gelembiuk,G.W. and Mertz,J.E. (1990) *J. Virol.*, **64**, 453–457.
46 Kozak,M. (1991) *Gene Exp.*, **1**, 111–115.
47 Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
48 Kozak,M. (1986) *Cell*, **44**, 283–292.
49 Strubin,M., Long,E.O. and Mach,B. (1986) *Cell*, **47**, 619–625.