Multiple Pathways for Steel Regulation Suggested by Genomic and Sequence Analysis of the Murine Steel Gene

Mary A. Bedell, Neal G. Copeland and Nancy A. Jenkins
Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

Manuscript received November 14, 1995
Accepted for publication December 4, 1995

ABSTRACT

The Steel (Sl) locus encodes mast cell growth factor (Mgf) that is required for the development of germ cells, hematopoietic cells and melanocytes. Although the expression patterns of the Mgf gene are well characterized, little is known of the factors which regulate its expression. Here, we describe the cloning and sequence of the full-length transcription unit and the 5' flanking region of the murine Mgf gene. The full-length Mgf cDNA consists of a short 5' untranslated region (UTR), a 0.8-kb ORF and a long 3' UTR. A single transcription initiation site is used in a number of mouse tissues and is located just downstream of binding sites for several known transcription factors. In the 5' UTR, two ATGs were found upstream of the initiator methionine and are conserved among different species, suggesting that Mgf may be translationally regulated. At least two Mgf mRNAs are produced by alternative use of polyadenylation sites, but numerous other potential polyadenylation sites were found in the 3' UTR. In addition, the 3' UTR contains numerous sequence motifs that may regulate Mgf mRNA stability. These studies suggest multiple ways in which expression of Mgf may be regulated.

Mutations at the Dominant White Spotting (W) and Steel (Sl) loci identify two genes essential for the development of several cell lineages: neural crest-derived melanocytes, germ cells, hematopoietic stem cells, mast cells and erythroid cells (reviewed by Silvers 1979a,b). The products of both loci have been cloned and together comprise a pathway for intercellular communication for these diverse cell populations. The product of the W locus is a receptor tyrosine kinase, called Kit, that is closely related to the platelet-derived growth factor receptor family (Chabot et al. 1988; Geissler et al. 1988). The Sl locus encodes mast cell growth factor [Mgf, also known as Kit ligand (KL), stem cell factor (SCF) and Steel factor (SLF)], the ligand for Kit (Copeland et al. 1990; Flanagan and Leder 1990; Huang et al. 1990; Williams et al. 1990; Zsebo et al. 1990).

While Kit is expressed on the surface of cells affected by W mutations (Nocka et al. 1989; Manova et al. 1990; Orr-Urtreger 1990), Mgf is expressed by stromal cells that support the growth and differentiation of the affected cells (Matsui et al. 1996; Keshet et al. 1991). This receptor-ligand interaction is thought to be required for guiding the migration and survival of certain migratory cell populations during embryogenesis. For example, Kit is expressed by primordial germ cells and neural crest-derived melanocytes that populate the genital ridges and skin, respectively, while Mgf is abundantly expressed in mesodermal cells along the pathway of migration and at the final destination of these cells. Several lines of evidence support the notion that the level of Mgf expression in both embryonic and adult tissues is critical to its function. All Sl mutant alleles exert semidominant phenotypes that have been attributed to haploinsufficiency (Silvers 1979a). In the postnatal ovary, a threshold level of Mgf is thought to be required for initiation and maintenance of follicle growth (Huang et al. 1993; Bedell et al. 1995). In addition, expression of Mgf mRNA in the mature ovarian follicles may be regulated by hormones of the estrous cycle (Motro and Bernstein 1993). Surprisingly, both Mgf and Kit are expressed in tissues not known to be affected in Sl or W mutants. In particular, these genes are expressed in a complementary pattern in the central nervous system of both embryos and adults (Matsui et al. 1990; Keshet et al. 1991; Motro et al. 1991), suggesting that this signaling pathway may be involved in development and/or function of the nervous system. Although the spatial and temporal patterns of expression of these genes are now well documented, little is known of the elements that control expression of either gene.

Mgf cDNAs that range in size from 1 to 2 kb have been isolated from a variety of species, including mouse (Anderson et al. 1990; Huang et al. 1990; Zsebo et al. 1990), human and rat (Martin et al. 1990), pig (Zhang and Anthony 1994), and chicken (Zhou et al. 1995). The largest of these clones is a murine Mgf cDNA (Mgf10, see Figure 1; Anderson et al. 1990) that contains a 0.8-kb open reading frame (ORF) with 1 kb of...
3' untranslated region (UTR). However, the major Mg transcript seen on Northern blots from tissues of all species is much larger and has been estimated to be 5.5–6.5 kb with mouse and chicken displaying minor transcripts that ranged in size from 3–4.6 kb (Anderson et al. 1990, 1991; Huang et al. 1990; Martin et al. 1990; Zsebo et al. 1990; Zhou et al. 1993; Zhang and Anthony 1994). These results indicate that a considerable portion of the Mg mRNA was not cloned in cDNA form. Furthermore, it was not known whether the Mg transcripts of various size were produced by alternative sites of transcription initiation or termination or by alternative splicing.

Here we describe the full length Mg transcript and 5' flanking sequence. These studies identify multiple pathways that may be involved in Sf gene regulation. In a companion paper, we use this cDNA and genomic clones for deletion analysis of various homozygous lethal Sf alleles.

MATERIALS AND METHODS

Libraries, clones and sequencing: A genomic library of C57BL/6J DNA was prepared by partial Sau3A digestion and insertion into Adashe vector (Stratagene, Inc., La Jolla, CA). An oligo-dT primed cDNA library of C57BL/6J brain RNA cloned into x Zap was purchased from Stratagene (La Jolla, CA). Portions of MGFl0 (Anderson et al. 1990; see Figure 1) were used to screen the genomic and cDNA libraries. Positive clones were isolated, characterized and subcloned into pSkl plasmid (Stratagene) using conventional methods. Sequencing was performed using either a dideoxy method (United States Biochemical, Cleveland, OH) or an automated DNA sequencing system (Applied Biosystems, Foster City, CA). Computer analysis of the nucleotide sequence was performed using the Wisconsin Sequence Analysis package (Genetics Computer Group, Madison, WI).

Analysis of Mg mRNA: Total RNA from various tissues of wild-type adult mice was prepared using RNAzol (Tel-Test, Inc., Friendswood, TX). PolyA+ RNA was prepared using a mRNA Purification Kit (Pharmacia Biotech, Piscataway, NJ). Northern blot analysis was performed as described previously (Bedell et al. 1995) using probes shown in Figure 1. The 5' ends of Mag mRNA were identified by a RACE technique using an Amplifinder kit (Clontech Laboratories, Inc., Palo Alto, CA). The position of RNA size markers are indicated at the left of the panel. Note that while MGFlO detects both the abundant 5.4 kb and minor 3.7 kb bands, only the former was screened with a 230-nt fragment (5') while the latter was screened with a 280-nt fragment (3'). The four cDNAs isolated using the 3' probe are shown below MGFlO. PolyA tract (An). (B) Northern blot analysis of Mg mRNA. PolyA+ mRNA of adult mouse brain (lane 1) and kidney (lane 2) was electrophoresed, blotted and hybridized with MGFlO, a DNA probe that contains the ORF and a portion of the 3' UTR (left) or ScPa, a probe containing a distal portion of the 3' UTR from pc1.2 (right). The position of RNA size markers is indicated to the left of the panel. Note that while MGFlO detects both the abundant 5.4 kb and minor 3.7 kb bands, only the former is detected by the ScPa probe.

RESULTS

Isolation of full-length Mg cDNA clones: To obtain cDNA clones representing full-length murine Mg transcripts, a wild-type mouse brain cDNA library was probed with a portion of the 3' UTR of MGFl0 (Anderson et al. 1990; see Figure 1A). Four overlapping cDNAs (pC4.1, pC16.1, pC7.1 and pC1.2) were isolated, sequenced and aligned with the known Mg cDNA sequence. A schematic of the full-length Mg mRNA, representing the composite of both 5' and 3' flanking sequences (see below), is shown at the top of Figure 1 with the relative positions of the four newly isolated cDNAs shown below. The sequence of the full-length Mg mRNA is shown in Figure 2. All four cDNAs terminate in polyA tracts. Three cDNAs (pC16.1, pC7.1 and pC1.2) have the same 3' ends but vary in the extent of 5' sequences, while the fourth cDNA (pC4.1) terminates 1.7 kb upstream of the ends of the other cDNAs.

Figure 1.—(A) Alignment of Mg cDNAs to full-length Mg mRNA. The composite, full-length Mg mRNA of 5.4 kb (see text) is shown at the top with the 813-nt coding region (□) and 5' and 3' UTRs of 197 and 4432 nt, respectively (— — — ). Fragments of MGFl0, a 2-ka DNA containing the ORF and ~1 kb of 3' UTR (Anderson et al. 1990) were used to screen genomic and cDNA libraries; the former was screened with a 230-nt fragment (5') while the latter was screened with a 280-nt fragment (3'). The four cDNAs isolated using the 3' probe are shown below MGFlO. PolyA tract (An). (B) Northern blot analysis of Mg mRNA. PolyA+ mRNA of adult mouse brain (lane 1) and kidney (lane 2) was electrophoresed, blotted and hybridized with MGFlO, a DNA probe that contains the ORF and a portion of the 3' UTR (left) or ScPa, a probe containing a distal portion of the 3' UTR from pc1.2 (right). The position of RNA size markers is indicated to the left of the panel. Note that while MGFlO detects both the abundant 5.4 kb and minor 3.7 kb bands, only the former is detected by the ScPa probe.
(Figure 1A). The 3' UTR of the latter clone is 2732 nt in length while all other cDNAs contain 3' UTRs of 4432 nt. None of the four cDNA clones contained novel 5' UTR or coding sequences compared with that of Mgf cDNAs described previously.

Restriction enzyme mapping, hybridization analysis and partial sequencing of a genomic clone encompassing the 3' flanking region demonstrated that the 3' termini of the cDNAs represent authentic polyadenylation sites. Alignment of the cDNAs to the composite Mgf mRNA (Figure 1A) indicates that the shorter 3' UTR may represent the 3' terminus of a small Mgf transcript while the longer 3' UTR is included in a larger Mgf transcript. To test whether this alignment is correct, Northern blots of adult mouse tissues were hybridized with either MGF10 or with a probe from the 3' terminus of pC1.2 (see Figure 1B). While MGF10 detects two Mgf transcripts of ~5.5 kb and ~4 kb, only the former transcript is detected by the 3' UTR sequence from pC1.2 (Figure 1B). These results indicate that the two Mgf mRNAs are produced by alternative usage of polyadenylation sites. Other bands of minor abundance are apparent with both probes and may reflect either additional, alternative Mgf mRNAs or cross-hybridization to related sequences.

To define the 5' end(s) of the Mgf mRNA, RACE was performed on cDNA using nested primers located in the 5' end of the Mgf ORF. The Mgf primers chosen were downstream of the first four introns of Mgf mRNA (Figure 2) and should allow any alternative splice products in this region to be detected. 5' RACE products of RNA from six tissues (brain, lung, kidney, heart, ovary and testes) of adult mice were analyzed. After one cycle of amplification using the RACE anchor primer and oligonucleotides B, C or D (Figure 2), the major band produced by each primer pair was identical in all six tissues (not shown). That these bands represent authentic Mgf sequences was confirmed by hybridization to an Mgf cDNA probe. Other, minor bands were observed but were not consistently produced during repetitive experiments. The major band from each tissue was then excised, purified and subjected to a second round of amplification using either oligonucleotide C or D and the PCR products derived from each tissue were directly sequenced. All of the RACE products were colinear except for the last three to four nucleotides at the 5' end. However, the sequence in this region was too heterogeneous to define a specific start site. To circumvent this, the RACE products from all six tissues were subcloned into plasmid and five or six individual clones from each tissue were isolated and sequenced. Of 35 different RACE clones that were sequenced, all initiated at one of three Gs at position 1, 2 or 4 shown in Figure 2. The majority of the 5' ends were at position 1 (23/35; 66%), with minor start sites at position 2 (7/35; 20%) and position 4 (5/35; 14%). These results identify the major start sites for Mgf transcription with no evidence for alternative start sites further upstream or alternative splicing within the first four coding exons of Mgf. The transcription start site identified by the RACE technique has been confirmed by RNase protection analysis (see below).

Because the 5' ends of Mgf transcripts vary by only a few nucleotides, the presence of transcripts of different size observed on Northern blots appears to reflect differences in the length of the 3' UTR. The longest Mgf mRNA would total ~5.4 kb and is comprised of a 5' UTR of 197 nt, an ORF of 818 nt and a 3' UTR of 4432 nt (Figure 2). This 5.4 kb mRNA most likely corresponds to the most abundant transcript observed on Northern blots of murine tissues (Figure 1B). A second Mgf mRNA of ~3.7 kb, with a 3' UTR of only 2732 nt, could be produced by alternative use of the polyadenylation site and differs from the 5.4 kb transcript only in the length of the 3' UTR. This 3.7-kb mRNA likely represents the smaller, less abundant transcript (Figure 1B). The total amount of mouse genomic DNA spanned by the larger Mgf mRNA was estimated to be ~50 kb by Southern blot analysis using various cDNA probes.

Although the Mgf 3' UTR contains a single intron (intron 9) at position 1056 (see Figure 2), the remainder of the 3' UTR is colinear with the genomic clone. Scanning of the nucleotide sequence in the 3' UTR revealed the presence of several sites that may represent alternative splicing sites: sequences that display a single mismatch from the consensus 5' donor splice site (A/GAGGTA/GAGT) are located at nucleotide positions 4886 while three possible 3' acceptor splice sites (C/T, NC/TAG) are located at positions 1547, 1828 and 2369 (Figure 2). In addition, the 3' UTR contains numerous motifs (ATTTA) that have been associated with decreased mRNA stability (Shaw and Kamen 1986; Brewer 1991). A total of 14 ATTTA motifs are found in the 3' UTR of the 5.4-kb Mgf transcript with nine of these motifs in the 3' UTR of the 3.7-kb transcript (Figure 2). In addition, nine different motifs that perfectly match the consensus polyadenylation sequence (AA-TAAA) are found in the Mgf 3' UTR (Figure 2). Two of these are immediately upstream of the polyadenylation tracts of the Mgf cDNAs and are likely to represent the signals for termination of the 3.7- and 5.4-kb Mgf transcripts. However, three other AATAAA motifs are found upstream of the 3.7-kb polyadenylation site and four AATAAA motifs are found upstream of the 5.4-kb polyadenylation signal (Figure 2). This raises the possibility that additional 3' termini may exist for Mgf mRNA.

Examination of the 5' UTR sequence revealed that there are three ATGs located at positions 88, 123 and 198, with the latter encoding the initiator methionine for Mgf (Figure 2). With few exceptions, eukaryotic mRNAs initiate translation at the ATG closest to the 5' end of the message (Kozak 1987). If translation were to initiate from the first ATG in Mgf mRNA, at position
Alternatively, initiation of translation from the second ATG would produce a 7 aa polypeptide. However, none of the three ATGs are in a sequence context that is known to be most favorable for translation initiation of the three ATGs are in a sequence context that is most critical; KOZAK (1987). Although these observations suggest a mechanism for translational regulation of Mg expression, there is at present no experimental evidence for such translational control. To gain some information on the functional significance of the 5' UTR of Mg, we have compared the sequence of this region of Mg mRNA isolated from five different species (Figure 3). Although the complete 5' UTR of each of these species has not been published, alignment of the available sequences revealed from 82 to 94% identity within the 110-nt interval between the first ATG at position 88 of the mouse sequence and the initiator methionine of Mg.

DNA was not known. To elucidate this, a genomic library of C57BL/6J DNA was screened with a probe encompassing the very 5' end of MGF10 (ANDERSON et al. 1990, see Figure 1A) and a 17-kb clone was identified and subcloned into Bluescript (p14.1, Figure 4). Cleavage of p14.1 with EcoRI produced four fragments (Figure 4), three of which (B, C and D) were found to contain unique fragments when used as probes on Southern blots of mouse genomic DNA. Hybridization of the EcoRI-B fragment (EcoB) to blots of mouse DNA prepared by pulsed field gel electrophoresis revealed the presence of an island of rare-cutting enzymes between the 5' flanking region and the Mg coding region (see BEDELL et al. 1996). This suggests the presence of a CpG island(s) that is frequently found upstream of genes (reviewed by BIRD 1987). Further analysis of p14.1 revealed that only the EcoRI-D (EcoD) fragment contains sequences that hybridize to the 5' Mg probe. The 1.7-kb EcoD fragment was subcloned and sequenced (Figure 4). The 3' end of EcoD is located within the 5' UTR of Mg, at position 179, just upstream of the initiator methionine. The remainder of the 5' UTR, including the transcription initiation site identified by RACE (shown as +1 in Figure 4), is colinear with the genomic sequence of the EcoD fragment. To confirm that the RACE products described above represent the site of transcription initiation, a portion of the 3' region of EcoD was used as a probe in RNase I protection experiments.
M. A. Bedell, N. G. Copeland and N. A. Jenkins

2 Kb

GAATTCCTCTCACAGCTGCGGAGAffiAATGCAGTCTTffiCAACAGCTGGTTTTCCCTCAACAATA(CATTTTGGACTTTTGACCCTCAACAGTAAGA

TGAACATGTGACCATTTACAACATTTATTCTGAGTTAAATTCTCGCAGCACAAGTAOOCAACTA~AAAG~AGTTTOOCATATTAACTTTTTGTTTTGT

TTGTTTTTTAACAATOOCAAGTOIOCTTCTCTCAAACCAAATTTffiAATACTCCTCTCTTATAAGAATGACTffiACAGCATACTTGTTCCAAAAAATTACTG

ATOOGQATTTGATA=TTC~ATAACAAAATGCCTTTT~TTTTCCTCAATTACCTTTAAAAATATTGAA~CTTCGAGAAAAAGA

CTCTTG~~~TC~TTTCTTA~~~~~TAACATGGA~~~~~C~CACACAC~~~~~C~C~CACAC~~~~~~~C~TGC~~~~ACACA~~~~C~T~C

ACATTCTTATTTffiCATATTATTTTTTTTTTAGffiCAGTATGTCCAAGGTTCTGGATTT~ATCACAACTCTC~~TTTTTTT~GAATTAAC

Figure 4.—Schematic and partial sequence of a genomic clone for the 5′ flanking region of Mgf. A 17-kb genomic clone (p14.1) containing four EcoRI fragments (A–D) is shown with the solid line. The EcoRI-B, C, and D fragments (EcoB, EcoC and EcoD) are unique. The EcoRI-D fragment (EcoD) was subcloned and sequenced. The transcriptional start site is indicated (?). The 3′ end of EcoD is at position 179 (EcoD II). The sequence is numbered with +1 as the site of transcription initiation and extends to the first intron of Mgf with the encoded amino acids shown below the nucleotide sequence. The first ATG in the 5′UTR is indicated (\*). Potential transcription factor binding motifs are in boxes (TATAAA) (SAWADOGO and SENTENAC 1990) or are underlined: SP1 (KADONAGA et al. 1986); TRE/Rev (reverse of TRE) (LEE et al. 1987; SASSONE-CORSI et al. 1990); NRE-1 (BANIAHMAD et al. 1987); H-APF-1 (MAJELLO et al. 1990). A motif, at position 1091, that has a single mismatch from the consensus 3′ splice site is double underlined. Restriction enzyme sites for BssHII and SmaI are underlined.

Discussion

In studies described here, we report the cloning of the full-length Mgf transcription unit as well as genomic sequences that extend 5′ to this transcribed region. Examination of the nucleotide sequences of the 5′ and 3′ flanking regions has revealed potential regulatory elements for Mgf expression. The full-length Mgf tran-
script is comprised of a short 5' UTR, a 0.8-kb ORF and a long 3' UTR that contains at least two polyadenylation sites. Differential use of polyadenylation signals appears to be the mechanism by which two Mgf transcripts of 3.7 and 5.4 kb, differing only in the length of the 3' UTR, are generated. Because the larger transcript is much more abundant than the smaller transcript (Figure 1B, ANDERSON et al. 1990; HUANG et al. 1990; ZSEBO et al. 1990), these differences in the 3' UTR may affect the relative stability of the two mRNAs. Examination of the sequence of the Mgf 3' UTR reveals numerous ATTA motifs, a motif that is known to affect the stability of mRNAs of many growth factors and cytokines (SHAW and KAMIN 1986; BREWER 1991). Regulation of mRNA stability may therefore be one mechanism for controlling the expression of the Mgf gene. The present studies also demonstrate that the 5' UTR of Mgf contains two upstream ATGs that are conserved in different species and may provide a means for regulation of Mgf translation.

A major transcription initiation site was identified for Mgf mRNA that functions in all adult tissues tested and is located 28 nt downstream from a consensus TATA motif, the binding site for TFIIID (reviewed by SAWADOGA and SENTRYNAC 1990). However, the sequence around the Mgf initiation site does not contain a consensus initiator sequence (JAVAHERY et al. 1994) that is thought to specify precise initiation. This may explain the heterogeneity observed in the 5' ends of the Mgf mRNA, with transcription initiating at one or three G5s located at position +1, +2 and +4. The Mgf 5' flanking sequence contains numerous potential binding sites for Sp1, located from 38 to 51 nt upstream of the transcription initiation site. Although the functional significance of these and other motifs, such as the TRE, NRE and H-APF-1 motifs (Figure 4) identified in the Mgfl5' flanking region remains to be determined, there is evidence for a potential role of one of these motifs in Mgf transcription. The H-APF-1 binding site is thought to be required for the interleukin (II)-6-induced transcription of the gene encoding human C-reactive protein and is found upstream of many II-6-responsive genes (MAJELLO et al. 1990). In stromal cell cultures, II-6 treatment has been shown to abrogate the stimulatory effect of II-7 on Mgf mRNA levels (PALACIOS and NISHIKAWA 1992). The mechanism by which Mgf expression is affected by these cytokines is not presently known but could perhaps be mediated through the H-APF-1 binding motif in the 5' flanking region of Mgf. The only other factors that appear to be clearly involved in regulation of Mgf mRNA expression act through the cyclic AMP pathway (ROSSI et al. 1993; PACKER et al. 1994). Although a cAMP-responsive element (TGACGTCA; see SASSONE-CORSI et al. 1990) was not found in the upstream genomic sequence of Mgf, it is of note that the cAMP-responsive element is very similar to the TPA-responsive element (TRE; TGACTCA) and in fact differs by only one nucleotide from the TRE/Rev motif in EcoD (TGAGTCA; see Figure 4). Molecular dissection of genomic fragments reported here should allow identification of cis-acting elements that control Mgf transcription. The regulation of Mgf transcription may, however, be complex as tissue specific effects on Mgf mRNA expression occur in Sp1+/+ and Sp1−/− mice as the result of rearrangements located 115 and 195 kb, respectively, upstream of the Mgfl coding region in these mutants (BEDELL et al. 1995). Because we have found no evidence for alternative Mgf transcriptional start sites located far upstream of the Mgfl coding region, the rearrangements may disrupt far-distant regulatory elements for Mgf expression. Alternatively, transcription of Mgf in some tissues may be susceptible to long-range position effects on chromatin structure.

We are grateful to DOUGLAS E. WILLIAMS and STEWART D. LYMANN for the Mgfl0 cDNA, MARILYN POWERS for synthesis of oligonucleotides and automated DNA sequencing, and EIRIKUR STEINGRIMSSON and CAROLYN M. HUSTAD for reading the manuscript. M.A.B. was supported by fellowships from the Foundation for Advanced Cancer Studies and the National Institute for General Medical Sciences, National Institutes for Health. This research was supported by the National Cancer Institute, Department of Health and Human Services, under contract with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

**Note added in proof:** The sequences have been deposited in GenBank under accession numbers U44724 and U44725.

**LITERATURE CITED**


BEDELL, M. A., C. L. BRANNAN, E. P. EVANS, N. G. COPELAND, N. A. JENKINS et al., 1995 DNA rearrangements located over 100 kb 5' of the Steel (Sl) cDNA region in Steel-panda and Steel-contrasted mice deregulate Sl expression and cause female sterility by disrupt- ing ovarian follicle development. Genes Dev. 9: 455–470.


COPELAND, N. G., D. J. GILBERT, B. C. CHO, P. J. DONOVAN, N. A. JENKINS et al., 1990 Mast cell growth factor maps near the Steel
locus on mouse chromosome 10 and is deleted in a number of Steel alleles. Cell 63: 175–183.


Communicating editor: R. E. GANSCHOW