Structure of the murine macrophage scavenger receptor gene and evaluation of sequences that regulate expression in the macrophage cell line, $P388D_1$

R. Paul Aftring and Mason W. Freeman[1]

Department of Medicine, Lipid Metabolism and Endocrine Units, Massachusetts General Hospital, 32 Fruit Street, Boston, MA 02114

Abstract The structure of the entire murine scavenger receptor gene was determined; it consists of eleven exons spanning more than 60 kilobases. Primer extension showed that transcription initiates at a cluster of sites unassociated with a TATAA element. DNA sequences adjacent to these transcription start sites are highly conserved in murine, human, and bovine genes. When transcriptional activity was tested using a luciferase reporter gene, a promoter fragment (−124 to +20) stimulated luciferase production in $P388D_1$ macrophage-like cells but not in non-macrophage COS-7 or 3T3 cells. A longer promoter fragment (approximately 5 kb) stimulated luciferase activity a further 10-fold in $P388D_1$ cells. However, using a series of fragments from −67 to −1500 bp, a 127 bp fragment (−67 to +50) was as active as a 1500 bp fragment in these assays. Mutation of a putative AP-1 element in the −67 to +50 promoter fragment reduced luciferase activity by 40%; mutation of a putative GATA factor element to TATA increased luciferase activity nearly 2-fold while mutation to AATA had no effect and deletion of the GATA sequence inhibited activity by about 50%. The results suggest that a scavenger receptor promoter fragment can confer cell-specific transcription and that the activity may be mediated in part by factors that recognize the AP-1 and GATA elements.—**Aftring, R. P., and M. W. Freeman.** Structure of the murine macrophage scavenger receptor gene and evaluation of sequences that regulate expression in the macrophage cell line, $P388D_1$. *J. Lipid Res.* 1995. **36:** 1305–1314.

Supplementary key words gene transcription regulation

Macrophage scavenger receptors (SRs) are trimeric, integral membrane glycoproteins that bind an unusually broad array of macromolecular ligands. They were first identified through their capacity to mediate the intracellular accumulation of cholesteryl ester in macrophages presented with an acetylated form of LDL (acetyl-LDL) (1, 2). Subsequently, other forms of LDL, such as oxidized LDL, as well as a diverse group of polyanionic ligands, including bacterial endotoxin (reviewed in ref. 3), were demonstrated to bind to SRs. The available evidence suggests that the expression of SRs is limited to cells of the monocyte/macrophage lineage and some populations of activated smooth muscle cells (4). Although the ability to bind acetyl-LDL is widely used as a marker for endothelial cells, there is evidence to suggest that the protein responsible for this activity may not be identical to the receptor that has been isolated and cloned from macrophages (4). In addition, the identification of additional macrophage receptors that appear to bind and internalize oxidized LDL but not acetyl-LDL (5), indicates that a family of scavenger receptor proteins probably exists that have functional but not sequence similarities. At present, the physiologic function of the original SRs remains unclear, but their expression on macrophage foam cells suggests a central role in atherogenesis, while their binding to bacterial cell wall components (6, 7) argues for their importance in host defense against bacterial pathogens.

Complementary DNAs encoding two forms of the SR, designated type I and type II, have now been cloned from bovine (8, 9), mouse (10, 11), human (12), and rabbit (4) sources. We have previously mapped the mouse gene to a region of chromosome 8 which is probably syntenic with human chromosome 8 (11), and a chromosome 8 location has been confirmed for the human gene (13). Overall there is remarkable conservation of the protein, with about 70% homology between the amino acid sequences of the four species (10). The deduced amino acid sequences can be divided into six protein domains: domain I, cytoplasmic tail; II, transmembrane; III, spacer; IV, α helical coiled-coil; V, collagen; and VI, carboxyl terminal

Abbreviations: SR, scavenger receptor; LDL, low density lipoprotein; SRCR, scavenger receptor cyteine-rich domain; kb, kilobase; bp, base pair; PCR, polymerase chain reaction. The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession number U13873.

[1]To whom correspondence should be addressed.

region. The VIth domain serves to distinguish the type I from the type II receptor in that the former protein has an highly conserved 110 amino acid cysteine-rich carboxy terminus (SRCR domain), while the latter has a much shorter (6–17 amino acids) segment that is not conserved. The elucidation of the structure of the human SR gene (13) has strengthened the argument for this putative six domain structure in that the breaks between the exons encoding the receptor are located at or very near to the transition points from one protein domain to the next.

In this paper, we have determined the structure of the entire SR mouse gene and cloned approximately 5000 bp of DNA 5' to the start site of transcription. As expected, this gene structure closely resembles that of the human (13), though there are differences in the two genes involving sequences in their cytoplasmic tail and collagen domains. The gene has 11 exons spanning approximately 60 kb of DNA. Exons 1–8 are common to the two receptor types, with the type II cDNA arising from the splicing of these common segments to exon 9, while the type I cDNA substitutes exons 10 and 11 in place of 9. We have used the promoter DNA coupled to a reporter gene to demonstrate the presence of sequences that appear to play an important role in regulating the tissue-specific expression of the SR. In addition, we have identified two sequence regions, one of which is downstream of the start site of transcription, that appear to contribute significantly to basal transcription activity. The second region, upstream of the transcriptional start site, includes a consensus sequence for the binding of the GATA transcription factors and a putative AP-1/ets site. The AP-1/ets site was found to affect both basal and phorbol ester-induced SR expression in THP-1 cells (14), a human monocytic leukemic cell line; in this paper, its importance to basal transcription in the mouse macrophage cell line, $P388D_1$, is shown. In addition, we have mutated the GATA sequence and demonstrated marked effects on reporter gene expression. As the GATA site is completely conserved in the murine, human, and bovine promoters, it may play a critical role in regulating SR gene expression in a variety of species.

MATERIALS AND METHODS

Gene cloning and mapping

cDNA clones for the type I and type II mouse scavenger receptors were prepared by PCR amplification from an oligo-dT primed cDNA pool synthesized from mRNA prepared from the murine macrophage-like cell line, $P388D_1$ (11). The oligonucleotide primers were based on the translated sequence (10); the two alternative cDNAs were amplified with a common 5' sense primer (5'-CGCGGGGAATTCATGACAAAAGAGATGACAGAG-3') which was paired with a specific antisense type I (5'-CGCGGGATCGATTTTTTTGGTTTCATAATTGTA-3'), or type II (5'-CGCGGGATCGATTTTTTTAAGTTTTTCTTAGAA-3') oligonucleotide. The amplified products were cloned into pBluescript II after digestion with EcoRI and ClaI, restriction sites present in the 5' extensions of the oligonucleotide primers. Radiolabeled probes were prepared from the cDNA clones by a random primer method (15). Mouse genomic lambda phage libraries from Clontech (Palo Alto, CA) and Promega (Madison, WI) were screened using conventional plaque hybridization methods (16). The synthetic oligonucleotides,

5'-TTAAAGGTGATCGGGACAAATTGGCTTCCCTGGA-3'

5'-GCCTCCTGGACCCCAAGGTGAAAAGGGAGACAGAG-3

5'-GCCTCCTGGACCCCAAGGTGAAAAGGGAGACAGAG-3'

were end-labeled with polynucleotide kinase and pooled to probe the phage libraries for clones containing the collagen-like domain. Overlapping phage clones were mapped by a combination of restriction mapping, nucleic acid hybridization, and PCR across introns, using exon specific oligonucleotide primers.

PCR amplifications were typically performed in a 0.05-mL reaction containing 20 pmol of each primer, 2 mM $MgCl_2$, 0.2 mM dNTPs and 2.5 U Taq polymerase in 1 × Taq polymerase buffer (Promega, Madison, WI) for 25 to 30 cycles using a program of 30 s denaturation at 95°C, 30 s annealing at 55°C, and 30 s extension at 72°C, except when long products were expected and the extension time was increased to 60 s per kilobase of amplified product. Anchored PCR was carried out using the method and synthetic oligonucleotides described by Frohman, Dush, and Martin (17). The cDNA pool was synthesized from $P388D_1$ RNA using the synthetic oligonucleotide primer, 5'-CTCTGTCATCTCTTTTGTCAT-3'. Inverse PCR was performed as described by Ochman, Gerber, and Hartl (18).

DNA sequencing was performed using Sequenase (US Biochemicals, Cleveland, OH) or Circumvent (New England Biolabs, Beverly, MA) kits according to the manufacturers' instructions. Nucleic acid and amino acid sequence comparisons and alignments were prepared using the GCG computer programs (Genetics Computer Group, Madison, WI). Restriction endonucleases and other enzymes were obtained from New England Biolabs, Boehringer-Mannheim (Indianapolis, IN) or Promega. Radiolabeled nucleotides were obtained from Dupont NEN (Boston, MA).

Reporter gene experiments

Luciferase reporter vectors (pGL2) were obtained from Promega and luciferase assays were conducted using the

Promega luciferase assay kit. Beta-galactosidase assays were performed either spectrophotometrically using *o*-nitrophenyl-β-galactoside as a substrate (19) or by luminescence using a luminescent substrate (20) according to the manufacturer's directions (Tropix, Bedford, MA). The control plasmid pCMVβgal was constructed by cloning the CMV promoter from pcDNA1 (Invitrogen, San Diego, CA) in the place of the SV40 promoter of pSVβgal (Promega).

Reporter plasmids coupling the SR promoter region to luciferase were constructed in pGL2 Basic. An initial reporter construct was prepared from the cloned sequence shown in Fig. 1B; the 3' end of this fragment, including the ATG and intron junction, was removed by digestion with BbsI and blunted with Klenow. This truncated DNA fragment, termed −120Δ, was cloned into pGL2Basic at the SmaI and XhoI sites. All other reporter constructs were prepared after mutating the ATG in exon 1 (Fig. 1B) by PCR using the mutant oligonucleotide 5'-AGCTTGT CGACGCTTGGTAAAGAAGACAG-3'; this mutation created a new SalI site in place of the ATG, permitting cloning into pGL2Basic. The −1500 and −5000 promoter fragments were then sequentially cloned into pGL2Basic as a 1500 bp KpnI-XhoI fragment (−1500) to which a 3500 bp EcoRV-KpnI fragment was added (−5000). Promoter constructs between −120 and −650 were prepared by truncating the −1500 construct by PCR using specific oligonucleotides to define the 5' termini. Site-directed mutagenesis of the reporter constructs was performed using the unique site elimination method of Deng and Nickoloff (21). All mutations were confirmed by DNA sequence and their activity was tested using at least two independent clones.

Cell culture and transfection

COS-7 (monkey kidney) cells were obtained from Dr. Brian Seed and 3T3 (mouse fibroblast) cells from Dr. Richard Mulligan; both cell lines were maintained in DMEM supplemented with 10% calf serum, penicillin (50 units/mL) and streptomycin (50 μg/mL). P388D_1 cells were obtained from Dr. Monty Krieger and maintained in F12 supplemented with 10% fetal calf serum, penicillin (50 units/mL), and streptomycin (50 μg/mL). Transfection of COS-7 cells was performed using the DEAE-dextran method of Seed and Aruffo (22). P388D_1 cells were transfected by electroporation (23). In brief, P388D_1 cells were harvested and resuspended at 10^7 cells per mL in complete medium without antibiotics. Cell suspension (0.5 mL) was mixed with the luciferase reporter plasmid and β-galactosidase control plasmid and electroporated at 300V, 1000 μfd in a 4 mm gap cuvette. Cells were harvested 20–24 h after electroporation for luciferase and β-galactosidase assays. 3T3 cells were transfected either by DEAE-dextran or a calcium phosphate method.

RESULTS AND DISCUSSION

Cloning and gene structure

Sequential screening of the genomic libraries with cDNA probes readily yielded multiple phage clones containing all of the SR cDNA sequence except for the 5' un-

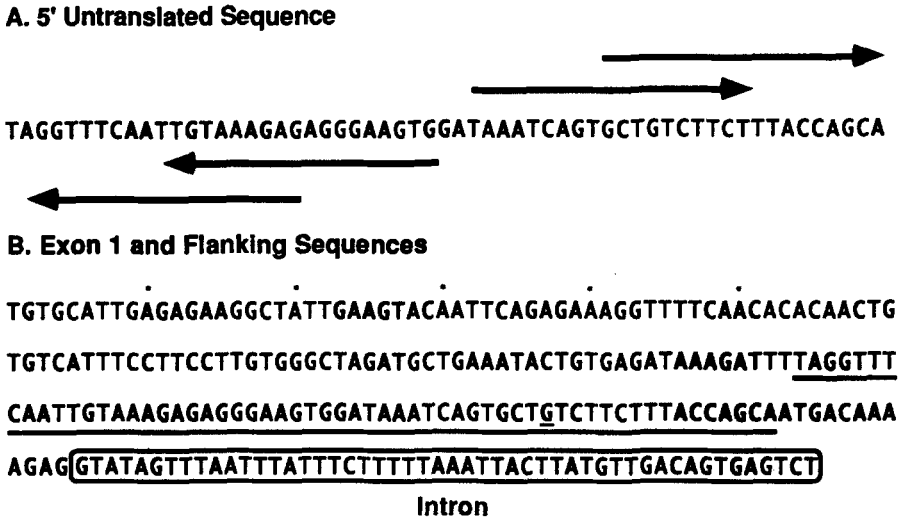

Fig. 1. DNA sequences of the 5' untranslated region and exon 1. Panel A shows the 5' untranslated sequence of the mouse SR as determined by anchored PCR. The arrows overlie the sequences of nested pairs of synthetic oligonucleotides used for inverse PCR, the 5' to 3' orientation of the oligonucleotides is indicated by the arrowhead. Panel B shows the sequence of exon 1 and its flanking sequences that were cloned by PCR using primers based on the sequences recovered by inverse PCR. The heavy underline indicates the location of the 5' untranslated sequence shown in panel A; the underlined G indicates the 3' terminus of the sequence used in the −120Δ construct used in reporter gene experiments.

translated region and the collagen-like domain. To obtain a specific probe for the 5' end of the SR gene, the 5' untranslated region of the SR message was cloned by an anchored PCR procedure (17). **Figure 1A** shows the sequence of the longest clone recovered by anchored PCR. This sequence was used to design a nested pair of oligonucleotide primers (Fig. 1A) for use in inverse PCR with $P388D_1$ genomic DNA as the template to allow cloning of the genomic sequences flanking the 5' untranslated sequence. The sequence of the cloned inverse PCR reaction products was used to design additional oligonucleotide primers, which were used to amplify the contiguous genomic sequence containing exon 1 and its flanking sequences, as shown in Fig. 1B. When the genomic library was screened using the sequence shown in Fig. 1B as a probe, a single phage clone containing exon 1, 5 kb of 5' flanking DNA, and 3 kb of the first intron (clone PRO4, **Fig. 2**) was obtained. To screen for the collagen-like domain, the data of Emi et al. (13) on the exon structure of the human SR collagen domain were used to design three synthetic oligonucleotides complementary to the mouse collagen domain, as described in Materials and Methods. The library was screened with a mixed pool of the three radiolabeled oligonucleotide probes; a single clone was obtained, which included two collagen exons and the type II carboxy terminus sequence (clone B1, Fig. 2).

The clones isolated from the genomic libraries allowed the construction of a map of the mouse scavenger receptor gene, as shown in Fig. 2. The coding sequence is divided into 11 exons distributed over more than 60 kb. The length of the introns between exons 6 and 7 and exons 9 and 10 are not defined by our clones; these introns are within the collagen-like domain and between the type II and type I specific ends, respectively. The order of the exons for the type I and II ends was deduced from the finding that exon 8 of the collagen domain is adjacent to the type II exon in a single phage isolate (clone B1, Fig. 2). This

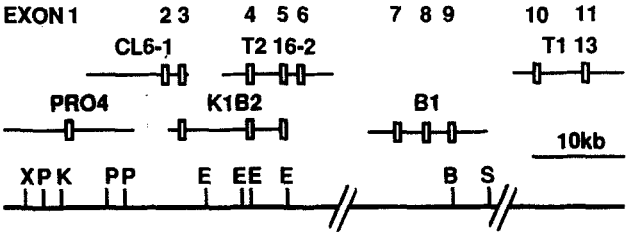

Fig. 2. Map of the mouse SR gene. Exons are numbered consecutively from 1 to 11 in the 5' to 3' direction. The relative location of each phage clone is shown along with its designation. Exon locations are designated by rectangles. A limited restriction map is shown at bottom of the figure but does not represent all sites for each enzyme within the gene. Interruptions in the gene designate unknown distances. X-XhoI, P-PstI, K-KpnI, E-EcoRI, B-BamHI, S-SalI.

matches the order in the human SR gene (13). The exon border splice junctions shown in **Table 1** are essentially the same as those determined for the human scavenger receptor gene, except in exons 1 and 7. Borders for exons 2, 3, 4, and 5 of the mouse gene were previously reported by Ashkenas et al. (10); their locations are confirmed here.

Sequence divergence in the cytoplasmic tail

A recent comparison of the published cDNA sequences of the mouse, bovine, human, and rabbit SR noted that the sequence of the mouse cytoplasmic domain diverged markedly from the other species (10). Fig. 1 and Table 1 show that an ATG is present in exon 1 which is suitable for translation initiation (24) and is in frame with the mature protein, suggesting that the mouse SR could have an extended cytoplasmic tail, relative to the length of the homologous region in the cow and man. This putative extended cytoplasmic tail is coded in part by nine nucleotides, adjacent to the first exon-intron splice junction; these bases have no homologue in the bovine and human genes (**Fig. 3**). The bovine and human SR genes, lacking a further upstream in-frame ATG, would be predicted to

TABLE 1. DNA sequences of intron-exon borders of the mouse scavenger receptor gene

Intron Sequence	Exon Sequence[a]	Intron Sequence	Exon #	Amino Acid at 3' End of Exon[b]
	ATG ACA AAA GAG	gtatagtttaattta...	1	glu 4
...aaatgtgtttgacag	ATG ACA GAG....CCT CAC A	gtaagtggaaaatga...	2	ser 40[c]
...tttgctttgttttag	GC ACT AAA....GTG ACA G	gtatgtatctgagat...	3	ala 78[c]
...tttattaaatattag	CT CAG CTT....CAA CAG GAG	gtaagccttcagaat...	4	glu 214
...tgactttgttttag	GAC ATC AGT....ATT CAA G	gtaaaatatcttgtc...	5	gly 277[c]
...tgttatttttaaag	GG CCT CCT....ATA AGA G	gtaaggctatagtaa...	6	gly 304[c]
...cttttaaccatgcag	GT ATT CCA....AGG TCG G	gtaggtattgtga	7	gly 334[c]
...ttgttttctcctaag	GA TCT CCT....GGA TCA A	gtaagttgatgactc...	8	arg 352 (Type II)[c] thr 352 (Type I)
...tcatatgcatgtcag	GA TCA GTA....		9	
...tttcctattttatag	CC CCC CTT....GGA CAA G	gtaatctgaactcta....	10	gly 415[c]
...atctgattgttgcag	GT ACT GGT....		11	

[a] Exon sequences are in capitals with the nucleotides grouped by codon.
[b] Amino acids are numbered relative to a putative ATG initiation codon in exon 1.
[c] Codon is interrupted at intron-exon border.

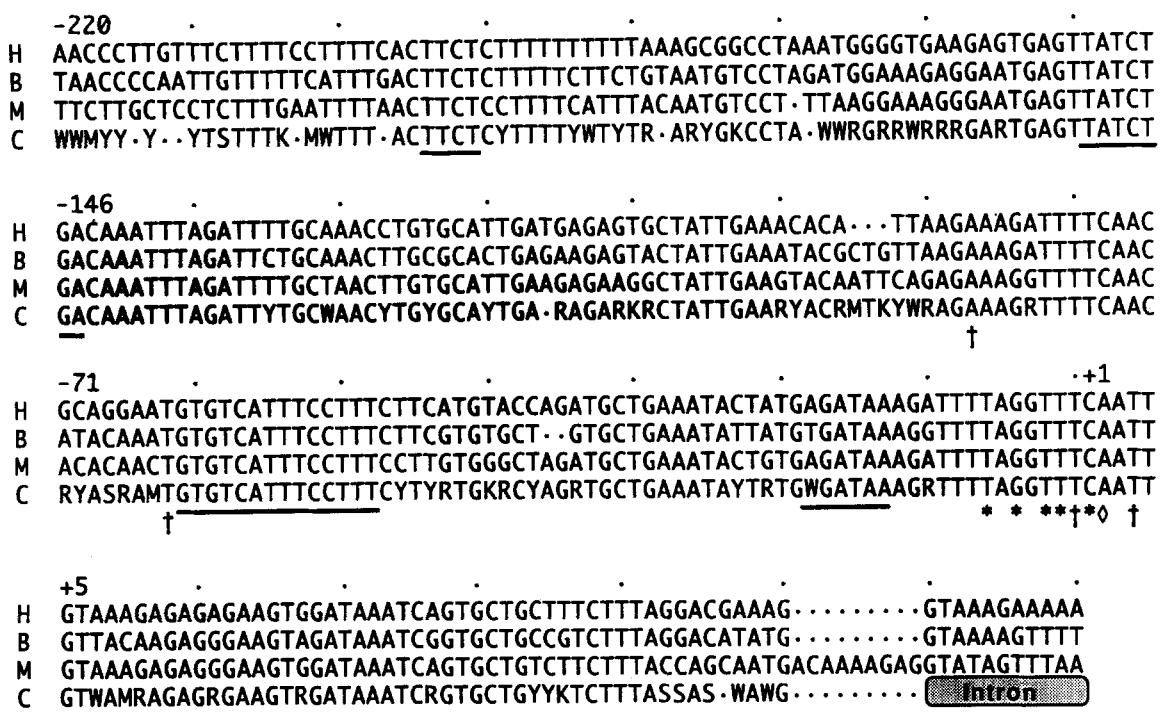

Fig. 3. Comparison of promoter sequences of the mouse (M), bovine (B) and human (H) SR genes. Underlined sequences represent potential promoter elements: −194, Pu.1 (33); −151, NF-Gmb (31); −64, AP-1/ets (14); −20, GATA (32). The start sites of transcription for the human (†), mouse (*) and bovine (◊) genes are indicated. The site (◊) is also a primary start site for the human and mouse transcripts. Human and bovine sequences and transcriptional start sites are from Moulton et al. (25). C designates the consensus promoter sequence which is shown using the IUB degeneracy codes except where degeneracy between the species was complete.

initiate translation in exon 2 (25). The divergence of amino acid sequence in the N-terminal cytoplasmic tail of the SR, with or without the putative murine extension, suggests that these sequences play no critical role in directing scavenger-receptor mediated endocytosis. More distal cytoplasmic tail sequences adjacent to the transmembrane domain are highly conserved among the species, and may therefore be critical for endocytosis, though they lack the characteristic NPXY motif found to be important in endocytosis mediated by the LDL and transferrin receptors (26, 27).

Collagen-like domain variations

Exons 6, 7, and 8 encode the collagen-like domain of the SR. Exon 6 of the mouse gene, like exon 6 of the human gene, is 81 bp in length, encoding 9 Gly-X-Y repeats. Exon 7 of the mouse is 90 bp in length and encodes 10 Gly-X-Y repeats, unlike human exon 7, which is 81 bp long and encodes 9 Gly-X-Y repeats. Comparisons of the nucleotide and deduced amino acid sequences of exon 7 from the human and mouse genes are shown in **Table 2;** these comparisons suggest that a Gly-X-Y triplet is deleted at the 3' end of human exon 7. SR sequences of other

TABLE 2. DNA and amino acid sequences of human and mouse scavenger receptor exon 7

```
        GlyProPro GlyLeuLys GlyAspArg GlyAlaIle GlyPhePro
Human    GTCCTCCG GGTCTTAAA GGTGATCGG GGAGCAATT GGCTTTCCT
          ••  •••  ••• ••••• ••••••••• •••  •••• ••••• •••
Mouse    GTATTCCA GGTGTTAAA GGTGATCGG GGACAAATT GGCTTCCCT
        GlyIlePro GlyValLys GlyAspArg GlyGlnIle GlyPhePro

        GlySerArg GlyLeuPro GlyTyrAla GlyArgPro
Human   GGAAGTCGA GGACTCCCA GGATATGCC GGAAGGCCA G
        •••••••••• •••  •••• •••   •  ••••• ••• • *
Mouse   GGAGGTCGA GGAAACCCA GGAGCACCA GGAAAGCCA GGGAGGTCG G
        GlySerArg GlyAsnPro GlyAlaPro GlyLysPro GlyArgSer
```

Human exon 7 sequence is from Emi et al. (13). The initial G in the first Gly codon is present at the end of exon 6 (Table 1). A G to T transversion in the mouse sequence at the marked site (*) would create a new potential splice-donor site. Dots (.) indicate nucleic acid identity between the mouse and human sequences.

mammals, especially lower primates, could provide interesting information on the evolutionary timing of such a mutation. Exon 8 of the human and mouse genes is 54 bp long and encodes 7 Gly-X-Y repeats in the mouse SR but only 6 Gly-X-Y repeats in the human SR. The lengths of exons 6, 7, and 8 of both the mouse and human genes represent exact multiples of 9 bp which is consistent with the findings that the exons of fibrillar collagen genes (28) and other collagen-like genes (29) consist of a nearly uniform pattern of exact multiples of 9 bp encoding Gly-X-Y triplets. Table 1 shows that the exons of the SR collagen-like domain (exons 6, 7, and 8) all begin with an interrupted glycine codon, which is consistent with the exon structure of other non-fibrillar collagen proteins (29) and is unlike the exon structure of the fibrillar collagen genes, where the majority of exons begin with a complete glycine codon at the 5' acceptor site and terminate with a complete Y codon at the 3' donor site (28). The conserved 54 bp length (six 9 bp elements) of exon 8 for both the mouse and human genes suggests that the collagen motif extended the full length of this exon in the ancestral SR gene, but that the terminal Gly-X-Y repeat was not essential for SR function, as it is mutated in the human gene.

SR subtypes

Exon 9 encodes the truncated type II carboxy terminus and the 3'-untranslated sequence. The DNA sequence of clone B1 (Fig. 2) includes a polyadenylation site previously identified in a mouse type II cDNA clone (10). Three different sized messages for the type II SR have been found in P388D$_1$ cells and there is data suggesting that these arose from alternate polyadenylation sites (10). Exons 10 and 11 encode the highly conserved type I carboxyl terminus, the SR cysteine-rich (SRCR) domain (11), and exon 11 also encodes the 3' untranslated sequence of the Type I message. Exon 9 and exon 10 represent alternative splice acceptor sites for the terminus of exon 8, thus yielding the type II and type I receptor mRNAs, respectively. The SRCR, type I specific C-terminal domain, is identifiable in a diverse collection of proteins, though its functional role remains obscure (30). The recent finding of differential binding of a bacterial endotoxin (ReLPS) by the murine type I and type II SR (10) is the first evidence to suggest a difference in function that could be mediated by the SRCR. As both the type I and type II receptor mRNAs appear to be present in cells expressing either message (4, 8, 9, 11, 12), there are no data available to suggest that differential regulation of this alternative splicing event can occur. Whether the generation of the two mouse receptor mRNAs is a stochastic or regulated phenomenon can now be investigated using these cloned gene fragments. In addition, the cloning of the murine type specific exons and their flanking sequences described here provides the necessary reagents to generate specific deletions of each receptor subtype in vivo by homologous recombination. This approach could provide a useful tool for examining any important functional differences between the type I and type II receptor and thereby clarify the role of the intriguing SRCR domain.

Promoter sequences and functional analysis

Primer extension analysis of RNA from P388D$_1$ cells mapped a cluster of transcription start sites distributed over a 9 bp distance (**Fig. 4**); the product of the anchored PCR procedure (Fig. 1A) corresponds to the longest of these transcripts. There are no identifiable TATAA-like or CAAT-like sequences in this region of the promoter (Fig. 3). These results are qualitatively similar to those described for the human SR promoter (25), where three, closely spaced, but distinct transcriptional start sites were identified in phorbol ester treated THP-1 cells. The start site found for the mouse gene corresponds to site A in the human promoter (25). The start site for the bovine gene, determined using bovine lung RNA, also mapped to this region, but was reported to be a discrete site. Emi et al. (13), also using phorbol ester-treated THP-1 cells, mapped a single discrete transcription start site for the human gene which corresponds to site B of Moulton et al. (25). There is no apparent explanation for the discrepancy in the results of Emi et al. (13) and Moulton et al. (25), regarding the transcription start site(s) of the human SR gene, but it is possible that the different transcription start sites are favored under different physiological conditions.

A striking feature of the mouse, human, and bovine SR genes is the high degree of conservation of the nucleic acid sequence in the proximal promoter and 5' untranslated

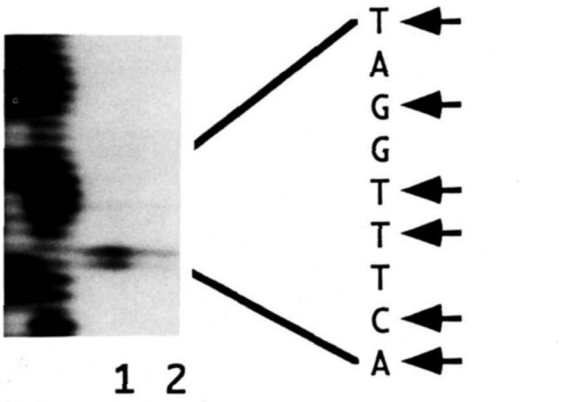

Fig. 4. Primer extension analysis of P388D$_1$ RNA. Total RNA was hybridized to the synthetic oligonucleotide 5'-GTCATTGCTGATAAGA AGAAGACAGC-3', and primer extension conducted as described (16). The extended products were electrophoresed on a 6% DNA sequencing gel with accompanying DNA sequencing reactions as length markers. Lanes 1 and 2 show the products using 75 and 30 μg of total RNA as template, respectively. A control with yeast tRNA as the template yielded no extended products (data not shown).

regions (Fig. 3). Within the highly conserved region of the promoter several conserved nucleic acid motifs were identified by comparison to the TFDSITE database (Dr. David Ghosh, NCBI, NLM). At position −151 the conserved sequence TCAGATA (antisense strand) may represent a cytokine response motif as found in the mouse and human GM-CSF and IL-3 genes (31). At position −64 there is a conserved sequence, TGTGTCATTTCCTTT, which was described as a combined ets and AP-1 site and demonstrated to participate in the basal transcription of a reporter gene in THP-1 cells and also to confer responsiveness to phorbol ester (14). The AGATAA sequence, located at −20, is conserved between the species, and is a potential binding site for the GATA transcription factors (32).

Promoter activity determined using a coupled reporter gene

In initial experiments to evaluate the function of the proximal promoter, a short fragment from −124 to +20 (−120Δ) was coupled to a luciferase reporter gene and its activity was examined in P388D$_1$, 3T3, and COS-7 cells, as shown in **Table 3.** This promoter fragment stimulated luciferase activity approximately 10-fold over the promoterless vector in the P388D$_1$ cells but did not stimulate luciferase activity in the other cell types. These results suggest that a portion of the essential information for macrophage-specific expression of the SR may reside in this short promoter fragment. Under these conditions, cell-specific expression in P388D$_1$ cells does not appear to require the Pu.1 site at position −194 which Moulton et al. (33) suggested to be involved in cell-specific expression using the human SR promoter. Whether the Pu.1 site contributes to cell-specific expression in the mouse will require further study.

TABLE 3. Stimulation of luciferase expression by the SR promoter in different cell types

Cell Line	Transfection Method	Relative Luciferase Activity
P388D$_1$	Electroporation	9.2 ± 2.2[a] (7)
COS-7	DEAE-dextran	1.3 ± 0.1 (3)
3T3	DEAE-dextran	0.3 ± 0.2 (3)
3T3	$CaPO_4$	2.7 ± 1.8 (3)

Cells were cotransfected with either the promoterless luciferase plasmid (pGL2Basic) or the −120Δ luciferase construct and the control plasmid pSVβgal. The luciferase activity in each transfection was normalized to the β-galactosidase activity from the control plasmid. The results shown are the ratios of luciferase activity from the −120Δ construct to the activity from the promoterless plasmid. P388D$_1$ cells were transfected by electroporation with 25 μg of each plasmid. COS-7 cells were transfected with 4 μg/mL of the luciferase plasmids and 1 μg/mL of pSVβgal in 400 μg/mL DEAE dextran. 3T3 cells were transfected with 4 μg/mL of the luciferase plasmids and 1 μg/mL of pSVβgal in 100 μg/mL DEAE dextran or with 10 μg of the luciferase plasmids and 2 μg of pCMVβgal in $CaPO_4$ precipitates. For COS-7 and 3T3 transfections, enzyme activities were measured 48 h after transfection. Results are shown as the mean ± standard error with the number of observations in parentheses.

[a] Significantly different from 1, $P < 0.01$.

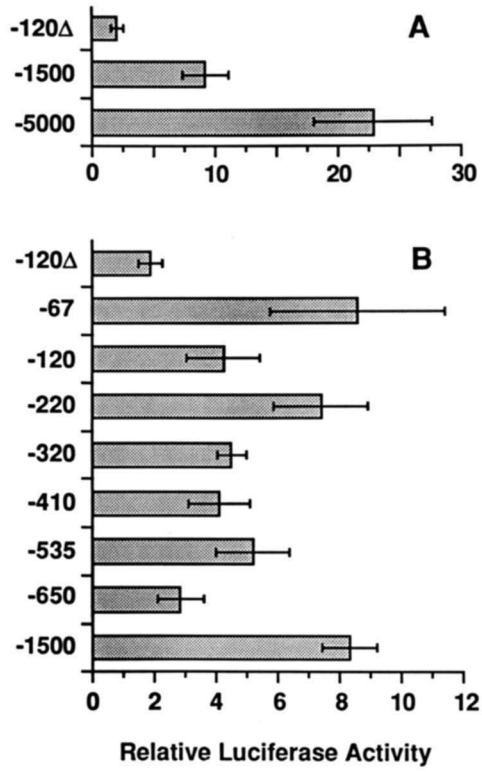

Fig. 5. Luciferase activity produced from SR promoter fragments transfected into P388D$_1$ cells. P388D$_1$ cells were transfected with 25 μg of each promoter construct by electroporation. Luciferase activity was normalized to β-galactosidase activity produced from the cotransfected plasmid pCMVβgal, 25 μg in panel A and 5 μg in panel B. Results represent the mean and standard error of 4 to 7 determinations.

To simplify later constructs, a directed mutation was introduced into the promoter sequence at +50 to alter the initial ATG of the SR transcript and allow promoter constructs to be created that include sequences immediately upstream of this site. Constructs including longer fragments of the SR promoter were then tested for their ability to stimulate luciferase expression in P388D$_1$ cells, as shown in **Fig. 5A.** There was a 5-fold and 10-fold increase in luciferase activity for promoter fragments of 1500 and 5000 bp, respectively, suggesting that additional enhancer sequences, which function to increase SR expression in P388D$_1$ cells, are present in the upstream region. The expression of the reporter gene under control of the −5000 bp promoter was also found to be restricted to P388D$_1$ cells as it had no activity when transfected into COS-7 or 3T3 cells (data not shown). To further define areas within the 1500 bp fragment that are important for SR expression, a series of truncations were prepared and tested for their ability to stimulate luciferase expression (Fig. 5B). A striking finding among these truncations was that a very short promoter fragment (−67) was essentially equivalent to the −1500 fragment. Within this series of promoter fragments only the −67 and −1500 constructs are significantly different from the −120Δ construct. The pat-

tern of activity from −120 to −650 suggests that other enhancer and silencer elements may be present in this region. However, these potential elements are either quantitatively less important for the activity of the SR promoter in $P388D_1$ cells than the sequences between −67 and +50 or the enhancing and silencing activities effectively negate the impact of each other. The −67 fragment includes the composite AP-1/ets site that Wu et al. (14) identified as an important element in the function of the human SR promoter in THP-1 cells. In addition, it can be inferred from the data of Fig. 5B that the sequences between +20 and +48 (absent in the −120Δ construct) may stimulate expression, though in the studies of the human promoter, sequences in this region were thought to be inhibitory (33). The differences in our data and those of Moulton et al. (33) may be due to the context in which the promoter elements are arrayed. We have used simple truncations whereas Moulton et al. (33) used site-directed mutations within a fixed length promoter fragment (−245 to +46). Differences between the cell types used could also account for the divergent results.

To further define the elements within the −67 fragment that stimulate transcription in $P388D_1$ cells, we have introduced directed mutations into the AP-1 site and to the consensus GATA element at position −20 (**Fig. 6**). Mutation of the AP-1 element reduced luciferase expression by about 40%. This mutation in the AP-1 element was previously shown to abrogate the function of this element in THP-1 cells (14). Mutation of the G in the GATA element to produce a consensus TATAAA sequence stimulated luciferase expression nearly 2-fold, while mutation of the G to an A had no effect on luciferase expression. Deletion of the four nucleotide GATA core reduced luciferase expression by about 50%. The effects of the G to T and G to A mutations are similar to the effects of identical mutations of a similarly placed GATA element in the platelet factor 4 promoter (35). The GATA transcription factors (GATA-1 to GATA-4) bind to a consensus element (WGATAW) and play a role in tissue specific expression of the globins (34), platelet factor 4 (35) and α-myosin heavy chain (36). The proximity of the GATA element to the start site of transcription also raised the possibility that it may be important for efficient transcription initiation; while this is consistent with the results we obtained with the GATA mutations, a more detailed analysis of this region is currently in progress to test this hypothesis. It has been shown previously that TBP and GATA-1 (and -2) can compete for the same binding site but that formation of a preinitiation complex was inhibited in the presence of GATA-1 (37). It was suggested that inhibition of transcription initiation by GATA-1 (or other GATA transcription factors) could be a mechanism for producing cell-specific expression. This would predict that mutation of the GATA to a consensus TATAAA element should enhance the activity of promoters with this feature and could eliminate cell specificity, as was observed in studies of the platelet factor 4 promoter (35). Studies are currently underway to determine whether the GATA consensus sequence indeed binds nuclear proteins from $P388D_1$ cells and whether these cells express any of the known GATA transcription factors.

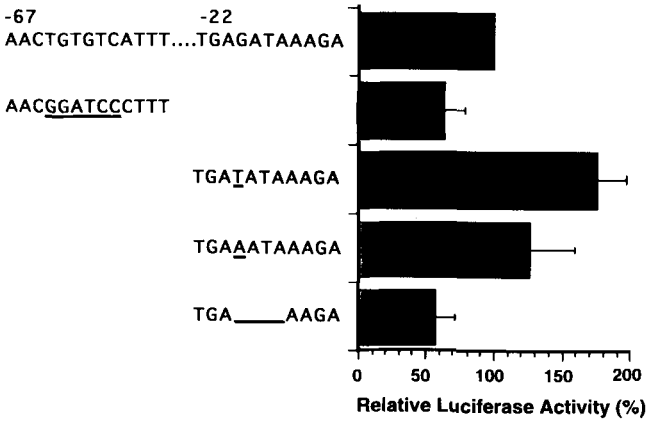

Fig. 6. Effect of mutations of the −67 SR promoter fragment on luciferase activity in $P388D_1$ cells. The native sequence is shown at the top of the figure on the left; sequence numbering corresponds to that shown in Fig. 3. Mutated sequences are shown adjacent to their respective luciferase activities with the mutated nucleotides underlined. $P388D_1$ cells were transfected by electroporation. Luciferase activities were normalized to β-galactosidase activity. Results shown are the mean and standard error of 4 to 8 determinations.

In this work, we have determined the structure of the entire mouse scavenger receptor gene and found it to extend over approximately 60 kb of genomic DNA. Eleven exons constitute the gene, with alternative splicing of exon 8 to either exon 9 or to exons 10 and 11 accounting for the generation of the type II and type I receptors, respectively. The organization of the mouse gene closely resembles that of the human, with minor differences noted around sequences encoding the N-terminus of the protein and within the collagen domains. Approximately 5 kb of DNA upstream of the start site of transcription of the mouse gene have been cloned and these sequences have been used to characterize regions of the SR gene promoter that are important in regulating basal transcription activity as well as tissue specific expression. The identification of a potential GATA binding site and the dramatic loss of promoter activity that followed its deletion suggests a role for GATA proteins in the regulation of SR activity. Further clarification of the control of scavenger receptor gene expression in macrophages should enhance our understanding of the modulation of gene expression that occurs in these cells during the process of atherogenesis. In conjunction with the development of better murine models of atherosclerosis (38, 39), it should now be possible to test the results of in vitro assays of SR gene activity under pathophysiologically relevant conditions in vivo. Finally, the cloning of the entire

mouse SR gene may enhance our understanding of the functional differences between type I and type II receptors, as the reagents are now available that would permit selective inactivation of the type I receptor by homologous recombination.

We thank Anne Sarbinowski and Jennifer Teicher for their expert technical assistance and Drs. Marie Demay, Joel Habener, and Henry Kronenberg for their comments and thoughtful advice. This study was supported by National Institutes of Health grant HL-45098 (to M.W.F.). R.P.A. was supported by a National Institutes of Health training grant awarded to the Endocrine Unit, Massachusetts General Hospital (DK07028-17) and a post-doctoral research fellowship from the American Heart Association, Massachusetts affiliate.

Manuscript received 2 November 1994 and in revised form 22 February 1995.

REFERENCES

1. Brown, M. S., S. K. Basu, J. R. Falck, Y. K. Ho, and J. L. Goldstein. 1980. The scavenger cell pathway for lipoprotein degradation: specificity of the binding site that mediates the uptake of negatively-charged LDL by macrophages. *J. Supramol. Struct.* **13:** 67–81.
2. Goldstein, J. L., Y. K. Ho, S. K. Basu, and M. S. Brown. 1979. Binding site on macrophages that mediates uptake and degradation of acetylated low density lipoprotein, producing massive cholesterol deposition. *Proc. Natl. Acad. Sci. USA.* **76:** 333–337.
3. Krieger, M. 1992. Molecular flypaper and atherosclerosis: structure of the macrophage scavenger receptor. *Trends Biochem. Sci.* **17:** 141–146.
4. Bickel, P. E., and M. W. Freeman. 1992. Rabbit aortic smooth muscle cells express inducible macrophage scavenger receptor messenger RNA that is absent from endothelial cells. *J. Clin. Invest.* **90:** 1450–1457.
5. Endemann, G., L. W. Stanton, K. S. Madden, C. M. Bryant, R. T. White, and A. A. Protter. 1993. CD36 is a receptor for oxidized low density lipoprotein. *J. Biol. Chem.* **268:** 11811–11816.
6. Hampton, R. Y., D. T. Golenbock, M. Penman, M. Krieger, and C. R. Raetz. 1991. Recognition and plasma clearance of endotoxin by scavenger receptors. *Nature.* **353:** 342–344.
7. Dunne, D. W., D. Resnick, J. Greenberg, M. Krieger, and K. A. Joiner. 1994. The type I macrophage scavenger receptor binds to Gram-positive bacteria and recognizes lipoteichoic acid. *Proc. Natl. Acad. Sci. USA.* **91:** 1863–1867.
8. Kodama, T., M. Freeman, L. Rohrer, J. Zabrecky, P. Matsudaira, and M. Krieger. 1990. Type I macrophage scavenger receptor contains a-helical and collagen-like coiled-coils. *Nature.* **343:** 531–535.
9. Rohrer, L., M. Freeman, T. Kodama, M. Penman, and M. Kreiger. 1990. Coiled-coil fibrous domains mediate ligand binding by macrophage scavenger receptor type II. *Nature.* **343:** 570–572.
10. Ashkenas, J., M. Penman, E. Vasile, S. Acton, M. Freeman, and M. Krieger. 1993. Structures and high and low affinity ligand binding properties of murine type I and type II macrophage scavenger receptors. *J. Lipid Res.* **34:** 983–1000.
11. Freeman, M., J. Ashkenasas, D. J. Rees, D. M. Kingsley, N. G. Copeland, N. A. Jenkins, and M. Krieger. 1990. An ancient, highly conserved family of cysteine-rich protein domains revealed by cloning type I and type II murine macrophage scavenger receptors. *Proc. Natl. Acad. Sci. USA.* **87:** 8810–8814.
12. Matsumoto, A., M. Naito, H. Itakura, S. Ikemoto, H. Asaoka, I. Hayakawa, H. Kanamori, H. Aburatani, F. Takaku, H. Suzuki, Y. Kobari, T. Miyai, K. Takahashi, E. H. Cohen, R. Wydro, D. E. Housman, and T. Kodama. 1990. Human macrophage scavenger receptors: primary structure, expression, and localization in atherosclerotic lesions. *Proc. Natl. Acad. Sci. USA.* **87:** 9133–9137.
13. Emi, M., H. Asaoka, A. Matsumoto, H. Itakura, Y. Kurihara, Y. Wada, H. Kanamori, Y. Yazaki, E. Takahashi, M. Lepert, et al. 1993. Structure, organization, and chromosomal mapping of the human macrophage scavenger receptor gene. *J. Biol. Chem.* **268:** 2120–2125.
14. Wu, H., K. Moulton, A. Horvai, S. Parik, and C. K. Glass. 1994. Combinatorial interactions between AP-1 and ets domain proteins contribute to the developmental regulation of the macrophage scavenger receptor gene. *Mol. Cell. Biol.* **14:** 2129–2139.
15. Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132:** 6–13.
16. Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1991. Current Protocols in Molecular Biology. John Wiley & Sons, New York.
17. Frohman, M. A., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA.* **85:** 8998–9002.
18. Ochman, H., A. S. Gerber, and D. L. Hartl. 1988. Genetic applications of an inverse polymerase chain reaction. *Genetics.* **120:** 621–623.
19. Rosenthal, N. 1987. Identification of regulatory elements of cloned genes with functional assays. *Methods Enzymol.* **152:** 704–720.
20. Jain, V. K., and I. T. Magrath. 1991. A chemiluminescent assay for the quantitation of β-galactosidase in the femtogram range: application to quantitation of β-galactosidase in lacZ-transfected cells. *Anal. Biochem.* **199:** 119–124.
21. Deng, W. P., and J. A. Nickoloff. 1992. Site-directed mutagenesis of virtually any plasmid by eliminating a unique site. *Anal. Biochem.* **200:** 81–88.
22. Seed, B., and A. Aruffo. 1987. Molecular cloning of the CD2 antigen, the T-cell erythrocyte receptor, by a rapid immunoselection procedure. *Proc. Natl. Acad. Sci. USA.* **84:** 3365–3369.
23. Chu, G., H. Hayakawa, and P. Berg. 1987. Electroporation for the efficient transfection of mammalian cells with DNA. *Nucleic Acids Res.* **15:** 1311–1326.
24. Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* **108:** 229–241.
25. Moulton, K. S., H. Wu, J. Barnett, S. Parthasarathy, and C. K. Glass. 1992. Regulated expression of the human acetylated low density lipoprotein receptor gene and isolation of promoter sequences. *Proc. Natl. Acad. Sci. USA.* **89:** 8102–8106.
26. Collawn, J. F., L. A. Kuhn, L. F. Liu, J. A. Tainer, and I. S. Trowbridge. 1991. Transplanted LDL and mannose-6-phosphate receptor internalization signals promote high-efficiency endocytosis of the transferrin receptor. *EMBO J.* **10:** 3247–3253.

27. Collawn, J. F., M. Stangel, L. A. Kuhn, V. Esekogwu, S. Q. Jing, I. S. Trowbridge, and J. A. Tainer. 1990. Transferrin receptor internalization sequence YXRF implicates a tight turn as the structural recognition motif for endocytosis. *Cell.* **63:** 1061–1072.
28. Perala, M., K. Elima, M. Metsaranta, R. Rosati, B. de Crombrugghe, and E. Vuorio. 1994. The exon structure of the mouse α2(IX) collagen gene shows unexpected divergence from the chick gene. *J. Biol. Chem.* **269:** 5064–5071.
29. Taylor, M. E., P. M. Brickell, R. K. Craig, and J. A. Summerfield. 1989. Structure and evolutionary origin of the gene encoding a human serum mannose-binding protein. *Biochem. J.* **262:** 763–771.
30. Resnick, D., A. Pearson, and M. Krieger. 1994. The SRCR superfamily: a family reminiscent of the Ig superfamily. *Trends Biochem. Sci.* **19:** 5–8.
31. Shannon, M. F., J. R. Gamble, and M. A. Vadas. 1988. Nuclear proteins interacting with the promoter region of the human granulocyte/macrophage colony-stimulating factor gene. *Proc. Natl. Acad. Sci. USA.* **85:** 674–678.
32. Arceci, R. J., A. A. J. King, M. C. Simon, S. H. Orkin, and D. B. Wilson. 1993. Mouse GATA-4: a retinoic acid-inducible GATA-binding transcription factor expressed in endodermally derived tissues and heart. *Mol. Cell. Biol.* **13:** 2235–2246.
33. Moulton, K. S., K. Semple, H. Wu, and C. K. Glass. 1994. Cell specific expression of the macrophage scavenger receptor gene is dependent on PU.1 and a composite AP-1/ets motif. *Mol. Cell. Biol.* **14:** 4408–4418.
34. Orkin, S. H. 1990. Globin gene regulation and switching: circa 1990. *Cell.* **63:** 665–672.
35. Ravid, K., T. Doi, D. L. Beeler, D. J. Kuter, and R. D. Rosenberg. 1991. Transcriptional regulation of the rat platelet factor 4 gene: interaction between an enhancer/silencer domain and the GATA site. *Mol. Cell. Biol.* **11:** 6116–6127.
36. Molkentin, J. D., D. V. Kalvakolanu, and B. E. Markham. 1994. Transcription factor GATA-4 regulates cardiac muscle-specific expression of the a-myosin heavy-chain gene. *Mol. Cell. Biol.* **14:** 4947–4957.
37. Aird, W. C., J. D. Parvin, P. A. Sharp, and R. D. Rosenberg. 1994. The interaction of GATA-binding proteins and basal transcription factors with GATA box-containing core promoters. *J. Biol. Chem.* **269:** 883–889.
38. Plump, A. S., J. D. Smith, T. Hayek, K. Aalto-Setala, A. Walsh, J. G. Verstuyft, E. M. Rubin, and J. L. Breslow. 1992. Severe hypercholesterolemia and atherosclerosis in apolipoprotein E-deficient mice created by homologous recombination in ES cells. *Cell.* **71:** 343–353.
39. Ishibashi, S., M. S. Brown, J. L. Goldstein, R. D. Gerard, R. E. Hammer, and J. L. Herz. 1993. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* **92:** 883–893.