Novel clustering of Sp1 transcription factor binding sites at the transcription initiation site of the human muscle phosphofructokinase P1 promoter

Jennifer L. Johnson and Alan McLachlan*

Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037, USA

Received June 23, 1994; Revised and Accepted October 17, 1994

EMBL accession no. Z34905

ABSTRACT

The regulatory sequence elements of the human muscle phosphofructokinase (HPFKM) p1 promoter from –655 to +78 were cloned and characterized. In the human cervical carcinoma cell line, HeLa S3, the HPFKM type C RNA initiated from a single predominant transcription initiation site and the HPFKM p1 promoter displayed transcriptional activity in transient transfection assays. The HPFKM p1 promoter region was shown to possess eight binding sites for the Sp1 transcription factor by DNase I footprinting and gel retardation analysis. The functional importance of these interactions was examined by transient transfection analysis in *Drosophila* SL2 and HeLa S3 cells. This analysis demonstrated that the HPFKM p1 promoter sequence between +12 and +78 retained Sp1-dependent transcriptional activity in *Drosophila* SL2 cells and retained promoter activity in HeLa S3 cells. These results suggest that the Sp1 binding site (site 8 between +12 and +21) immediately adjacent to the transcription initiation site represents an important regulatory element of this promoter at least in the context of the minimal HPFKM p1 promoter. However mutagenesis of the Sp1 site 8 demonstrated that, in the context of a larger HPFKM p1 promoter region containing Sp1 sites 1 to 7, it now contributed very little to the total promoter activity. Therefore it appears the Sp1 sites in the HPFKM p1 promoter display functional redundancy.

INTRODUCTION

Phosphofructokinase (ATP:D-fructose-6-phosphate 1-phosphotransferase, EC2.7.1.11; PFK) is an important regulatory enzyme of glycolysis which catalyzes the phosphorylation of fructose-6-phosphate to fructose-1,6-diphosphate (1–3). The human enzyme is a tetrameric 340-kDa protein generated by the random association of the PFK polypeptides encoded by the muscle (M), liver (L) and platelet (P) genes (4–13). The three PFK genes

are expressed at different times during development and have distinct tissue specificities (14–16). As a result of the differences in the expression patterns of the PFK genes, various tissues express different amounts of the three PFK polypeptides. This produces tetrameric PFK protein of different subunit composition and enzymatic properties which correlates with the relative levels of glycolysis and gluconeogenesis in different tissues (1,3,17). In tissues such as skeletal muscle, heart and brain where metabolic energy requirements are principally met by glycolysis, the predominant isoenzyme is PFKM whereas in organs such as liver and kidney where gluconeogenesis is active the PFKL isoenzyme is the predominant subunit (3,17–19).

The levels of the different PFK isoenzymes and their corresponding transcripts in various tissues are in reasonable concurrence (14–16,20–22). This observation suggests that the tissue distribution of the PFK isoenzymes may be determined, at least in part, by the level of transcription of the PFK genes. For the HPFKM gene, the highest abundance of transcripts was observed in muscle with intermediate levels of transcripts in kidney, lung and aorta (14,15). The very low levels of HPFKM transcripts in liver and peripheral blood mononuclear cells were detectable only by reverse transcription and subsequent polymerase chain reaction (PCR) analysis (14). In an attempt to understand the regulatory mechanisms controlling the tissue-specific expression of the HPFKM gene, this gene has been cloned and characterized (5–7,12,13,21,23,24). From this analysis, it is apparent that the expression of the HPFKM gene is controlled by two promoters, p1 and p2, separated by ~3-kbp (12,13). The HPFKM type A and B RNAs initiating from the p2 promoter and the HPFKM type C RNAs initiating from the p1 promoter encode identical polypeptides (12,13,23). However, the reason transcription from the HPFKM gene is regulated by two promoters is not known. It has been suggested that the two promoters might regulate expression of the HPFKM gene in a tissue-specific manner. Consistent with this proposal, it has been shown that the HPFKM type C RNA is relatively ubiquitously expressed whereas expression of the HPFKM type A and B RNAs appears to be restricted to muscle tissue (21). However,

*To whom correspondence should be addressed

in cell culture it appears that the HPFKM p2 promoter which directs the expression of the type A and B RNAs has a broad cell-type-specificity and binds several ubiquitous transcription factors (13).

Since the initial characterization of the functional properties of the HPFKM p2 promoter indicated that this promoter was transcriptionally active in a variety of cell types, it was of interest to investigate the transcriptional properties of the HPFKM p1 promoter. In this study, the HPFKM p1 promoter has been cloned and sequenced. The transcription initiation site for the HPFKM C type RNA present in HeLa cells has been determined and shown to be positioned between two Sp1 transcription factor binding sites. This unusual arrangement of Sp1 transcription factor binding sites and the major transcription initiation site appears to represent a novel promoter configuration.

MATERIALS AND METHODS

Plasmid constructions

The various steps in the cloning of the plasmid constructs used in the transfection experiments were performed by standard techniques (25). The HPFKM p1 promoter sequences in these constructs were derived from the plasmid, HPFKMp1(−655/+78)LUC, which contains the HPFKM sequences from −655 to +78 relative to the proposed transcription initiation site (+1) for the HPFKM type C RNA cloned into the plasmid, p19DLUC (26). The HPFKM p1 promoter is located directly 5' to the promoterless firefly luciferase (LUC) reporter gene such that the expression of the LUC gene is governed by this promoter. The HPFKM p1 promoter deletions were generated from the HPFKMp1(−655/+78)LUC plasmid by appropriate restriction endonuclease digestion, *Bal31* nuclease digestion, or PCR amplification. The HPFKM p1 promoter DNA fragment from −655 to +78 was generated by PCR using genomic DNA derived from the human cervical carcinoma cell line, HeLa S3, and the oligonucleotides cggatccgTGCTTCACTGCCAGAATCCCGC (coordinates −655 to −634) and CTTTGTAGTCTTACGCTCTCCTCCGCCGCTGC (coordinates +78 to +49) (12). The lower case letters indicate the linker sequence present in the oligonucleotide used for the PCR amplification of the HPFKM p1 promoter. The HPFKMp1(−655/+78)LUCmt and HPFKMp1(−6/+78)LUCmt plasmids containing clustered point mutations were generated by a two-step PCR method using an appropriate mutant oligonucleotide to make the desired mutation within the HPFKM p1 promoter sequence (27). The PCR-generated mutant HPFKM p1 promoter fragments were subsequently cloned into the luciferase vector p19DLUC. The HPFKMp1(−655/+78)LUCmt and HPFKMp1(−6/+78)LUCmt plasmids are derivatives of the HPFKMp1(−655/+78)LUC and HPFKMp1(−6/+78)LUC plasmids, respectively, and contain a 10-nucleotide mutation spanning the Sp1 site 8 promoter sequence located at +12 to +21. All PCR products and deletion break-points generated by *Bal31* nuclease digestion were confirmed by dideoxynucleotide sequencing (28).

Cells and transfections

The human cervical carcinoma cell line, HeLa S3, was grown in Dulbecco's modified Eagle's medium containing 4.5 mg of glucose per ml and 10% calf serum at 37°C in 5% CO₂/air. Transfections were performed as described previously (26,29,30).

The transfected DNA mixture comprised 15 µg of a LUC plasmid and 1.5 µg of pSV2CAT (31) which served as an internal control for transfection efficiency. pSV2CAT directs the expression of the chloramphenicol acetyltransferase (CAT) gene using the SV40 early promoter.

The *Drosophila* Schneider line-2 cells (SL2) were grown in Schneider's *Drosophila* media (Gibco Laboratories) containing 10% heat-inactivated fetal calf serum at 25°C. The SL2 cells were transfected as described (32) with a DNA mixture comprising 1 µg of a LUC plasmid, 200 ng of the Sp1 expression vector, pPacSp1 (32), or 200 ng of the control expression vector, pPacU (32), and 10 ng of p5CCAT (33). p5CCAT directs the expression of the chloramphenicol acetyltransferase (CAT) gene using the *Drosophila* actin 5C promoter and served as an internal control for transfection efficiency. Cell extracts were prepared 40–48 h after transfection and assayed for luciferase and chloramphenicol acetyltransferase activity as previously reported (26,30,34).

Nuclear extracts, gel retardation analysis and DNase I footprinting

Nuclear extracts were prepared from HeLa S3 cells essentially as described (35). Gel retardation analysis was performed as described previously (36). Gel retardation supershift analysis was performed by incubating the HeLa S3 nuclear extract with 2 µl of a rabbit affinity purified polyclonal antibody raised against residues 520 to 538 of the human Sp1 protein (37) (Santa Cruz Biotechnology) for 1 h at 0°C prior to the addition of the ³²P-labelled double-stranded oligonucleotide. The oligonucleotide pairs used to generate the p1A to p1H and Sp1 double-stranded oligonucleotides were TCGAAATCCCCGCCCC and TCGAGGGGCGGGATT (oligo p1A, coordinates −641 to −630), TCGACATCCCTCCCCA and TCGATGGGGAGGGATG (oligo p1A3', coordinates −631 to −619), TCGAGACACGCCCTTC and TCGAGAAGGGCGGTGTC (oligo p1B, coordinates −593 to −581), TCGAGGGGAGGAGG and TCGACCTCCTCCCC (oligo p1C5', coordinates −130 to −121), TCGAGAGGAGGAAG and TCGACTTCCTCCTC (oligo p1C3', coordinates −127 to −118), TCGAGGAGGCGGAGCC and TCGAGGCTCCGCCTCC (oligo p1D, coordinates −97 to −83), TCGAGGAGGCGGACTT and TCGAAAGTCCGCC-TCC [oligo p1D(alt), coordinates −97 to −82], TCGATCCCCCTCCCTC and TCGAGAGGGAGGGGGA (oligo p1E, coordinates −38 to −26), TCGAGGCTCCTCCCCA and TCGATGGGGAGGAGCC (oligo p1F, coordinates −13 to −1), TCGAATCACCCCTC and TCGAGAGGGGTGAT (oligo p1G5', coordinates +9 to +18), TCGAACCCCTCCCC and TCGAGGGGAGGGGT (oligo p1G3', coordinates +12 to +21), TCGAAGCGGCGGAGGA and TCGATCCTCCGCCGCT (oligo p1H, coordinates +51 to +62), and TCGAGGGGCGGGGC and TCGAGCCCCGCCCC [oligo Sp1 (38)]. The DNase I footprinting reactions were performed as reported (30,38).

Isolation of cytoplasmic RNA and primer extension analysis

Cytoplasmic RNA was isolated from HeLa S3 cells as described (25). Primer extension analysis was performed as described (39) using 50 µg of RNA and 2 ng of the oligonucleotide primer C-TTTTGTAGTCTTAGCTCTCCTCCGCCGCTGC (coordinates +78 to +49) (12). The mixture was incubated for 2 h at 42°C with 200 units of the Moloney murine leukemia virus reverse transcriptase.

RESULTS

Sequence of the HPFKM p1 promoter and mapping the transcription initiation site

The HPFKM p1 promoter region from -655 to +78 was PCR amplified using HeLa S3 genomic DNA and oligonucleotide primers based on the previously reported promoter sequence (12). The promoter sequence was determined for two clones from independent PCR reactions. The sequences were identical to each other but differed from the previously reported sequence at four positions (Fig. 1). At nucleotide coordinates -316 the previously reported sequence had an additional A whereas it lacked the C and GC residues at nucleotide coordinates 579, -86 and -85. It is possible that the differences between these two sequences represent polymorphisms in the HPFKM p1 promoter.

The initiation site of the HPFKM C type RNA transcribed from the HPFKM p1 promoter was determined using HeLa S3 cytoplasmic RNA and primer extension analysis (Fig. 2). A single predominant transcription initiation site was identified. The location of this transcription initiation site was in good agreement with the most upstream initiation site mapped previously using total human muscle RNA. In the case of the muscle RNA, three RNA initiation sites were mapped to -3, +25 and +36, respectively (12). In addition, the identified transcription initiation

site is in good agreement with the sequence of the longest cDNA sequence corresponding to the HPFKM C type RNA which terminated at +15 (5). Therefore it appears likely that the initiation site for transcription identified in HeLa S3 RNA represent the predominant site utilized in this cell line.

Examination of the HPFKM p1 promoter sequence had previously suggested that the Sp1 transcription factor might interact with two sequence elements of this promoter (12). One of these sequence elements, region D (Fig. 1), is located in the region of the promoter where differences between the two promoter sequences were observed. Further examination of the nucleotide sequence of the HPFKM p1 promoter indicated that there were several regions of this promoter which showed homology to the Sp1 consensus sequence (Table 1). These observations suggested that the role of the Sp1 transcription factor in the regulation of the expression of the HPFKM gene from the p1 promoter should be examined.

Identification of the Sp1 transcription factor binding sites in the HPFKM p1 promoter

The location of the Sp1 transcription factor binding sites in the HPFKM p1 promoter was examined by DNase I footprinting analysis using purified recombinant Sp1 protein (Fig. 3).

Figure 1. Nucleotide sequence of the HPFKM p1 promoter region (EMBL accession no. Z34905). The sequence is numbered from the transcription initiation site (+1) identified by primer extension analysis of HeLa S3 RNA (Fig. 2). The location of an additional A residue in the previously reported HPFKM p1 promoter sequence (12) is indicated (\$). Nucleotides absent from the previously reported sequence (12) are indicated with asterisks. Lower case letters represent nucleotides protected from DNase I cleavage by the Sp1 transcription factor. The location of the Sp1 sites was determined from Fig. 3. The locations of the double-stranded oligonucleotides spanning the Sp1 sites (p1A, p1A3', p1B, p1C3', p1C5', p1D, p1D(alt), p1E, p1F, p1G5', p1G3' and p1H) used in the gel retardation analysis are indicated. The proposed Sp1 recognition sites 1 to 8 are underlined.

Table 1. Comparison between the Sp1 consensus sequence and the HPFKM p1 promoter Sp1 binding sites and related sequences

Sp1 site	Sequence ^a	Identity ^b
Sp1 consensus ^c	ATTA TA GCCCCGCCCC	
p1A region/site 1 (N ^d)	ATCCCGCCCC	10/10
p1A3' region/site 2 (N ^d)	ATCCCTCCCC	9/10
p1B region/site 3 (N ^d)	GaCACGCCct	8/10
p1C5' region/site 4 (C ^d)	cCTCCtCCCC	8/10
p1C3' region (C ^d)	cTTCCtCCTC	8/10
p1D region/site 5 (C ^d)	GCTCCGCCTC	10/10
p1D(alte) region (C ^d)	AgTCCGCCTC	9/10
p1E region/site 6 (N ^d)	cCCCCtCCCC	7/10
p1F region/site 7 (N ^d)	GCTCCtCCCC	9/10
p1G5' region (N ^d)	ATCACcCCTC	9/10
p1G3' region/site 8 (N ^d)	ACCCctCCCC	9/10
p1H region (C ^d)	cTCCGCCgC	8/10

^aLower case letters indicate the nucleotides which differ from the Sp1 consensus sequence.

^bIdentity indicates the number of nucleotides which are identical to the 10 nucleotide Sp1 consensus sequence.

^cFrom reference (38).

^dC and N indicate that the sequence is derived from the coding or noncoding strand of the HPFKM p1 sequence, respectively.

^eFrom the HPFKM p1 sequence reported previously (12).

Figure 2. Primer extension analysis of the HPFKM C type RNA expressed in HeLa S3 cells. Primer extension analysis was performed using 50 μ g of HeLa S3 cytoplasmic RNA (lane 1) or 50 μ g of yeast tRNA (lane 2) and the oligonucleotide primer CTTTAgTCTTAGCTCTCCTCCGCGCTGC (coordinates +78 to +49) (Fig. 1). An arrow indicates the predominant 78 nucleotide primer extension product. An unrelated DNA sequence (GATC) is adjacent to lane 1 and was used to determine the size of the primer extension product.

Protection of two regions (sites 1 to 3) of the noncoding and five regions (sites 4 to 8) of the coding strand was observed. The protection of the site 2 footprint on the noncoding strand was rather poor. This may have resulted from the proximity of sites 1 and 2 which were separated by a single nucleotide (Fig. 1). The Sp1 footprint sites 1 to 8 span nucleotide coordinates -641 to -616 (sites 1 and 2), -596 to -577 (site 3), -136 to -107 (site 4), -102 to -82 (site 5), -45 to -24 (site 6), -19 to +2 (site 7) and +4 to +28 (site 8), respectively (Fig. 1). Consistent with these observations, it was noted that the central ten nucleotides of the sites 1 to 8 footprints showed between 70 and 100% identity with the Sp1 consensus sequence (38) suggesting these sequences may represent the Sp1 recognition elements in the HPFKM p1 promoter (Table 1). However, in the case of the Sp1 binding sites 4 and 8, it was not clear exactly which nucleotide sequence was being recognized by the Sp1 protein because two sequence elements within the footprinting region showed homology to the Sp1 consensus sequence (Table 1).

Figure 3. (A) DNase I footprinting analysis of the noncoding strand of the HPFKM p1 promoter. The 733 nucleotide HPFKM p1 promoter DNA fragment from -655 to +78 was 5' end labelled at -655 and incubated (lane 1) in the presence of 4 units of the Sp1 transcription factor, or (lane 2) in the absence of factor, before DNase I digestion. An unrelated DNA sequence (GATC) is adjacent to lane 1 and was used as a size standard to localize the Sp1 sites 1 to 3 footprints (Fig. 1). (B) DNase I footprinting analysis of the coding strand of the HPFKM p1 promoter. The 733 nucleotide HPFKM p1 promoter DNA fragment from -655 to +78 was 5' end labelled at +78 and incubated (lane 1) in the presence of 4 units of the Sp1 transcription factor, or (lane 2) in the absence of factor, before DNase I digestion. An unrelated DNA sequence (GATC) is adjacent to lane 1 and was used as a size standard to localize the Sp1 sites 4 to 8 footprints (Fig. 1).

In an attempt to resolve the issue of the sequence elements recognized by the Sp1 transcription factor and to demonstrate that these regulatory elements of the HPFKM p1 promoter recognized the Sp1 transcription factor in nuclear extracts, gel retardation analysis of these individual regulatory elements was performed. Initially, double-stranded oligonucleotides corresponding to the putative HPFKM p1 promoter Sp1 transcription factor sites were examined for their ability to interact with proteins present in HeLa S3 nuclear extracts (Fig. 4). This analysis demonstrated that the p1A, p1A3', p1B, p1C5', p1D, p1E, p1F and p1G3' regions of the HPFKM p1 promoter specifically interacted with a HeLa S3 nuclear protein, forming a complex which comigrated with the complex formed between the Sp1 consensus sequence and a HeLa S3 nuclear protein. In addition, all the complexes formed between the double-stranded oligonucleotides and the HeLa S3 nuclear extract were inhibited by the use of unlabeled Sp1 double-stranded oligonucleotide as competitor (data not shown). This indicated that the Sp1 transcription factor was specifically binding the p1A, p1A3', p1B, p1C5', p1D, p1E, p1F and p1G3' regions of the HPFKM p1 promoter. The p1D(alt) region examined in this analysis was derived from the previously described HPFKM p1 promoter sequence and does not appear to specifically bind the Sp1 transcription factor.

The nature of the DNA-protein complexes formed between the p1A, p1A3', p1B, p1C5', p1D, p1E, p1F, p1G3' and Sp1 ³²P-labelled double-stranded oligonucleotides and proteins in the HeLa S3 nuclear extract was investigated using an antibody against the Sp1 protein (37). The major gel retardation complexes were supershifted by the anti-Sp1 antibody (Fig. 4). This indicated that almost all of the complexes contained the Sp1 transcription factor. A minor complex which migrates slightly faster than the major complex can be observed with some of the oligonucleotide probes. This complex may indicate the presence of additional nuclear factors in the HeLa S3 cells which can associate with

Sp1 recognition sequences as has been reported previously (40–43).

Regulation of transcription from the HPFKM p1 promoter by the Sp1 transcription factor

Functional studies using the HPFKM p1 promoter Sp1 binding sites indicated that they are capable of mediating Sp1-dependent transcription from a minimal promoter (data not shown). However, the role these Sp1 sites have in regulating the level of transcription from the HPFKM p1 promoter was not apparent. To investigate this issue the HPFKM p1 promoter construct, HPFKMp1(−655/+78)LUC, was examined for its activities in the presence or absence of exogenously expressed Sp1 using transient transfection experiments in *Drosophila* SL2 cells. In the absence of exogenously expressed Sp1 the HPFKM p1 promoter was transcriptionally inactive whereas the promoter was transactivated by exogenously expressed Sp1 (Fig. 5). Subsequently a 5′ deletion series of the HPFKM p1 promoter was examined for promoter activity in the presence of exogenously expressed Sp1 to determine the minimal promoter element necessary for Sp1-dependent transcription (Fig. 5). The deletion of sequences to +12 [construct HPFKMp1(+12/+78)LUC] did not eliminate Sp1-dependent transcription from the HPFKM p1 promoter indicating that Sp1 sites 1 to 7 are not essential for promoter activity in this system although it appears that they do contribute to the total promoter activity. The deletion of the HPFKM p1 promoter sequence to +20 [construct HPFKMp1(+20/+78)LUC] essentially eliminated Sp1-dependent transcriptional activity, indicating the importance of the Sp1 site 8 for this effect. However mutation of the Sp1 site 8 in the context of the larger HPFKM p1 promoter sequence demonstrates that the loss of the Sp1 site 8 can be compensated for by the other Sp1 sites in the promoter [compare HPFKMp1(−655/+78)LUC and HPFKMp1(−655/+78)LUCmt]. In the context of the minimal promoter mutation of the Sp1 site 8 eliminates promoter

Figure 4. Gel retardation and complex supershift analysis of the HPFKM p1 promoter regulatory regions. The ³²P-labelled double-stranded oligonucleotide probes p1A, p1A3', p1B, p1C5', p1C3', p1D(alt), p1D, p1E, p1F, p1G5', p1G3', p1H and Sp1 were analyzed for their ability to form supershifted complexes with DNA-binding proteins present in HeLa S3 nuclear extracts. The rabbit anti-Sp1 antibody (37) was incubated with the extract analyzed in lanes 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 and 26. An irrelevant rabbit antibody was incubated with the extract analyzed in lane 27. The unlabeled Sp1 consensus sequence double-stranded oligonucleotide was used as competitor DNA in lanes 28 and 29 to demonstrate the specificity of the observed complexes.

Figure 5. Deletion analysis of the HPFKM p1 promoter. The arrow indicates the position and direction of transcription from the HPFKM p1 promoter. The boxes represent the mutation of Sp1 site 8 located between +12 and +21 in the HPFKM p1 promoter. Horizontal lines indicate the HPFKM p1 promoter sequences present in the various luciferase constructs. The construct HPFKMp1(-655/+78)LUC contains the HPFKM p1 promoter sequence from -655 to +78 located directly 5' to the promoterless firefly luciferase (LUC) reporter gene such that the expression of the LUC gene is governed by the HPFKM p1 promoter. Similarly, the designations of the other 5' deletions indicate the extent of the promoter region present in each construct. The relative activity of the HPFKM p1 promoter in the presence of exogenously expressed Sp1 derived from the expression vector, pPacSp1, determined in *Drosophila* SL2 cells is indicated (SL2). In the absence of exogenously expressed Sp1, no transcription is observed from any of the HPFKM p1 promoter constructs examined. The internal control used to correct for transfection efficiencies in SL2 cells was pSCCAT. The relative activity of the HPFKM p1 promoter in HeLa S3 cells is shown (HeLa). The internal control used to correct for transfection efficiencies in HeLa cells was pSV2CAT. The relative luciferase activities were calculated as described (26). The mean values for the relative luciferase activities are shown with the standard errors in parentheses. The relative activity of the HPFKM(-655/+78)LUC construct was designated 1.00 and represents an average of 200 000 and 80 000 light units in *Drosophila* SL2 and HeLa S3 cells, respectively.

activity [compare HPFKMp1(-6/+78)LUC and HPFKMp1(-6/+78)LUCmt] as expected based on the 5'-deletion series results (Fig. 5).

Using transient transfection experiments in the human cervical carcinoma cell line, HeLa S3, the possible role of the Sp1 sites in determining the level of transcription from the HPFKM p1 promoter was also examined (Fig. 5). The deletion of sequence elements from -127 to -105 [compare constructs HPFKMp1(-127/+78)LUC to HPFKMp1(-104/+78)LUC], -61 to -14 [compare constructs HPFKMp1(-61/+78)LUC to HPFKMp1(-13/+78)LUC] and -12 to -7 [compare HPFKMp1(-12/+78)LUC and HPFKMp1(-6/+78)LUC] resulted in a two- to three-fold reduction in the level of transcription from the HPFKM p1 promoter. Although it is not possible to determine the nature of the transcription factor(s) responsible for modulating promoter activity from this analysis, it is apparent that the deletion of sequences -61 to -14 and -12 to -7 removed the Sp1 sites 6 and 7, respectively. Therefore it is possible that the loss of Sp1 binding to these regions of the promoter are responsible for the observed modulation of promoter activity. It also appears that Sp1 sites 1 to 5 contribute very little to the level of transcription from the

Figure 6. Nucleotide sequence of the HPFKM p1 promoter region around the transcription initiation site. The overlapping regions homologous to the Sp1 (38) and YY1 (48) transcription factor consensus sequences are indicated.

HPFKM p1 promoter in this transient transfection system. However, it appears that the Sp1 site 8 is an essential element of the minimal HPFKM p1 promoter since deletion or mutation [construct HPFKMp1(-6/+78)LUCmt] of the sequence between +12 and +19 resulted in the complete loss of promoter activity (Fig. 5). As observed in the *Drosophila* transfection system, the Sp1 site 8 only marginally influences the activity of the HPFKM p1 promoter in the context of the larger promoter sequence [compare HPFKMp1(-655/+78)LUC and HPFKMp1(-655/+78)LUCmt]. This suggests the Sp1 sites in this promoter display functional redundancy. These results indicate that the Sp1

transcription factor has an important role in determining the level of transcription from the HPFKM p1 promoter. In addition the absence of a requirement for promoter sequences upstream of the major transcription initiation site to observe low levels of transcription [construct HPFKMp1(+12/+78)LUC, Fig. 5] suggests that the minor transcription initiation sites located at approximately +25 and +36 (12) are also utilized in HeLa S3 cells.

DISCUSSION

The HPFKM p1 promoter region has been cloned and sequenced, and the role of the Sp1 transcription factor in determining the level of transcription from this promoter has been examined. The sequence of the HPFKM p1 promoter was shown to be identical to the previously reported sequence (12) except the HeLa S3 DNA sequence included three additional nucleotides and was deleted for one nucleotide (Fig. 1). A single major transcription initiation site was identified by primer extension using HeLa S3 RNA which was in good agreement with one of the major initiation sites identified in muscle RNA (12) and the sequence of a HPFKM placental cDNA sequence (5). Further analysis of the HPFKM p1 promoter sequence indicated multiple regions which were homologous to the Sp1 consensus sequence (Table 1) (12,38), suggesting this transcription factor might be an important regulator of the activity of this promoter.

The possibility that the HPFKM p1 promoter might contain Sp1 transcription factor binding sites was investigated by DNase I footprinting and gel retardation analysis. DNase I footprinting analysis identified eight sites in the promoter which bound the purified recombinant Sp1 protein (Fig. 3). Gel retardation analysis verified that these eight sites defined by regions p1A (site 1), p1A3' (site 2), p1B (site 3), p1C5' (site 4), p1D (site 5), p1E (site 6), p1F (site 7) and p1G3' (site 8) bound the Sp1 transcription factor present in HeLa S3 nuclear extracts (Figs 1 and 4).

The role of the Sp1 transcription factor in modulating the level of transcription from the HPFKM p1 promoter was examined using transient transfection assays in *Drosophila* SL2 and HeLa S3 cells. In both systems, the sequence from +12 to +19 was shown to represent an essential element of the minimal HPFKM p1 promoter but could be mutated in the context of a larger promoter region without much effect on promoter activity. This region of the HPFKM p1 promoter contains Sp1 site 8 located in region p1G3'. The location of this Sp1 recognition element from +12 to +21 represents a novel position for an Sp1 sites relative to the transcription initiation site especially since the Sp1 site 8 is located within the 5'-untranslated region of the HPFKM C type RNA. The binding sites for the Sp1 transcription factor are usually located upstream from the transcription initiation site, often distal to a TATA-box element located at approximately -30 relative to the RNA start site (44).

The lack of a TATA-box element in the HPFKM p1 promoter suggested that an 'initiator' like-element (45-47) located around the transcription initiation site might be involved in regulating transcription from this promoter. Examination of the sequence around the mapped transcription initiation site reveal a sequence element identical to the binding site for the transcription factor YY1 (Fig. 6) (48). The transcription factor YY1 (49), also known as NF-E1 (50), δ (51), UCRBP (52) and F-ACT1 (48), is a multi-functional transcription factor which has been identified as binding the transcription initiation sites of the adeno-associated virus type

2 P5 promoter (49) and the cytochrome c oxidase subunit Vb gene (53). The YY1 recognition sequence element has been shown to be necessary and sufficient for accurate basal transcription of these genes (53,54). In the case of the cytochrome c oxidase subunit Vb gene the YY1 recognition element overlaps an Sp1 recognition element in a manner very similar to that observed in the HPFKM p1 promoter (53). The similar proximity of the overlapping Sp1 and YY1 recognition sequences to the transcription initiation site in both the cytochrome c oxidase subunit Vb and HPFKM p1 promoters, and the ability of these two transcription factors to physically interact (55) implies that this arrangement of transcription factors may be important in determining the position of initiation of transcription from these promoters.

The observation that the HPFKM p1 promoter has multiple functional Sp1 transcription factor binding sites has implications for the expression of the HPFKM gene from this promoter. As Sp1 is a ubiquitous transcription factor (32,38,56), it may account for the relative broad cell-type and tissue distribution of the HPFKM RNA (14,21). However, since the HPFKM p2 promoter is transcriptionally active in a variety of cell lines and the HPFKM p2 promoter also contains multiple functional Sp1 transcription factor binding sites (13), it is unclear what role this factor might have in controlling the level of the HPFKM A, B and C type RNAs in various cells and tissues. The HPFKM gene presumably has two promoters so it can differentially regulate the expression of the HPFKM gene, possibly in a tissue-specific manner or in response to a particular stimulus. The role of the Sp1 transcription factor at the two promoters may be to interact with additional factors which have highly restricted expression patterns so as to mediate the appropriate transcription of the HPFKM gene. Further analysis of the expression of the different types of HPFKM RNAs in various tissues and under different physiological conditions may resolve these issues.

ACKNOWLEDGEMENTS

We are grateful to Robert Tjian (University of California, Berkeley) for the plasmids pPacU and pPacSp1, and Beverley Bond-Matthews (University of California, Los Angeles) for the plasmid p5CCAT. This work was supported by Public Health Service grant DK33445 from the National Institutes of Health and the Stein Endowment Fund.

REFERENCES

1. Bloxham, D.P. and Lardy, H.A. (1973) in *The enzymes*, Boyer, P.D. Ed. pp. 239-278, Academic Press, New York.
2. Uyeda, K. (1979) in *Advances in enzymology*, Meister, A. Ed. pp. 193-244, John Wiley and Sons, New York.
3. Dunaway, G.A. (1983) *Mol. Cell. Biochem.* **52**, 75-91.
4. Elson, A., Levanon, D., Brandeis, M., Dafni, N., Bernstein, Y., Danciger, E. and Groner, Y. (1990) *Genomics* **7**, 47-56.
5. Sharma, P.M., Reddy, G.R., Vora, S., Babiorek, B.M. and McLachlan, A. (1989) *Gene* **77**, 177-183.
6. Nakajima, H., Noguchi, T., Yamasaki, T., Kono, N., Tanaka, T. and Tarui, S. (1987) *FEBS Lett.* **223**, 113-116.
7. Valdez, B.C., Chen, Z., Sosa, M.G., Younathan, E.S. and Chang, S.H. (1989) *Gene* **76**, 167-169.
8. Simpson, C.J. and Fothergill-Gilmore, L.A. (1991) *Biochem. Biophys. Res. Commun.* **180**, 197-203.
9. Levanon, D., Danciger, E., Dafni, N. and Groner, Y. (1986) *Biochem. Biophys. Res. Commun.* **141**, 374-380.
10. Levanon, D., Danciger, E., Dafni, N. and Groner, Y. (1987) *Biochem. Biophys. Res. Commun.* **147**, 1182-1187.

11. Levanon, D., Danciger, E., Dafni, N., Bernstein, Y., Elson, A., Moens, W., Brandeis, M. and Groner, Y. (1989) *DNA* **8**, 733–743.
12. Yamasaki, T., Nakajima, H., Kono, N., Hotta, K., Yamada, K., Imai, E., Kuwajima, M., Noguchi, T., Tanaka, T. and Tarui, S. (1991) *Gene* **104**, 277–282.
13. Vaisanen, P.A., Reddy, G.R., Sharma, P.M., Kohani, R., Johnson, J.L., Raney, A.K., Babior, B.M. and McLachlan, A. (1992) *DNA Cell Biol.* **11**, 461–470.
14. Sharma, P.M., Reddy, G.R., Babior, B.M. and McLachlan, A. (1990) *J. Biol. Chem.* **265**, 9006–9010.
15. Gehnrich, S.C., Gekakis, N. and Sul, H.S. (1988) *J. Biol. Chem.* **263**, 11755–11759.
16. Vora, S. (1982) in *Isozymes: Current topics in biological and medical research*, Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. Eds pp. 119–167, Alan R. Liss, Inc., New York.
17. Tsai, M.Y., Gonzalez, F. and Kemp, R.G. (1975) in *Isozymes*, Market, C.L. Ed. pp. 819–835, Academic Press, New York.
18. Dunaway, G.A., Kasten, T.P., Sebo, T. and Trapp, R. (1988) *Biochem. J.* **251**, 677–683.
19. Kahn, A., Meienhofer, M.C., Cottreau, D., Lagrange, J.L. and Dreyfus, J.C. (1979) *Hum. Genet.* **48**, 93–108.
20. Vora, S. (1983) in *Isozymes: Current topics in biological and medical research*, Rattazzi, M.C., Scandalios, J.G. and Whitt, G.S. Eds pp. 3–23, Alan R. Liss, Inc., New York.
21. Nakajima, H., Kono, N., Yamasaki, T., Hamaguchi, T., Hotta, K., Kuwajima, M., Noguchi, T., Tanaka, T. and Tarui, S. (1990) *Biochem. Biophys. Res. Commun.* **173**, 1317–1321.
22. Hotta, K., Nakajima, H., Yamasaki, T., Hamaguchi, T., Kuwajima, M., Noguchi, T., Tanaka, T., Kono, N. and Tarui, S. (1991) *Eur. J. Biochem.* **202**, 293–298.
23. Nakajima, H., Yamasaki, T., Noguchi, T., Tanaka, T., Kono, N. and Tarui, S. (1990) *Biochem. Biophys. Res. Commun.* **166**, 637–641.
24. Vora, S., Hong, F. and Olender, E. (1986) *Biochem. Biophys. Res. Commun.* **135**, 615–621.
25. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: a laboratory manual*, Cold Spring Harbor Laboratory Press, New York.
26. Raney, A.K., Milich, D.R., Easton, A.J. and McLachlan, A. (1990) *J. Virol.* **64**, 2360–2368.
27. Landt, O., Grunert, H.P. and Hahn, U. (1990) *Gene* **96**, 125–128.
28. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl Acad. Sci. USA* **74**, 5463–5467.
29. McLachlan, A., Milich, D.R., Raney, A.K., Riggs, M.G., Hughes, J.L., Sorge, J. and Chisari, F.V. (1987) *J. Virol.* **61**, 683–692.
30. Raney, A.K., Milich, D.R. and McLachlan, A. (1989) *J. Virol.* **63**, 3919–3925.
31. Gorman, C.M., Moffat, L.F. and Howard, B.H. (1982) *Mol. Cell. Biol.* **2**, 1044–1051.
32. Courey, A.J. and Tjian, R. (1988) *Cell* **55**, 887–898.
33. Bond-Matthews, B. and Davidson, N. (1988) *Gene* **62**, 289–300.
34. De Wet, J.R., Wood, K.V., DeLuca, M., Helinski, D.R. and Subramani, S. (1987) *Mol. Cell. Biol.* **7**, 725–737.
35. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.* **11**, 1475–1489.
36. Raney, A.K., Easton, A.J., Milich, D.R. and McLachlan, A. (1991) *J. Virol.* **65**, 5774–5781.
37. Jackson, S.P., MacDonald, J.J., Lees-Miller, S. and Tjian, R. (1990) *Cell* **63**, 155–165.
38. Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) *Science* **234**, 47–52.
39. Santoro, C., Marone, M., Ferrone, M., Costanzo, F., Colombo, M., Minganti, C., Cortese, R. and Silengo, L. (1986) *Nucleic Acids Res.* **14**, 2863–2876.
40. Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M. and Fujii-Kuriyama, Y. (1992) *EMBO J.* **11**, 3663–3671.
41. Kingsley, C. and Winoto, A. (1992) *Mol. Cell. Biol.* **12**, 4251–4261.
42. Hagen, G., Müller, S., Beato, M. and Suske, G. (1992) *Nucleic Acids Res.* **20**, 5519–5525.
43. Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H. and Fujii-Kuriyama, Y. (1993) *Nucleic Acids Res.* **21**, 1527–1532.
44. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
45. Smale, S.T., Schmidt, M.C., Berk, A.J. and Baltimore, D. (1990) *Proc. Natl Acad. Sci. USA* **87**, 4509–4513.
46. Smale, S.T. and Baltimore, D. (1989) *Cell* **57**, 103–113.
47. Weis, L. and Reinberg, D. (1992) *FASEB J.* **6**, 3300–3309.
48. Lee, T.-C., Shi, Y. and Schwartz, R.J. (1992) *Proc. Natl Acad. Sci. USA* **89**, 9814–9818.
49. Shi, Y., Seto, E., Chang, L.S. and Shenk, T. (1991) *Cell* **67**, 377–388.
50. Park, K. and Atchison, M.L. (1991) *Proc. Natl Acad. Sci. USA* **88**, 9804–9808.
51. Hariharan, N., Kelley, D.E. and Perry, R.P. (1991) *Proc. Natl Acad. Sci. USA* **88**, 9799–9803.
52. Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B.Z., Apella, E. and Ozato, K. (1992) *Mol. Cell. Biol.* **12**, 38–44.
53. Basu, A., Park, K., Atchison, M.L., Carter, R.S. and Avadhani, N.G. (1993) *J. Biol. Chem.* **268**, 4188–4196.
54. Seto, E., Shi, Y. and Shenk, T. (1991) *Nature* **354**, 241–245.
55. Lee, J.S., Galvin, K.M. and Shi, Y. (1993) *Proc. Natl Acad. Sci. USA* **90**, 6145–6149.
56. Saffer, J.D., Jackson, S.P. and Annarella, M.B. (1991) *Mol. Cell. Biol.* **11**, 2189–2199.