Regulation of Transcription of the TATA-less Human Complement Component C4 Gene


J Immunol 1998; 160:4353-4360; ;
http://www.jimmunol.org/content/160/9/4353

References

This article cites 49 articles, 22 of which you can access for free at:
http://www.jimmunol.org/content/160/9/4353.full#ref-list-1

Subscription

Information about subscribing to The Journal of Immunology is online at:
http://jimmunol.org/subscription

Permissions

Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts

Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts
Regulation of Transcription of the TATA-less Human Complement Component C4 Gene


The 5'-sequences flanking the human complement component C4 genes (C4A and C4B) have been analyzed for their ability to direct expression of a reporter gene in cell lines that constitutively express or do not express C4. No difference in the level of reporter gene expression was detected in cells transfected with C4A- or C4B-specific constructs. A series of reporter constructs containing progressively truncated C4 promoter fragments transfected into the hepatocyte Hep G2 cell line, identified the sequence contained within the region −178 to −39 as that associated with maximal reporter gene expression. This region contains consensus binding motifs for nuclear factor 1 (−110 to −97), Sp1 (−57 to −49), and three basic helix-loop-helix (−137 to −132, −98 to −93, and −78 to −73)-like transcription factors. Electromobility shift assays and DNase I footprinting analysis showed specific DNA-protein interactions of the C4 promoter at the nuclear factor 1, two E box (−98 to −93 and −78 to −73), and Sp1 binding domains. Site-directed mutagenesis of the Sp1 binding site resulted in total abrogation of reporter gene expression and mutation of the E box (−78 to −73) resulted in a 8-fold reduction in expression. We conclude that the Sp1 binding site at position −57 to −49 is critical for accurately initiated, basal transcription of C4. The Journal of Immunology, 1998, 160: 4353–4360.

The fourth component of the complement system plays a pivotal role in the activation of the classical and recently described mannan-binding lectin (1) pathways as a constituent of the C3 convertase. Two C4 isotypes have been described in humans (C4A and C4B). These share >99% nucleotide identity (2, 3), and both generate C4b upon activation. C4b derived from C4A has been shown to opsonize protein-containing immune complexes more efficiently than that derived from C4B (4). This may explain the high prevalence of C4A null alleles in SLE patients (5), where failure to opsonize immune complexes in the periphery could lead to deposition and inflammation (reviewed in Ref. 6). It also raises the possibility that subtle defects in the regulation of C4 expression may have pathogenic relevance.

C4 is encoded by tandemly duplicated loci, located in the MHC class III region (7, 8). The nonhomologous duplication event that resulted in formation of these tandem loci also involved regions upstream of the genes (9). An equivalent duplication is seen in the mouse H-2 region, giving rise to the C4 isotypes, C4 and C4-sex-limited protein (10). There is significant identity of the sequence (>76%) up to position −150 between the human and mouse C4 promoter regions, and both share common consensus binding motifs for potential regulatory elements. Both human and murine C4 genes belong to an expanding group, transcribed by RNA polymerase II, which lack consensus TATA and CCAAT boxes. In the mouse, three sequence motifs, an initiator element (−1 to +12), an E box (−75 to −70), and a nuclear factor 1 (NF-1) (−112 to −87) binding site have been shown to be functionally important in directing high level, accurately initiated C4 expression (11, 12). These sites are conserved in the human, but to date there are no data regarding their role in the regulation of human C4 expression. We have previously demonstrated tissue-specific DNase I-hypersensitive sites immediately 5' of the C4A and C4B cap sites (13), suggesting that this region contains key transcriptional elements. A similar DNase I-hypersensitive site has also been identified in the mouse C4 promoter (14) which is consistent with functional data that have localized the transcription factor binding sites to the 5'-flanking regions within 150 bp of the transcription start site.

The liver is the main source of serum C4, and this is reflected in the high level of C4 mRNA and protein expression by the human hepatoma cell line Hep G2 (15, 16). Additional cell types that synthesize and secrete C4 include PBMC (17), glomerular epithelial cells (18), proximal tubular cells of the kidney (19), fibroblasts (20), and synoviocytes (21). During the acute phase response, serum concentration of C4 increases by ~50%. This is predominantly mediated by IFN-γ, which is the only cytokine to induce C4 expression in Hep G2 cells (16), II-1, II-6, and TNF-α having no effect (16, 22, 23). We have previously shown that in Hep G2 cells the major effector mechanism by which IFN-γ up-regulates C4 expression is through stabilization of C4 mRNA (24).

To define the cis-acting elements necessary for high level tissue-specific expression of human C4, we have determined the capacity of progressively truncated C4 promoter fragments to drive expression of a reporter gene in two cell lines, Hep G2 (hepatocyte) and HeLa (epithelial) cells. In vitro DNA-protein interactions were studied using DNase I footprinting and electromobility shift assays.
A promoterless β globin fragment was amplified using the PCR from a plasmid, pNPb1.1 (25), containing the entire β-globin gene. Primers Y1 (ggactctgcagGTGTTCACTAGCAACCTC, +322 to +393) and EGmut (5'-ggactctgcagGTGTTCACTAGCAACCTC, +1670 to +1653) were used to amplify a 1650-bp amplicon that spanned the β-globin gene (primer sequence complementary to the β-globin gene is in upper case), including 29 bp of the 5'-untranslated region and terminating 64 bp downstream of the polyadenylation site. This fragment was cloned via Pst I restriction sites in the primers sequence into the plasmid pBlueScript SK+ (Stratagene, Cambridge, U.K.). This reporter construct was termed pβ.

C4 5'-flanking regions were generated using the PCR from cosmid containing the C4 genes, as described previously (13). A series of seven 5' primers, progressively closer to the C4 cap site and a common 3' primer (CAGGTGCACCATGGTGTCTGTTTG, 5'-GGTGCCCCCACCACTCTGGGC, 5'-GGAGGAGCAGG (C4-1), 5'-GGTTTCCCAGC (C4-2), 5'-GACCTTCACGTGCCGACCTC (C4-3), 5'-GGCGCAAGAATGCACTGACCTCCCAC, 5'-TACCACCACATGCTTCTGCACCAC, 5'-GCTTCAAGTCCAATGATTCGGAG, 5'-AGTTTCCCACCACCCCGAGGAGCAGG, 5'-ACACCTAGTGCTTCTGCTTATTG, 5'-CGGTGAGCTTGGGAAAGTGAC, 5'-GCTGAGAGCAGG, 5'-GGTTTCCCAGC, 5'-GGAGGAGCAGG (C4-9), the E box at position 126 to 139 was used) and one 3' primer, progressively closer to the C4 cap site and a common 5' primer (CAGGTGCACCATGGTGTCTGTTTG, 5'-GGTGCCCCCACCACTCTGGGC, 5'-GGAGGAGCAGG (C4-1), 5'-GGTTTCCCAGC (C4-2), 5'-GACCTTCACGTGCCGACCTC (C4-3), 5'-GGCGCAAGAATGCACTGACCTCCCAC, 5'-TACCACCACATGCTTCTGCACCAC, 5'-GCTTCAAGTCCAATGATTCGGAG, 5'-AGTTTCCCACCACCCCGAGGAGCAGG, 5'-ACACCTAGTGCTTCTGCTTATTG, 5'-CGGTGAGCTTGGGAAAGTGAC, 5'-GCTGAGAGCAGG, 5'-GGTTTCCCAGC, 5'-GGAGGAGCAGG (C4-9)) was used to amplify C4 fragments, respectively, and were amplified using the primers described above. Primers were end-labeled using polynucleotide kinase with [γ32P]ATP, before use in the PCR. After phenol/chloroform extraction of the PCR amplicons, probe concentrations were determined by agarose gel electrophoresis. Binding reactions contained 1 to 2 ng of DNA probe (2 x 10^6 cpm) and 5 to 10 µg of nuclear extract in a 25-µl reaction containing 2 µg of poly(dI-dC)·poly(dI-dC) double-stranded DNA (Pharmacia Biotech, Milton Keynes, U.K.). In an initial round of PCR, 1.5 µM of each primer was used.

C4-1 (C4.1, AGGTCCAAGAATGCACTGACCTCCCAC, 5'-GGCGCAAGAATGCACTGACCTCCCAC, 5'-TACCACCACATGCTTCTGCACCAC, 5'-GCTTCAAGTCCAATGATTCGGAG, 5'-AGTTTCCCACCACCCCGAGGAGCAGG, 5'-ACACCTAGTGCTTCTGCTTATTG, 5'-CGGTGAGCTTGGGAAAGTGAC, 5'-GCTGAGAGCAGG, 5'-GGTTTCCCAGC, 5'-GGAGGAGCAGG (C4-9)) was digested with Not I and EcoRI. These fragments were gel purified, and 10 pmol of each were used in a labeling reaction with [α-32P]ATP using Moloney murine leukemia virus reverse transcriptase (Life Technologies) to end-fill the 5' overhang generated by EcoRI291 or NotI. Probes were labeled to a sp. act. of ~1 x 10^6 cpm/10 fmol. "A + G ladders" were prepared as size markers, and DNase I footprinting assays were performed as described (31) with crude nuclear extracts, prepared as described above.

Results

In a previous study, we mapped DNase I-hypersensitive sites around the C4 gene loci including regions proximal and distal to the C4 promoters (13). A single DNase I-hypersensitive site was
identified, for each of the C4 genes, located within 500 bp upstream of the cap sites, only in a cell line expressing C4 (Hep G2). No other hypersensitive sites were identified downstream or within the body of the gene. The investigation reported here has therefore concentrated on the region containing this hypersensitive site and extended the original findings by functionally characterizing the cis-acting sequences regulating the C4 promoter.

Promoter efficiency of the C4 5′-flanking region

A series of fusion genes consisting of progressively truncated C4 5′-flanking sequences linked to a β-globin reporter gene (see Materials and Methods) were constructed to identify the functional consensus binding motifs regulating transcription of the C4 gene in the hepatocyte cell line, Hep G2 (Fig. 1). The 3′ end of the C4 sequence in all the constructs was +62, and the 5′ end was variable. The longest constructs, specific for C4B and C4A, extended −1805 and −1668 upstream of the C4 cap site, respectively. Each β4 test construct was cotransfected with the control plasmid Hβ (see Materials and Methods) by calcium phosphate precipitation. Transcription of Hβ and β4 constructs resulted in cDNA extension products of 62 and 112 nucleotides, respectively. These data confirmed that the majority of the transcripts for both test and control constructs were being initiated from the previously documented endogenous cap sites for C4 (26) and β-globin (32). Transient transfections and primer extension analysis were performed in duplicate in four independent experiments.

Levels of transcription from reporter and control constructs were evaluated by scanning densitometry of the 112- and 62-bp bands. Densitometric values of the control plasmid (Hβ) were used as an internal control to normalize for variations in transfection efficiency. The effect of C4 5′ truncation on reporter gene expression, relative to the level of transcription determined from the full-length C4 fragment (−1558 to +62) in construct pβC4A, is summarized in Table I. No significant difference in transcriptional efficiencies of the C4A (pβC4A) and C4B (pβC4B) promoter regions was observed. Maximal promoter efficiency was associated with two constructs pβ4.4 (−178 to +62) and pβ4.3 (−126 to +62) with mean relative promoter efficiencies (± SE) of 1.9 ± 0.09 and 1.9 ± 0.12, respectively. These data are also indicative of the presence of a negative regulatory element within the region −178 to −1043, as the mean promoter efficiency of construct pβ4.8 was 0.9 ± 0.04 compared with 1.9 ± 0.09 for pβ4.4. Deletion of sequence from −126 to −81 (pβ4.2) reduced transcription of β-globin (0.8 ± 0.04). Total abrogation of gene transcription occurred with the construct pβ4.1 (−39 to +62), demonstrating that the sequence from −81 to −39 is essential for promoter function and directs 42% of maximal transcriptional activity.

DNA-protein interactions of the C4 promoter

To complement the promoter function analyses described above, we determined nuclear DNA-protein interactions of the C4 5′-flanking regions using EMSAs (see Materials and Methods).

![Diagram](https://example.com/diagram.jpg)

**FIGURE 1.** Design of pβ C4 reporter and Hβ control constructs. The C4 5′-flanking regions used in each construct are indicated. All fragments have a common 3′ end (position +62 in the C4 gene), which was cloned into an Smal site immediately upstream of a promoterless β-globin reporter gene. Hβ contains the entire human β-globin gene (HpaI-PstI fragment), driven by an SV40 enhancer. Primer extension analysis, using a β-globin-specific primer (P) was used to determine levels of β-globin expression. The length of the cDNAs generated from the test and control constructs (bold dashed line) were 112 and 62 bases, respectively. The cap site for the constructs is indicated by an arrow. The test and control constructs are not drawn to scale.

### Table I. Mean promoter efficiencies of the C4 reporter constructs

<table>
<thead>
<tr>
<th>pβC4 Reporter Construct</th>
<th>Mean Promoter Efficiency (± SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pβC4B (−1805 to +62)</td>
<td>0.9 ± 0.1</td>
</tr>
<tr>
<td>pβC4A (−1558 to +62)</td>
<td>1.0</td>
</tr>
<tr>
<td>pβ4.8 (−1043 to +62)</td>
<td>0.9 ± 0.04</td>
</tr>
<tr>
<td>pβ4.4 (−178 to +62)</td>
<td>1.9 ± 0.09</td>
</tr>
<tr>
<td>pβ4.3 (−126 to +62)</td>
<td>1.9 ± 0.12</td>
</tr>
<tr>
<td>pβ4.2 (−81 to +62)</td>
<td>0.8 ± 0.04</td>
</tr>
<tr>
<td>pβ4.1 (−39 to +62)</td>
<td>0.0</td>
</tr>
<tr>
<td>pβSp1mut (−81 to +62)</td>
<td>0.0</td>
</tr>
<tr>
<td>pβEmut (−81 to +62)</td>
<td>0.1 ± 0.04</td>
</tr>
<tr>
<td>pβEGmut (−81 to +62)</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* Densitometric analysis of the cDNA transcripts from the primer extension analysis of the RNA derived from transiently transfected Hep G2 cells. An arbitrary value of 1.00 was assigned to the level of transcription directed by the construct pβC4A. Correction for transfection efficiency and RNA loading was made by normalization of values against the control construct Hβ.
Three probes were generated, termed P4 (−178 to +62), P2 (−81 to +62), and P1 (−39 to +62) as described (see Materials and Methods). A SIGNAL SCAN search for transcription factor consensus motifs (33) was conducted on the C4 5′-flanking sequence associated with maximal expression (up to position −178). This identified putative binding domains for several factors (illustrated in Fig. 2A) including NF-1 (−110 to −97), Sp1 (−57 to −49), and three E box binding sites termed 5′ E box (−137 to −132), middle E box (−98 to −93), and 3′ E box (−78 to −73).

In agreement with the data from the transient transfection analysis of construct pβ4.1 (see Table I), probe P1 containing the same C4 promoter sequence (−39 to +62) did not form any DNA-protein complexes in EMSAs (data not shown). When probe P2 was used in EMSAs with nuclear protein extracts from Hep G2
cells, and in the absence of any competitor, three distinct complexes were observed (shown in Fig. 2B, lanes 2, 5, and 9). These binding complexes were termed P2A, P2B, and P2C. The addition of excess unlabeled probe P2 to the binding reactions effectively competed all DNA-protein interactions, whereas a nonspecific competitor in the form of EGmut (P2 with mutated Sp1 and E box binding sites) had no effect on complex formation (data not shown). Using a combination of specific competitors and the mutant P2 probes (i.e., Sp1mut and Emut), it was possible to demonstrate that the three electromobility shifts with probe P2 were due to specific interactions with factors binding to the Sp1 (−57 to −49) and 3’ E box (−78 to −73) consensus binding sites (Fig. 2B, lanes 3 to 11). The addition of the excess unlabeled Sp1 oligonucleotide abrogated the binding of P2 with factors corresponding to P2A and P2B, whereas P2C was altered but not abolished (Fig. 2B, lanes 4 and 8). These data were supported using Sp1mut as probe (Fig. 2B, lanes 3 and 10). This suggested that P2C represented more than one DNA-protein complex. The complex remaining after Sp1 competition was termed P2C2. When the Emut probe was used, there was no obvious alteration in complex formation (Fig. 2B, lane 6). However, when the E box and Sp1 competitors were used with Sp1mut and Emut probes, respectively, all DNA-protein complexes were effectively competed (Fig. 2B, lanes 7 and 11). It is possible that Sp1 binding masks the effect of E box competition or there is competition for access to binding sites. DNase I footprinting analysis (see below) was used to address this issue.

Probe P4 bound factors present in crude nuclear protein extracts from Hep G2 cells, in addition to Sp1 and the E box binding factor. These were specifically competed away using a 50-fold molar excess of unlabeled double-stranded consensus oligonucleotide specific for NF-1 (data not shown). These data suggest a positive role for the NF-1 site at positions −110 to −97, which is supported by the enhanced levels of transcription directed by the reporter construct (pB4.4) containing this site (see Table I).

DNase I footprinting analysis (see Materials and Methods) was used to confirm the data obtained from transient transfection analyses and EMSAs and to determine whether binding to the 3’ E box and Sp1 sites occurred simultaneously. End-labeled probes corresponding to pB4.4 (P4), pB4.2 (P2), pβSp1mut, pβEmut, and pβEGmut were used to define the position of the binding sites occupied by nuclear protein extracts from Hep G2 cells. P4 contains the consensus binding motifs for NF-1 (−110 to −97), Sp1 (−57 to −49), and three E box binding sites: the 5’ E box (−137 to −132), middle E box (−98 to −93), and 3’ E box (−78 to −73). Footprint analysis showed strong binding at the NF-1, middle E box, 3’ E box, and Sp1 binding sites (see Fig. 3). The 5’ E box binding motif was not associated with a footprint. The promoter region associated with pβ4.2 and containing the Sp1 and 3’ E box consensus binding motifs confirmed that both these sites were associated with strong footprints (shown in Fig. 3) and occupancy of the sites occurred simultaneously. In addition, a strong footprint was present at the extreme 3’ end of the probe in the region +35 and beyond. Since the transient transfection analyses demonstrated that this region was not functional, further characterization of this footprint was not pursued.

An Sp1 binding site at position −57 to −49 is essential for minimal C4 promoter activity

The data from the transient transfection analyses and the EMSAs led us to suggest that C4 promoter activity for the region −81 to +62 was achieved entirely through binding to the Sp1 and 3’ E box consensus motifs. To test this hypothesis, we prepared C4/β-globin reporter constructs containing mutated Sp1 and 3’ E box consensus binding domains (see Materials and Methods).

FIGURE 3. In vitro footprints of the region −178 to +62 (probe P4) of the C4 gene. The DNase I-treated probe is shown in the absence (−) (lanes 1 and 5) and in the presence of increasing amounts of an Hep G2 nuclear extract (lanes 2–4). A Maxim-Gilbert A + G ladder (lane 6) serves as a size marker. Protein binding sites are shown on the left and are denoted as open boxes.

These chimeric genes were tested by transient transfection into Hep G2 cells. Primer extension analysis of the RNA harvested from the transfected cells was used to determine the levels of β-globin RNA transcribed by the constructs. The experiment was performed in duplicate on four independent occasions. A representative example of such an experiment is shown in Figure 4, and the derived densitometric data from these experiments are shown in Table I. Comparison of the transcription efficiencies of the three mutant constructs pBSp1mut, pβEmut, and pβEGmut (all positions −81 to +62) with the wild-type promoter fragment (pB4.2) showed that the Sp1 binding site was critical for minimal promoter activity given that no reporter gene expression was detected from the constructs that contained a mutated Sp1 binding site (pBSp1mut and pβEGmut). Mutation of the E box alone (pβEmut), led to a considerably reduced transcription efficiency (0.1 ± 0.04) when compared with pB4.2, supporting a major role
for a member of the basic helix-loop-helix (bHLH) family of DNA-binding proteins in hepatic regulation of the human C4 gene. Results consistent with these observations were obtained when the probes containing the same mutated binding sites were used in EMSAs (shown in Fig. 2B). No footprints associated with Sp1 and/or E box binding were observed when these sites were mutated (data not shown). Furthermore, Sp1 binding was apparent in the absence of 3’ E box binding using the probe Emut, and the converse was true when the Sp1mut probe was used. Thus, the data clearly show that both Sp1 and 3’ E box transcription factors are able to activate their respective binding sites independently of each other, but both bind simultaneously to provide maximum transcriptional activity.

The −81 to +62 promoter region contains the elements necessary for tissue-specific expression of C4

We next attempted to address the issue of tissue-specific expression of C4. Two additional cell lines, HeLa and K562, were used for this analysis. Northern blot analysis of total RNA derived from these cell lines showed that Hep G2 cells express abundant amounts of C4 mRNA, but no mRNA was detected in either HeLa or K562 cells (data not shown). Reporter constructs and control plasmid Hb were transiently transfected into HeLa cells by electroporation (see Materials and Methods). These experiments were performed in duplicate on three occasions. In contrast to the data obtained when the same experiments were performed in Hep G2 cells, reporter gene expression from the test constructs was not detected in the HeLa cell line, while levels of expression of the control plasmid Hb in HeLa cells were similar to those observed in Hep G2 cells (data not shown). These data therefore suggest that the C4 reporter construct, pβ4.2, which in Hep G2 cells confers promoter activity, also contains the sequence(s) necessary to direct tissue-specific expression of C4. This led us to examine the possibility that the specificity of DNA-protein interactions with probe P2 may be different when EMSAs and DNase I footprinting assays were performed using crude nuclear protein extracts prepared from HeLa and K562 cells. Surprisingly, the complexes formed with probe P2 and the mutant probes using extracts from these nonexpressing cell lines appeared to be identical with those seen with Hep G2 cell extracts (Fig. 5). This was confirmed by the DNase I footprint analysis (data not shown). The binding sites associated with Sp1 and 3’ E box using Hep G2 nuclear extracts were very similar when nuclear extracts from K562 and HeLa cell lines were used. In summary, despite observing a tissue-specific pattern of expression of the C4 reporter constructs in transient transfection analysis, we were unable to demonstrate any differences in actual binding of nuclear proteins derived from those cell lines that did or did not express C4.

Discussion

We have used a combination of transient transfection studies and DNA-protein binding assays to define the promoter elements that regulate transcription of the human C4 gene. The hepatocyte cell
line. Hep G2, has been shown to express high levels of C4A and C4B mRNA (15) and is therefore the cell line of choice as a model in vitro system for studying C4 regulation. In this study, Hep G2 cells were transfected with reporter constructs containing between 39 bp and 1.8 kb of the C4 5′-flanking region. Variations in the level of expression of C4A and C4B genes have been reported in human synovial macrophages, in the early monocyto cell line U937, and in Hep G2 cells (21). Transient transfection into Hep G2 cells of the C4A (pβC4A) and C4B (pβC4B) reporter constructs detected no significant difference in promoter efficiency (Table I). This is consistent with data from our previous study in which we reported that DNase I-hypersensitive sites associated with the C4A and C4B genes are identical (13). It therefore appears unlikely that the variation in the levels of expression of C4A and C4B is due to transcriptional differences. We have recently shown that up-regulation of C4 expression by IFN-γ is a posttranscriptional event achieved by mRNA stabilization (24). Since the C4A and C4B mRNAs differ in the length of their 3′-untranslated region, it is possible that posttranscriptional differences in processing may account for tissue variations in C4A and C4B expression.

We found that maximal reporter gene expression was associated with the region up to position −126 of the C4 cap site (no change in activity up to −178). These data are supported by Tee et al. (34), who examined the regulation of the XA gene and identified powerful liver-specific sequences located within the first 84 bp upstream of the C4B cap site. This region downstream of −126 contains potential binding sites for ubiquitously expressed transcription factors of the bHLH family, NF-1 and Sp1. DNase I footprinting analysis (Fig. 3) showed a strong footprint in the region corresponding to the NF-1 binding site, and the data also suggested that the middle E box which overlaps the NF-1 binding site is occupied. The likelihood that the NF-1 consensus motif is functional in the human C4 promoter is also supported by our indirect evidence that a consensus NF-1 oligonucleotide was effective at competing away a DNA-protein complex formed in an EMSA (data not shown). Furthermore, increased reporter gene expression was associated with constructs pβ4.4/pβ4.3 (Table I) containing the NF-1 consensus binding motif. In the mouse C4 promoter, the NF-1 binding domain is conserved and has been shown to bind an NF-1-like factor, while double-stranded oligomers encompassing the NF-1 region specifically inhibited the mouse C4 promoter in vitro (12).

Our data also demonstrated that 42% of C4 promoter activity is still retained when the sequence from −126 to −81 (containing the NF-1 and middle E box binding domains) is deleted (Table I). However, no reporter gene expression or DNA-protein interactions were associated with the promoter fragment containing the sequence from −39 to +62. This strongly suggests that the Sp1 and 3′ E box binding sites are the only functional elements that effect C4 transcriptional activity within this compact region. These data contrast with the murine data (12) where no role for Sp1 has been identified. This could reflect the differing methodologies used or may be genuine differences between the human and murine promoters. The 3′ E box has the structure (−78) CACGGT (−73) which is a characteristic binding site for a large family of the class B subset of bHLH proteins. This motif is also conserved in the mouse and rat C4 promoters. In the mouse, this bHLH binding domain has been shown, along with other consensus binding motifs, to be essential for maximal promoter activity. A novel E box activator termed E-C4 has been isolated from rat liver that binds this domain in the murine promoter (35). This factor is distinct from, but appears to be closely related to, upstream stimulatory factors (USF) originally characterized from HeLa cell nuclear extracts (36).

These experiments led us to define the region containing the Sp1 binding site (up to position −57) as the minimal human C4 promoter. This region lacks orthodox TATA and CCAAT boxes. This raises important questions regarding C4 transcriptional start site selection and promoter activation because for the majority of tissue-specific eukaryotic genes, this follows nucleation of the TATA box with basal transcriptional machinery (reviewed in Ref. 37). Certain genes appear to lack a TATA box in the orthodox −30 position but either possess cryptic sequences (38), which may localize TATA box binding proteins, or have displaced TATA boxes. In the case of C4, a cryptic TATA box immediately upstream of the cap site seems unlikely, since the construct pβ4.1 was unable to direct any reporter gene expression. The evidence presented here identifies the Sp1 binding site at position −57 to −49 as the critical element for initiating gene transcription which is in common with a number of other recently studied TATA-less tissue-specific promoters (39–42). Recognition of such TATA-less promoters has led to the investigation and delineation of novel promoter mechanisms (reviewed in Ref. 37). Smale and Baltimore (43) have explored the hypothesis that Sp1 binding sites may function as surrogate TATA motifs where the initial localization of Sp1 leads to recruitment of the TATA box-binding proteins (TBP) through Sp1-TBP interactions (44). This suggestion has been supported by functional studies of tissue-specific TATA-less promoters (39–42). In almost all cases, an Sp1 consensus binding motif, located ∼50 bp upstream of the cap site, was critical to promoter function with respect to both level and accuracy of transcription. Indeed a survey of all known eukaryotic promoters determined that the peak location of Sp1 binding sites was at −50 (45) as for C4.

Our data also suggest that the C4 promoter fragment from position −81 to +62 contains the necessary elements to confer tissue-specific C4 expression. Transfection of HeLa cells, which do not express C4, with the C4 reporter constructs failed to demonstrate any detectable reporter gene expression. These data suggest that either the Sp1 or the 3′ E box, or indeed an association between the factors binding these elements, may confer tissue specificity. There is well-documented evidence illustrating the existence of a family of proteins able to bind the Sp1 consensus binding motif (46–48), and it has been suggested that a complex interaction of several factors at this site may be involved in tissue-specific gene expression. Similar interactions between ubiquitously expressed class A bHLH proteins and tissue-specific class B bHLH factors have also been reported (49–51) to confer tissue-specific gene expression. Although the promoter fragment (−81 to +62) confers tissue-specific expression, functional analysis using EMSAs and DNase I footprinting demonstrated that the Sp1 and 3′ E box binding sites bound specific Sp1 and E box binding proteins in nuclear extracts derived from both HeLa and K562 cell lines (Fig. 5). This is perhaps not surprising because factors capable of binding these consensus motifs are ubiquitously expressed. In supershift assays using three different Abs to USF-1 and USF-2, all were found to interact with E box binding proteins bound to probe P2, which were present in all three nuclear extracts (HeLa, K562, and Hep G2) examined (T. J. Mitchell, S. J. Rose, and B. J. Morley, unpublished observations). Findings similar to those reported here have been observed in the promoter of the MUC1 gene which codes for mucin and which is expressed mainly at the apical surface of glandular epithelial cells. Analysis of the 5′ sequences flanking the human MUC1 gene has demonstrated both Sp1 and E box binding factors to be involved in positive regulation of MUC1 in cell types that express mucin and in transcriptional repression of the gene in cell lines that do not express mucin (52, 53). Further work will be required to elucidate whether tissue specificity of human C4 transcription is due to restricted expression of specific
E box- and/or Sp1-regulatory proteins or whether other proteins are involved.

References