Promoter Elements of the Mouse Complement C4 Gene Critical for Transcription Activation and Start Site Location*

(Yuko Miyagoc, Marie-Dominique Galibert, Elena Georgatsou, Geneviève Fourrel, and Tommaso Meo)

From the Unité d’Immunogénétique et Institut National de la Santé et de la Recherche Médicale, U. 276, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France

We have explored the template and factor requirements for transcription of the gene encoding the murine complement component C4, expressed predominantly but not exclusively in liver and mononuclear phagocytes. Competition experiments in transcription assays with liver nuclear extracts show that the regions upstream of the transcription initiation site are largely dispensable for obtaining basal levels of accurately initiated transcription. Activated transcription, however, depends on three upstream regulatory factors, two of which interact with target sites seemingly related to NF-1 (region -112/-87) and USF (region -85/-64), respectively. A third upstream regulatory factor has been detected by the surprising finding that double-stranded oligomers covering sequences proximal to the cap site (position -48 to -7) stimulate transcription from the C4 promoter specifically. Results of nucleotide deletions and site-directed mutations argue that the C4 initiator, that is, the most critical element for basal and accurate transcription of the gene, overlaps the cap site and extends into the transcribed sequences (1 to +12). Immediately downstream of this region lies a last regulatory element (within the +5 to +43 boundaries) indispensable for high levels of transcription. These data assume wider interest because the C4 promoter does not contain TATA or CAAT boxes and does not feature any of the elements characteristic of the TATA-less genes so far reported.

*This work was supported in part by the Association pour la Recherche sur le Cancer and the Fondation pour la Recherche Médicale. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† Recipient of a Poste Vert fellowship awarded by the Institut National de la Santé et de la Recherche Médicale (INSERM) and a fellowship from the Fondation pour la Recherche Médicale. Present address: Department of Molecular Immunology, Medical Institute of Bioregulation, Kyushu University, Maidashi 3-1-1, Higashiku, Fukuoka 812, Japan.

‡ Present address: Institute of Molecular Biology and Biotechnology, Fondation de Research and Technology, P.O. Box 1527, Heraklion 71110 Crete, Greece.

∥ Present address: Unité de Recombination et Expression Génétique (INSERM U. 163), Institut Pasteur, Paris.

To whom correspondence should be addressed: Unité d’Immunogénétique, Institut Pasteur, 25, Rue du Docteur-Roux, 75724 Paris Cedex 15, France. Tel.: 33-1-45-88-85-95; Fax: 33-1-40-61-92-96.

The major site of synthesis of C4 is the liver (2). However, like many other complement components, C4 is synthesized in several extrahepatic sites (2, 3). C4 plasma concentration increases 2–3-fold during acute inflammatory or tissue injury (4), and its synthesis is stimulated by interferon-γ in many cell types (reviewed in Ref. 5).

Interest in the regulation of C4 gene expression is highlighted by the observation in man that complete deficiency of C4 has proved fatal in some patients and is otherwise associated with systemic or discoid lupus erythematosus and chronic immunocomplex diseases of the kidney (6). Furthermore, as in man, C4 exists in the mouse in two isoforms encoded by two nonallelic genes, tandemly arranged in the 5 region of the H-2 major histocompatibility complex (7), one of which C4-Slp (sex-limited protein) owes its discovery to its peculiar dependence on androgens for expression (see Refs. 8 and 9 for reviews). We have shown elsewhere that C4-Slp expression is regulated by nuclear factors and that rather than testosterone an intermittent rise in growth hormone is necessary and sufficient for the expression of the gene (10). Nevertheless, in spite of the extreme structural similarity between C4 and C4-Slp, the determinant of this isotype-specific hormonal regulation remains elusive (11–14).

The complete structure of C4 genes is known in man (15) (for upstream regions (16)) and mouse (17, 18), and in particular the 5′ upstream sequences of the mouse C4 and C4-Slp genes have been determined in several H-2 haplotypes (11, 19–22). Typically, C4 promoters lack a canonical TATA-box which is commonly found at promoters of genes transcribed by RNA polymerase II. It is generally acknowledged that the binding of the general transcription factor TFIID to the TATA box initiates an ordered assembly of other general factors at the promoter. Nevertheless, several promoters lack this essential element, and, at least in some of them, other sequences termed initiators (23) are thought to take up this critical role for basal promoter functions (i.e. positioning at the initiation site and mediation of preinitiation complex assembly) (23–29). An alternative assembly pathway, different from that proposed for TATA-box-containing promoters, has been suggested for initiator-containing promoters (28, 29). Thus, it is of interest to identify the core promoter elements of the C4 gene and elucidate their role both in the basal and activated mode of transcription. Understanding the properties of the C4 promoter will likely be rewarding for dissecting the mechanisms of the tissue-specific, developmentally and hormonally regulated C4-Slp gene expression in vivo.

We and others previously investigated the role of sequences in the 5′-flanking region of the C4 gene in trascriptonal regulation by means of gene transfer experiments (10–14, 22, 30).

1 E. Georgatsou and T. Meo, unpublished results.
Our results showed that the first 190 bp upstream of the C4 gene are necessary and sufficient for accurate and optimally efficient transcription in the human hepatoma cell line HepG2 and in L cell fibroblasts.

To characterize the functional sites of the C4 promoter and the factors associated with them, we exploited a cell-free transcription system with crude rat liver extracts, as originally developed by Gorski et al. (31) and used to analyze the promoters of many liver-specific genes (see e.g. Refs. 31-33). In an initial study, a 5' deletion analysis showed that the region -119 to -60 is functionally important for maximum and accurate in vitro transcription from the C4 promoter but is dispensable for basal expression (22). Here, we show that the C4 promoter has a simple architecture and represents a novel class of TATA-less promoters.

**EXPERIMENTAL PROCEDURES**

**Oligonucleotide Synthesis**—Single-stranded oligonucleotides were synthesized with a Beckman oligonucleotide synthesizer and purified on a 20% polyacrylamide, 7 M urea gels. Double-stranded oligonucleotide oligomers were obtained by mixing equal amounts of complementary oligonucleotides, heating to 85 °C for 5 min, and annealed by slow cooling.

**DNA Constructs**—The promoter-less 17BG vector was generated by inserting at the SaI site of pSBG (34) containing the following pairs of mutated oligomers for D-L M3 (Stratagene) cut with SalI for the 3' deletion mutants. To obtain C4/pBS(D-M3) and C4/pBS(E-M2), were generated from each HindIII-linearized template using T7 DNA polymerase and DNA ligase. For site-directed mutagenesis, the overlap extension PCR method was employed (36). The following primers were used for the first PCR reaction: for C4-17BG(M-I), 5'-GATATCGAATTCGTCGACCTCCCCAGCTC- TGTGCTTGGGGCCA-3' and (as 3' primer) 5'-GATATACAGACCTGGTC-GAGCGATCCAGGA-3' for D-l M3, pCAJ6 carrying the rabbit globin driven by its natural promoter and strengthened by the SV40 enhancer was a generous gift of P. Herombel and has been structurally and functionally described (35).

All 5' and 3' deletion mutants of the C4 promoter were obtained using the polymerase chain reaction (PCR). The sequences of the oligonucleotides used for PCR-generated constructs are as follows. For the C4-17BG(-198/+48), 5'-GATATCGAATTCGTCGACCTCCCCAGCTC-TGTGCTTGGGGCCA-3' and (as primer) 5'-GATATACAGACCTGGTC-GAGCGATCCAGGA-3' for D-l M3, pCAJ6 carrying the rabbit globin driven by its natural promoter and strengthened by the SV40 enhancer was a generous gift of P. Herombel and has been structurally and functionally described (35).

For site-directed mutagenesis, the overlap extension PCR method was employed (36). The following primers were used for the first PCR reaction: for C4-17BG(M-I), 5'-GATATCGAATTCGTCGACCTCCCCAGCTC-TGTGCTTGGGGCCA-3' and (as primer) 5'-GATATACAGACCTGGTC-GAGCGATCCAGGA-3' for D-l M3, pCAJ6 carrying the rabbit globin driven by its natural promoter and strengthened by the SV40 enhancer was a generous gift of P. Herombel and has been structurally and functionally described (35).

Our results showed that the first 190 bp upstream of the C4 gene are necessary and sufficient for accurate and optimally efficient transcription in the human hepatoma cell line HepG2 and in L cell fibroblasts.

To characterize the functional sites of the C4 promoter and the factors associated with them, we exploited a cell-free transcription system with crude rat liver extracts, as originally developed by Gorski et al. (31) and used to analyze the promoters of many liver-specific genes (see e.g. Refs. 31-33). In an initial study, a 5' deletion analysis showed that the region -119 to -60 is functionally important for maximum and accurate in vitro transcription from the C4 promoter but is dispensable for basal expression (22). Here, we show that the C4 promoter has a simple architecture and represents a novel class of TATA-less promoters.
Fig. 1. Footprinting data, nucleotide sequence of the 5' flanking region of the mouse C4 gene (H-2k), and positions of synthetic ds-oligonucleotides used in transcription and binding experiments. Panel A, deoxyribonuclease I footprinting of a 405-bp Taq1-BamHI fragment of the mouse C4 promoter, extending from -363 to +42, labeled at the BamHI site using Klenow enzyme (coding strand). C+T, Maxam and Gilbert sequencing reactions; lanes 2-5, digestion of the probe after incubation without extract (lane 2) or increasing amounts of nuclear extract (lane 3, 6 μg; lane 4, 13 μg; lane 5, 27 μg). Brackets and numbers indicate, respectively, the extent and positions of the footprinted regions, and arrowheads indicate the hypersensitive sites. Footprints consistent with, although weaker than, the ones described on the coding strand were obtained using a probe labeled on the non-coding strand.

Panel B, diagrammatic summary of the protected sequences and representation of the oligonucleotide used in the study. Boxes indicate the extent on the upper strand of protections against DNase I by factors present in mouse liver extracts. Full arrowheads indicate nucleotides next to DNase I hypersensitivity bonds. Transcription of the C4 gene starts at two adjacent sites (GTTGATCGAGAGGA), 56 and 57 bases upstream of the translation initiation codon, as seen in vivo (3, 19). In this report, the second base, A is denoted +1. Percent values denote relative transcription yields (22) of template constructs deleted upstream from the nucleotide marked by a half square bracket.

which marks off the minimal 5'-flanking region required for maximum transcription (22). Footprint II, from -112 to -64 distinctly circumscribed by runs of DNase I hypersensitive sites resembles a large uninterrupted domain. However, the presence of a DNase I hypersensitive bond at position -86 and the incremental contribution of subfragments of the region to the strength of the promoter, suggests that the footprinted area is a composite of two independent protein-binding domains. Footprint III, from -53 to -20 covers the site at position -30 characteristically occupied by the TATA motif in the vast majority of RNA pol II promoters (40). The site is required to bind factor TFIID and thereby to provide a critical nucleation site for the assembly of the typical transcription complex (for reviews, see Refs. 28 and 41). Nevertheless, the lack of an eventual
cryptic TATA motif in this region is reinforced by the observation that ablation of the region reduces transcription efficiency but does not affect start site positioning. To evaluate the contribution of these DNA-protein interactions to the expression of the C4 gene and to map out the regulatory elements of these regions, we carried out transcription competition experiments in the presence of double-stranded oligonucleotides (Fig. 2). Oligomer I covers footprint region I, the deletion of which had little effect on promoter activity. Overlapping oligonucleotides II-A and II-B were designed to encompass all putative regulatory sites of footprint region II, while oligonucleotide HNF-1 contains the binding site for mouse transcription factor HNF-1, which regulates a large number of genes expressed primarily in the liver (42), and was used as a negative control. Oligonucleotide III-C1 covers the 5' part of footprint region III and is similar to the consensus sequence, (G/T)(G/C)(A/T)(A/G)/G(T/T)/(G/C)(A/T)(A/G), of factor HNF-4 (43). Two templates were co-transcribed in each assay, the C4-β-globin construct, C4-17BG(-189/+48) (see Fig. 2, lower part) and the internal control template pCAJ6, carrying the rabbit β-globin driven by its homologous promoter strengthened by the SV40 enhancer (35). Unlabeled transcripts were detected by RNase protection analysis with a 32P-labeled RNA probe. Addition to the reactions of a 75-fold molar excess of oligomer II-A or oligomer II-B but not of the other oligonucleotides severely diminished transcriptional activity (Fig. 2, lanes 3 and 4). Moreover, the inhibition is specific for the C4 template, indicating that general transcription factors are not titrated out in the reaction and that footprint region II binds distinct factors able to activate specifically the C4 promoter. Consistent with the negative results of the deletion analysis (22), oligomer I, covering the most distal footprint does not seem to affect transcription. Oligo-
nucleotide III-C1 designed to test the functional relevance of the footprint region closest to the cap site did not affect reaction yields but induced part of the transcripts to initiate at an upstream site about position -25.

**Proteins That Bind Footprint Regions II-A and II-B**—Since oligomer II-A and oligomer II-B clearly inhibited transcription from the C4 construct, we tried to detect the putative transcription activating factors directly by electrophoretic mobility shift assays. Fig. 3A shows the protein-DNA complexes revealed with oligonucleotide 11-A. Though heterogeneous in mobility, these protein-DNA complexes are specific in that a 50-fold excess of unlabeled oligonucleotide II-A efficiently inhibits the labeled probe. This DNA fragment contains a sequence similar to the consensus binding site of the transcription factor CTF/NF-1 (44). As shown in Fig. 3A, a 50-fold excess of oligonucleotides containing the adenoviral CTF/NF-1 binding site (44) (see legend to Fig. 3) competes away all complexes completely. Extracts from different tissues revealed that such CTF/NF-1-like factors are present in much reduced quantities in kidney or spleen extracts. This difference might account for the preferential expression of C4 in hepatocytes.

When 32P-labeled oligomer II-B was used as a probe (Fig. 3B), only one major DNA-protein complex was formed. The complex is competed away by an excess of unlabeled oligomer, and the factor bound to site II-B seems to be widely expressed and relatively abundant. Although a minor faster migrating complex was seen with all extracts, its electrophoretic mobility seems to differ with the type of tissue extract used. A closer scrutiny of site II-B revealed that this sequence contains a short palindromic motif 5'-CACGTG-3' related to the E box consensus 5'-CANNTG-3' (45), typically recognized by members of the helix-loop-helix (HLH) protein family (46). Upon further characterization of this region, we also noted that oligonucleotide II-B partially competes the formation of complex II-A when added in a 50-fold excess to the binding reactions, whereas oligonucleotide II-A does not compete the binding of factor II-B (not shown). This observation indicates that oligomer II-B contains a relatively low affinity binding site for the CTF/NF-1-like factor detected with probe 11-A. Access to the II-B site is probably hindered by the E box factor, thus II-B complexes display a more uniform electrophoretic distribution than II-A complexes. But when used in large excess probe II-B can titrate out the CTF/NF-1 like factor and prevent it from binding to probe II-A or to the templates in transcription reactions.

**A Dinucleotide Mutation Abolishes the Binding of Factor II-B and Severely Reduces C4 Promoter Activity**—To validate the functional relevance of the E box core sequence found in frag-
abolishes the binding of factor II-B and severely reduces C4 promoter activity in vitro. A, electrophoretic mobility shift assay was carried out as described in Fig. 3. ³²P-End-labeled oligomer II-B (wild type, lanes 1-4) and a mutated oligomer II-B(mut) (lanes 5-7) were incubated with 1 or 2 µg of nuclear liver protein. The sequence of site II-B and the mutated nucleotides are shown in the lower part. Control reactions contained no protein (lanes 1 and 5). Competitors indicated on the top of each lane were added to the binding reaction in a 50-fold molar excess over the probe (lanes 3 and 4).

B, the same base pair substitutions were introduced into plasmid C4-17BG (−119/+48) (see “Experimental Procedures”) and tested in a transcription reaction. The transcriptional activities of the wild-type (lane 1) and of the mutated template (lane 2) were compared under the same conditions used in Fig. 2 and normalized to that of the internal control β-globin plasmid. Transcripts were quantitated by a PhosphoImager, and relative transcription levels are shown under the lanes. The dinucleotide substitution creates a double mismatch with the probe, which is recognized by the RNases and produces the band labeled NS (lane 2) corresponding to upstream transcripts (spurious start sites and read through). Markers (lane M) are pBR322 digested with HpaII.

Fig. 4. A dinucleotide mutation abolishes the binding of factor II-B and severely reduces C4 promoter activity in vitro. A, electrophoretic mobility shift assay was carried out as described in Fig. 3. ³²P-End-labeled oligomer II-B (wild type, lanes 1-4) and a mutated oligomer II-B(mut) (lanes 5-7) were incubated with 1 or 2 µg of nuclear liver protein. The sequence of site II-B and the mutated nucleotides are shown in the lower part. Control reactions contained no protein (lanes 1 and 5). Competitors indicated on the top of each lane were added to the binding reaction in a 50-fold molar excess over the probe (lanes 3 and 4).

B, the same base pair substitutions were introduced into plasmid C4-17BG (−119/+48) (see “Experimental Procedures”) and tested in a transcription reaction. The transcriptional activities of the wild-type (lane 1) and of the mutated template (lane 2) were compared under the same conditions used in Fig. 2 and normalized to that of the internal control β-globin plasmid. Transcripts were quantitated by a PhosphoImager, and relative transcription levels are shown under the lanes. The dinucleotide substitution creates a double mismatch with the probe, which is recognized by the RNases and produces the band labeled NS (lane 2) corresponding to upstream transcripts (spurious start sites and read through). Markers (lane M) are pBR322 digested with HpaII.
Fig. 5. 3' deletions affect C4 transcription in vitro. A, 3' deletion mutants were constructed as described under "Experimental Procedures." Each template (350 ng) was transcribed with liver nuclear extracts (60 pg) and analyzed with RNA probes. The 3' deletion point of mutant templates is shown on the top of each lane. The expected protected fragments corresponding to correctly initiated transcripts are indicated on the right. Strong bands (about 90 nucleotides long in lane 3) are the products of nonspecific initiation. Markers (lane M) are HpaII-digested pBR322.

B, upper part, summary of correctly initiated transcription rates of 3' deletion mutants. The transcription rates from C4/pBS (+36), C4/pBS (+24), and C4/pBS (+12) relative to that from C4/pBS (+49) are given on the right. RTL, relative transcription level obtained by quantitating the radioactivity in the correctly initiated transcripts with a Phospholmager. Equivalent results were obtained in three independent assays. Lower part, 32P-labeled antisense RNA probes were generated by transcribing each HindIII-linearized C4/pBS plasmid with T7 RNA polymerase.

Fig. 6A. Each deletion entailed a progressive decrease in transcription activity (see relative transcription levels in Fig. 5B). However, loss of fragment +12/+24 and especially of fragment +24/+49 caused also a shift of initiation to spurious start sites around position −20/−25 (lane 3). These results indicate that sequences within the region bracketed by nucleotides +13/+49 play a critical role in fixing precisely the "natural" transcriptional start site of the C4 promoter and affect the levels of transcription.

Double-stranded Oligonucleotides Homologous to the Cap Site Proximal Regions Alter C4 Transcription—To demonstrate more conclusively the direct role of the region downstream of the initiation site, a series of in vitro transcription competition experiments were performed (Fig. 6). Addition to the reaction mixture of oligomer D-L, which spans nucleotides +5 to +43 downstream of the cap site, resulted in a substantial decrease of correctly initiated transcripts together with the appearance of discrete downstream initiations centered at about nucleotide +20 (lanes 4 and 5).

For sake of completion, we also tested the effects on transcription of two overlapping DNA fragments III-C1 and III-C2 that cover the −30 site (Fig. 6A, upper part), where the TATA motif is usually found in TATA-containing genes (40). At variance with all oligomers used to inhibit C4 transcription, fragment III-C2 paradoxically but reproducibly increased the amount of transcripts by 60−80% without affecting the transcription of the control β-globin template. These results indicate that the region around position −30 rather than the interacting with the basal transcription factor TFIIID is able to outcompete a promoter specific transcriptional repressor. It will be interesting to find out if this repressor works by blocking the effect of one of the activators of the C4 promoter or by interfering directly with the initiation complex formation.

Arkhipova and Ilyin (47) noted that several TATA-less promoters of Drosophila possess a "downstream promoter sequence" critical for transcription initiation and able to bind a nuclear factor called downstream initiation factor. This sequence is characterized by a conserved 4-bp sequence, ACA/G(C) or CGCT, about 30 bp downstream from the CAP site. The C4 gene presents two such "ACAG motifs" in the downstream region (close to the stretches marked D-M1 and D-M2 in Fig. 6B), and we tested the role of the distal one using as competitor, oligonucleotides covering the natural sequence (+26 ACAG +29) with oligomer D-L wild type, or the mutated one (+26 AAGG +29) with D-M1. Like the wild-type DNA fragment D-L (Fig. 6A, lanes 4 and 5), the mutated oligomer, D-M1, outcompeted important transcription factors essential for correct initiation on the C4 template but not for the control β-globin promoter (Fig. 6C, lanes 2 and 6). Transcription competition experiments with a second series of mutated oligomers (see Fig. 6B) showed that certain mutations, i.e. D-M2 and D-M3, abolish the inhibitory activity of the oligonucleotides (Fig. 6C,
Fig. 6. Transcription competition assays with ds-oligonucleotides abutting the initiator nucleotide and transcriptions of mutated templates. A, competitions with oligonucleotide III-C1 (−49/−24) (lane 6), oligonucleotide III-C2 (−48/−7) (lanes 2 and 3), oligomer D-L (+5/+43) (lanes 4 and 5), and oligomer HNF-1 (lane 7, see the legend to Fig. 2) and salmon sperm DNA (lanes 1 and 8). The positions of these double-stranded oligonucleotides are shown on the top. Plasmid C4-17BG (−189/+48) (270 ng) and a control plasmid pCAJ6 (70 ng) were co-transcribed in vitro with HpaII-digested pBR322.

B, nucleotide sequence near the cap site. The position of ds-oligomer D-L (39 bp) is also shown. The sites of mutations are indicated. Markers (lane M) are (D-M1 to D-M4) introduced into ds-oligomer D-L or into plasmid C4/pBS (+49) are underlined. For mutation D-M1, the underlined dinucleotides CA were changed to AG. For the other three mutations, underlined nucleotides were changed to T. C, competitions with ds-oligomer D-L (wild-type) (lane 6) and its mutated derivatives (lanes 2-5). 250 ng of mutated oligomers were added to the transcription reactions as described in Panel A. Competitions with the single-stranded oligonucleotides (ss-D-L, the coding strand, and the non-coding strand of oligomer D-L) are also shown (lanes 8 and 9).

D, mutations D-M3 and D-M4 were introduced into C4/pBS (+49) and tested for template activity. Transcription of C4/pBS (+49) (wild type) (lane 1) and of the two mutated plasmids, C4/pBS(D-M3) (lane 2) and C4/pBS(D-M4) (lane 3) were carried out and analyzed with RNA probes generated from each linearized template as described in Fig. 5. RTL, relative transcription level.

These results establish that transcription inhibition results from a sequence-specific interaction of a novel factor with the competing DNA. The position of mutations D-M2 and D-M3 indicate that the regulatory region can be roughly mapped from +11 to +22, but we have no experimental data concerning its relationship to the downstream initiation factor binding site of Drosophila promoters.

To prove that the functional region identified with DNA fragment D-L and its substitution mutants has a genuine regulatory role within the context of the C4 promoter, mutations D-M2 and D-M3 indicate that the whole region can be roughly mapped from +11 to +22, but we have no experimental data concerning its relationship to the downstream initiation factor binding site of Drosophila promoters.

Identification of an Initiator-like Element Adjacent to the Cap Site—Several promoters, generally but not necessarily, lacking a TATA-box, have an initiator element which can replace or reinforce the role of the TATA sequence in directing the location of a transcription start site (23–27, 29). To analyze protein-DNA interactions around the cap site, we prepared a set of wild-type and mutated oligomers (shown in Fig. 7A) and used them as competitors in the transcription reactions. A 50-fold
Fig. 7. An internal "initiator-like element" borders on the cap site. A, sequences of the ds-oligomer E and mutants thereof (E-M1 to E-M5). Mutations E-M1 and E-M2 were also introduced into C4/pBS(+49) and tested for template activity. Only changed nucleotides are indicated. B, transcription competitions with a series of mutated double-stranded oligonucleotides. A 50-fold excess of unlabeled wild-type oligomer E (lane 4).
excess of oligomer E (wild-type) over the C4 template reduced the level of C4 transcription but also affected transcription of the internal control (Fig. 7B, lane 4). We tested a set of mutated oligonucleotides in the attempt to delineate the regulatory sequence in this region. Fig. 7B shows that while mutations upstream of the cap site (see E-M1) did not affect the inhibitory activity of oligomer E all other 3' proximal mutations impaired it in varying degrees. As all oligomers able to outcompete transcription from C4 also inhibited transcription on the internal control template, we set out to test the direct influence of the mutations when present within the context of the C4 template. For this, we introduced mutations E-M1 and E-M2 into the plasmid C4/pBS(E-M2). The first mutation strongly affected the transcription yield without interfering with start site selection (Fig. 7C, lane 2). Alternatively, when the mutated plasmid C4/pBS(E-M2) was used as a template, transcription initiation from several new start sites took over completely that from the natural initiation site (Fig. 7C, lane 3). Therefore, we conclude that the integrity of the sequence 3' of the +1 site is essential to ensure accurate initiation.

Certain initiator elements are thought to be recognized by specific initiator-binding proteins (25, 48, 49). To examine their possible relationship with the C4 promoter we tested whether double-stranded oligonucleotides specific for two well characterized initiator elements namely the Inr-TdT of the mouse terminal deoxynucleotidyl transferase gene which binds factor TFII-I (49) and the P5 +1 element of the P5 promoter of the adenovirus-associated type 2 which binds factor YY1 (48) would inhibit transcription from the C4 template (Fig. 7D). Neither an excess of oligomer Inr-TdT (lane 2) nor of oligomer Inr-P5 (lane 3) inhibits transcription of the C4 template, indicating that the C4 initiator-element and these two initiators are recognized by different factors.

**DISCUSSION**

Transcription by RNA polymerase II can be regulated at multiple levels, including higher order chromatin structure not reconstituted in conventional nuclear extracts. However, the combination of cell transfections and cell-free transcription assays using liver nuclear extracts is a valuable approach to sort out functionally relevant DNA-protein interactions at promoters sites.

Our results show that the basal expression of the C4 gene is potentiated by two promoter-specific transcriptional activators interacting with regions -112 to -87 and -85 to -64 relative to the transcriptional initiation site (Fig. 8). To the first, more distal, region binds a factor which has CTF/NF-1-like binding specificity and produces electrophoretically heterogeneous DNA-protein complexes. Interestingly, this binding activity is enriched in liver extracts, suggesting that the CTF/NF-1-like factor may play a role in the preferential C4 expression in hepatocytes. Deletion of the binding site for this factor from a C4 transcriptional template results only in a two fold reduction of the promoter activity in vitro. Furthermore, nuclease differences in the CTF/NF-1 site exist between the C4 promoter and the promoter of C4-Slp, which may account for the weaker footprint observed in domain IIA of the latter promoter (data not shown) and for the lower transcriptional activity of the C4-Slp promoter (22).

The factor that activates C4 transcription through the region from -85 to -64, is more widely expressed and abundant. This site bears the E box near the cap site. Although this motif is sufficiently well characterized (451, also a hepatitis C virus reporter consensus can be defined (53), the hypothesis has been put forward that specific initiation requires that the preinitiation complex be anchored at two points, one in close proximity and the other upstream or downstream from the cap site (54).

We have identified two kinds of promoter elements downstream of the cap site of the C4 gene using site-directed mutations and deletion analysis in combination with oligonucleotide competition assays. The first element (-1/+42) designated C4-Inr in Fig 8 seems to function as a typical initiator element in that (a) it does not require upstream sequences for accurate initiation and (b) mutations introduced in this region caused multiple initiation sites in vitro (Fig. 7C). Therefore, we conclude that this element plays an important role in the precise determination of the transcriptional initiation site. Oligomer competition experiments (Fig. 7B) and gel mobility shift assays of DNA-protein complexes (not shown) suggest the existence of
sequence-specific initiator binding factor(s) for this element, which cannot be assimilated to any of the initiators so far known (55, 56).

The sequence of the non-coding strand of the C4-Inr (5'-CTGCTCTTCATCCTG-3') is similar in pyrimidine-content and location to that described for the TdT gene (5'-CTTACCTTCTAGG-3') (23) or the adenovirus-associated type 2 P5 promoter (5'-GGTGTCTGAATTTGACCG-3') (48). However, double-stranded oligonucleotides containing these initiator sequences in vitro did not inhibit transcription from the C4 gene constructs (see Fig. 7D), indicating that different factors interact with these three "initiators." We also noticed that transcription from the C4-Inr is downshifted if carried out in the presence of a pyrimidine-rich single-stranded DNA from the non-coding strand (−1 to +12) but not of its complementary purine rich strand (data not shown). We have not ascertain whether this effect seen also with ss-oligomer D-L (Fig. 6), is due to occlusion of factor binding by the formation of triple-helical complexes.

The element immediately 3' of the C4 initiator element appears to influence both efficiency of transcription and start site selection. For instance, bacterially expressed PYBP (58) failed to bind to oligomers E (data not shown). Although other experiments are required to decide whether these effects are mediated by triple-helical DNA complex, it should not be neglected that factors acting in proximity of the +1 site of certain promoters might have high binding affinity to single-stranded DNA, since this region is expected to undergo a critical conformational change prior to transcriptional initiation.

The element immediately 3' of the C4 initiator element appears to influence both efficiency of transcription and start site selection. The topology of this element can be better defined by the preparation of triple-helical complexes (57) or to genuine competition. But we note that start site sliding induced by the pyrimidine oligonucleotide has two elements of specificity. First of all it is template specific in that it does not interfere with RNA initiation on the β-globin internal control. In addition, the phenomenon does not seem to depend on proteins generally capable to bind pyrimidine-rich oligomers. For instance, bacterially expressed PYBP (58) failed to bind to oligomers E (data not shown). Although other experiments are required to decide whether these effects are mediated by triple-helical DNA complex, it should not be neglected that factors acting in proximity of the +1 site of certain promoters might have high binding affinity to single-stranded DNA, since this region is expected to undergo a critical conformational change prior to transcriptional initiation.

The element immediately 3' of the C4 initiator element appears to influence both efficiency of transcription and start site selection. The topology of this element can be better defined by the preparation of triple-helical complexes (57) or to genuine competition. But we note that start site sliding induced by the pyrimidine oligonucleotide has two elements of specificity. First of all it is template specific in that it does not interfere with RNA initiation on the β-globin internal control. In addition, the phenomenon does not seem to depend on proteins generally capable to bind pyrimidine-rich oligomers. For instance, bacterially expressed PYBP (58) failed to bind to oligomers E (data not shown). Although other experiments are required to decide whether these effects are mediated by triple-helical DNA complex, it should not be neglected that factors acting in proximity of the +1 site of certain promoters might have high binding affinity to single-stranded DNA, since this region is expected to undergo a critical conformational change prior to transcriptional initiation.

The element immediately 3' of the C4 initiator element appears to influence both efficiency of transcription and start site selection. The topology of this element can be better defined by the preparation of triple-helical complexes (57) or to genuine competition. But we note that start site sliding induced by the pyrimidine oligonucleotide has two elements of specificity. First of all it is template specific in that it does not interfere with RNA initiation on the β-globin internal control. In addition, the phenomenon does not seem to depend on proteins generally capable to bind pyrimidine-rich oligomers. For instance, bacterially expressed PYBP (58) failed to bind to oligomers E (data not shown). Although other experiments are required to decide whether these effects are mediated by triple-helical DNA complex, it should not be neglected that factors acting in proximity of the +1 site of certain promoters might have high binding affinity to single-stranded DNA, since this region is expected to undergo a critical conformational change prior to transcriptional initiation.

The element immediately 3' of the C4 initiator element appears to influence both efficiency of transcription and start site selection. The topology of this element can be better defined by the preparation of triple-helical complexes (57) or to genuine competition. But we note that start site sliding induced by the pyrimidine oligonucleotide has two elements of specificity. First of all it is template specific in that it does not interfere with RNA initiation on the β-globin internal control. In addition, the phenomenon does not seem to depend on proteins generally capable to bind pyrimidine-rich oligomers. For instance, bacterially expressed PYBP (58) failed to bind to oligomers E (data not shown). Although other experiments are required to decide whether these effects are mediated by triple-helical DNA complex, it should not be neglected that factors acting in proximity of the +1 site of certain promoters might have high binding affinity to single-stranded DNA, since this region is expected to undergo a critical conformational change prior to transcriptional initiation.
Transcriptional Analysis of the Mouse C4 Promoter