Start Site Selection at the TATA-less Carbamoyl-phosphate Synthase (Glutamine-hydrolyzing)/Aspartate Carbamoyltransferase/Dihydroorotase Promoter*

(Received for publication, June 14, 1993, and in revised form, October 4, 1993)

Richard Kollmar†‡†, Kristine A. Sukow§, Stephen K. Sponagle¶, and Peggy J. Farnham**

From the †Program in Cell and Molecular Biology and the §McArtrade Laboratory for Cancer Research, University of Wisconsin, Madison, Wisconsin 53706

Transcription of the carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase (CAD) gene from the Syrian hamster, Mesocricetus auratus, starts at a single major site (2, 3). To study the regulation of CAD transcription, we examined how RNA polymerase II is positioned at this start site.

We recently have proposed that three types of positioning elements direct site-specific initiation at promoters that are transcribed by RNA polymerase II (4). (i) About 80% of the promoters contain a TATA box around –30. The TATA box directs initiation about 30 bp downstream by binding the basal transcription factor TFIID, which assembles the other basal factors and RNA polymerase II into a preinitiation complex. (ii) At promoters without a TATA box, an activator protein such as Sp1 may direct site-specific initiation near its binding site by tethering TFIID to the template and by stabilizing the preinitiation complex through protein-protein interactions. (iii) About 60% of the promoters transcribed by RNA polymerase II contain a consensus sequence, the initiator, at the start site. The initiator appears to interact with a common component of the preinitiation complex, possibly RNA polymerase II itself, and to determine the exact start site if it occurs at an appropriate distance from a TATA box or an activator binding site.

The CAD promoter contains two of these three types of positioning elements. Here we show by sequence alignment that there is a consensus initiator at the start site (+1), but no TATA box around –30; we have previously shown that there are Sp1 binding sites at –49 and –70 (2) and that Sp1 stabilizes preinitiation complexes at the CAD promoter (5). To examine the positioning function of these elements, we determined the transcriptional activities and start sites for various mutant CAD promoters in vitro and in vivo. Sequences from –81 to +26 were necessary and sufficient for accurate and efficient initiation of transcription. The proximal Sp1 site at –49 substituted for the missing TATA box as the primary positioning element and directed initiation about 50 bp downstream. A binding site around –17 for a ubiquitous activator protein did not participate in start site selection. The consensus initiator at +1 was not required nor were other consensus initiators near +1 sufficient for transcriptional activity.

MATERIALS AND METHODS

Plasmid DNAs—Promoter fragments were inserted with standard cloning techniques (6) into the vector pBS* (Stratagene) or, as indicated by the suffix GL, upstream of the luciferase reporter gene into the vector pGL2Basic (Promega). Plasmids with the suffix GLSV contained the origin of replication between positions 5171 and 132 of simian virus 40, which was isolated from plasmid pGEMHindIIIIC and inserted at the BamHI or SalI site downstream of the luciferase gene. Details of the cloning schemes have been described (7). The cloning junctions and mutated regions of all plasmids were sequenced. Plasmid DNA was purified by equilibrium centrifugation in CsCl-ethidium bromide gradients or by chromatography over Qiagen columns (Qiagen). DNA concentrations were determined by fluorometry (5).

* This work was supported by United States Public Health Service Grants CA07175 and CA25076 (to P. J. F.), a Hilldale Undergraduate Research Fellowship (to K. A. S.), and a Tewartha Undergraduate Research Grant (to S. K. S.). 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.

† Present address: Howard Hughes Medical Institute, Center for Basic Neuroscience Research, University of Texas Southwestern Medical Center, Dallas, TX 75235-9117.

‡ To whom correspondence should be sent: McArland Laboratory for Cancer Research, University of Wisconsin, 1400 University Ave., Madison, WI 53706. Tel.: 608-262-2071; Fax: 608-262-2924.

§ The abbreviations used are: CAD, carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase; bp, base pairs; kb, kilobase pair(s); TFIIID, transcription factor IID.
Four plasmids containing CAD promoter fragments in pBS* were derived from plasmid pC76B (2), which contains a BglII fragment with CAD sequences from -1051 to +406 in the BamHI site of pBS*, as follows. Plasmid p-C76B/1/483, a 1.2-kb EcoRI-BglII fragment of pC76B was inserted into EcoRI-SmaI cut pBS*. p-C76B/2/26, a 1.1-kb BstUI-EcoRI fragment of pC76B was inserted into EcoRI-SmaI cut pBS*. p-C76B/3 and p-C76B/4, a 1.5-kb NcoI-SmaI fragment of pC76B/1/26 was filled in and religated. The CAD sequences from these four plasmids were then inserted into pGt2Basic to create the following plasmids: p-C76B/1/483L and p-C76B/2/26L, a 1.1-kb KpnI-XbaI fragment of pC76B/1/483 or pC76B/2/26 was inserted into KpnI-Nhel cut pGt2Basic. p-C76B/3L and p-C76B/4L, a 0.2- or 0.1-kb KpnI-BamHI fragment of pC76B/3 or pC76B/4 was inserted into KpnI-BglII cut pGt2Basic.

Plasmids that contain mutants of the minimal CAD promoter between -81 and +26 are indicated by the prefix pCM and were derived from p-C76B/1/26 or p-C76B/2/26L. Linker-scanner mutants (see Table 1) were created by oligonucleotide-directed mutagenesis with uracil-containing single-stranded DNA; the oligonucleotides contained between 11 and 17 additional nucleotides on both sides of the mutated sequence. The internal deletion mutant pCM-38 [CM-38] was created by inserting a 1.6-kb SacI-BglII fragment of pCM-131-1/bgl into pG2-1/26L to create pCM-38L. The insertion mutant pCM-45L was created by inserting a 1.8-kb SacI-BglII fragment of pCM-38L into pG2-1/26L to create pCM-45L. The deletion mutant pCM-50L was created by inserting oligonucleotide-directed mutations with uracil-containing single-stranded DNA and the oligonucleotide, d(AGCGGCAGCTTAATGCTGGAGCACGACTG), where X denotes a random equimolar mixture of all four nucleotides.

Plasmids that contain synthetic promoters were derived from plasmids IV and VII (8) as follows: PTIGL and pSTIGL, a 0.1- or 0.2-kb HpaI-BamHI fragment of plasmid IV or VII was inserted into Small-BglII cut pG2-1/26L. p(Honk)TIGL and p(Shuffle)TIGL, two Honk or Shuffle oligonucleotides (Fig. 2) were inserted at the BglII site into pTIGL, p(Honk)GL and p(Shuffle)GL, an EcoRV/HindIII fragment containing the TATA box and the initiator was deleted from p(Honk)TIGL or p(Shuffle)TIGL.

Transfections—BH2-21 baby hamster kidney cells (American Type Culture Collection) were maintained in Dulbecco-Vogt's modified Eagle's medium (Life Technologies Inc.) supplemented with 10% (v/v) defined-supplemented bovine calf serum (HyClone), 100 units of penicillin/ml, and 100 μg of streptomycin/ml (Life Technologies Inc.) at 37°C and 5% (v/v) CO2 in a humidified incubator. The cells were plated at about 105 cells/cm² and passaged after 3 days with trypsin-EDTA (Life Technologies Inc.); their doubling time was between 11 and 13 h. One day before transfection, 105 cells in 1 ml of medium were plated into 22-mm wells in 12-well tissue culture plates. The cells were transfected with 1 μg of plasmid DNA per well by calcium-phosphate coprecipitation as described (9), except that the glycerol treatment was omitted from the buffers and that dialysis was against the same buffer as for the BS-4 nuclear extract. Protein binding to the radiolabeled Honk oligonucleotide was measured in an electrophoretic mobility shift assay (15).

**RESULTS**

The Minimal CAD Promoter—As a first step towards characterizing the positioning elements in the hamster CAD promoter, we determined the shortest fragment that supported accurate and efficient initiation in vitro and in vivo. Sequences upstream of -75 or downstream of +26 could be deleted without changing the transcriptional activity or the start site pattern in nuclear extract from hamster BS-4 cells. We then transfected four promoter fragments that had upstream ends at -1051 or -81 and downstream ends at +83 or +26 that were fused to a luciferase reporter gene, into hamster BHK-21 cells (Table 1). The activity of the shortest fragment was about one-fourth of that of the longest fragment (p < 0.0001), but was still 200-fold higher than the vector background. We concluded from these results that a minimal CAD promoter from -81 to +26 was necessary and sufficient for accurate and efficient transcription and that additional cis-acting elements were located both upstream and downstream of the minimal CAD promoter.

To locate the cis-acting elements within the minimal CAD promoter, we assayed the activities of a series of linker-scanner mutants in vivo and in vitro (Fig. 1). Mutations in 6 out of 10 regions changed the activity. One of the mutations between -30 and -22 increased the activity in vitro about 4-fold; however, this was probably because the linker and the adjacent wild-type sequence created a TATA box (-32 CTatagat -25) at the appropriate distance upstream of the start site. Mutating either one of the Sp1 binding sites at -70 and -49 reduced transcription in vivo and in vitro; mutating both sites essentially abolished transcription. Mutations between -21 and -13 reduced transcription only in vivo, whereas mutations from -2 to +4 reduced transcription only in vitro. We concluded from this linker-scanner analysis that at least four cis-acting elements might control transcription from the minimal CAD promoter: two Sp1 binding sites at -70 and -49, the region between -21 and -13, and the region between -2 and +15. Because the activity of several promoter mutants was different in vivo and in vitro and because initiation in vitro was heterogeneous (Fig. 9), R. Kollmar, unpublished observations.

**TABLE 1**

<table>
<thead>
<tr>
<th>Promoter Fragment</th>
<th>Relative Activity</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>5' End</td>
<td>3' End</td>
<td>Average ± S.E. (n)</td>
</tr>
<tr>
<td>-1051</td>
<td>+83</td>
<td>100 ± 17 (12)</td>
</tr>
<tr>
<td>+83</td>
<td>-1051</td>
<td>52 ± 9 (12)</td>
</tr>
<tr>
<td>+26</td>
<td>-1051</td>
<td>42 ± 7 (12)</td>
</tr>
<tr>
<td>+83</td>
<td>+26</td>
<td>28 ± 5 (12)</td>
</tr>
<tr>
<td>Vector</td>
<td>No DNA</td>
<td>0.13 ± 0.02 (12)</td>
</tr>
</tbody>
</table>

* Four low values were excluded from the analysis of variance because of their large residuals; the relative activity of p-C76B/26L may therefore be overestimated.
Start Site Selection at the CAD Promoter

A

\[
\begin{array}{cccccccccccc}
-80 & -70 & -60 & -50 & -40 & -30 & -20 & -10 & +1 & +10 & +20 \\
\end{array}
\]

1B), unlike initiation at the cellular CAD promoter (2), we continued the analysis of the CAD promoter entirely in vivo.

The Honk Site—We noticed that the sequence between -21 and -13, which activated transcription in vivo, was in the middle of the region between -30 and -4 that binds nuclear protein from HeLa cells (2) and was part of an imperfect 18-bp palindrome between -30 and -13 (Fig. 2). Because several transcription factors bind to palindromic sites (16), we examined whether this palindrome, which we called a Honk site, was sufficient to bind protein and to activate transcription from a heterologous promoter. In an electrophoretic mobility shift assay (EMSA), we concluded from the shift in the autoradiograph of primer extension products with the linker-scanner mutants. The major start site in lane 5 was at about -20. The extension products between -1 and +3 were quantitated together; initiation was always heterogeneous in vivo, even at a CAD promoter fragment from -1051 to +83.3

B

Fig. 1. cis-Acting elements of the minimal CAD promoter. A, linker-scanner mutations were introduced throughout the minimal CAD promoter. Top, the wild-type sequence of the non-template strand; the bases substituted in each mutant are aligned underneath. Underline, bases that are protected by nuclear proteins against DNase I digestion (2). Relative activity in vivo, BHK-21 cells were transfected with the promoter constructs in vector pGL2Basic (plasmids pCM-761[-65(GL), etc.], and luciferase activities were determined. Relative activity in vitro, the promoter constructs in vector pBGS (plasmids pCM-761[-65(GL), etc.) were transcribed in B5-4 nuclear extract, and the transcripts were analyzed by primer extension. S, significantly different from wild type (p < 0.004); n.d., not determined. B, autoradiograph of primer extension products from one of three in vitro transcription experiments with the linker-scanner mutants. Left, the size of each marker in nucleotides. Arrowheads, the expected 131-nucleotide extension product. The major start site in lane 5 was at about -20. The extension products between -1 and +3 were quantitated together; initiation was always heterogeneous in vivo, even at a CAD promoter fragment from -1051 to +83.
Protein binding to the Honk site. The Honk oligonucleotide with its complement contained the palindrome in the CAD promoter from -30 to -13 between cohesive BglII ends. Binding of protein from the indicated sources to the radiolabeled Honk probe was measured in an electrophoretic mobility shift assay. Concatemerized competitors were the Honk oligonucleotide itself; the Shuffle oligonucleotide, which had the order of the nucleotides changed, but the palindrome structure with two mismatches maintained; and the Middle3 oligonucleotide, which had only the middle of each half of the palindrome altered. Lower case, nucleotides that differ from the CAD sequence. At the top of the autoradiograph are the wells of the gel, at the bottom is free probe; arrowheads, the specific complex of protein and probe.

**Table II**

**Activation of a synthetic promoter through Honk sites**

Two Honk or Shuffle oligonucleotides (Fig. 2) were inserted in the same orientation into vector pGL2Basic or in plasmid pTIGL, about 25 bp upstream of a synthetic promoter that comprised the TATA box from the major late promoter of human adenovirus 2 and the initiator from the promoter of the mouse terminal T-cell leukemia virus. BHK-21 cells were transfected with these plasmids, and luciferase activities were determined.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relative activity (average ± S.E.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No DNA</td>
<td>5.6 ± 0.5</td>
</tr>
<tr>
<td>pGL2Basic</td>
<td>56 ± 7</td>
</tr>
<tr>
<td>p(Shuffle)GL</td>
<td>58 ± 7</td>
</tr>
<tr>
<td>p(Honk)GL</td>
<td>98 ± 12a</td>
</tr>
<tr>
<td>pTIGL</td>
<td>100 ± 12a</td>
</tr>
<tr>
<td>p(Shuffle)TIGL</td>
<td>127 ± 15</td>
</tr>
<tr>
<td>p(Honk)TIGL</td>
<td>390 ± 50a</td>
</tr>
<tr>
<td>pSTIGL</td>
<td>(1.77 ± 0.23) × 10^{10}</td>
</tr>
</tbody>
</table>

* Significantly more active than pGL2Basic (p ≤ 0.02).
* Significantly more active than pTIGL and p(Honk)GL (p < 0.0001).

...heterogeneous than at the cellular CAD promoter (see above), the start sites in the following experiments were mapped in vivo. Initially, we could not detect transcripts from the minimal CAD promoter in transfected BHK-21 cells. We therefore inserted the origin of replication from simian virus 40 downstream of the luciferase reporter gene; the resulting constructs were transfected into COS cells, which express the viral T antigen and amplify plasmids carrying this origin of replication to a high copy number. Using this in vivo assay, we observed initiation at the wild-type CAD promoter not only at +1, but also at about -19 (Fig. 4, lane 4). Replacing the sequences between -12 and +14 with a shorter linker shifted the major start site in vivo downstream (lane 5, -13 to +15). This result suggested that the start site was determined by sequences between -81 and -12, which included two Sp1 binding sites around -70 and -49 and the Honk site between -30 and -13. To determine whether Sp1 or Honk directed initiation at +1, we inserted a linker at -44, between the proximal Sp1 site around -49 and the Honk site. This insertion shifted the major start site upstream (lane 6, -45 to +15), suggesting that one or both of the Sp1 sites upstream of -44 determined the start site. To test whether the proximal Sp1 site at -49 directed initiation at +1, we replaced it with two different linker sequences (-55 to -44 and -55 to -44), suggesting that one or both of the Sp1 sites upstream of -44 determined the start site. In mutant 678, which had a good match to the consensus at +1, the start site shifted accordingly to a new position 1 bp upstream (lane 6). In mutants 692 and 703 with no good matches to the initiator consensus, transcription was reduced (lanes 8 and 9). However, in mutant 692 with a good match to the consensus at +4, transcription also was reduced; there was less initiation at +4 than at a poor match to the consensus at +7 and at another good, but more distant match at -5 (lane 7). These results indicated that a consensus initiator at or near +1 in the minimal CAD promoter was not sufficient for accurate and efficient initiation. However, three of the four substitutions of the -2 to +6 region were deleterious, suggesting that this region contributed to the activity of the minimal CAD promoter.
CAD gene at a single major site (2, 3). Our analysis demonstrated that start site selection at the hamster CAD promoter was not controlled by the most common positioning element, a TATA box around –30. First, the ability of a sequence to recognize the DNA-binding subunit of TFIIID and to direct initiation 30 bp downstream correlates with its similarity to the TATA box consensus (4); the profile of sequence similarity to the TATA box consensus showed no significant matches in the minimal CAD promoter, in particular not around –30. Second, mutating the region between –30 and –22 did not affect transcription in vivo. The fortuitous creation of a better TATA box in plasmid pCM–31 (78% similarity to the consensus) increased promoter activity, but only in vitro. In vivo, the activity of this mutant may have been identical to that of the wild-type promoter because the DNA was less accessible and other steps than binding of TFIIID to the CAD promoter were limiting in the cell. Initiation at the CAD promoter, as at other TATA-less promoters (4), probably still requires TFIIID. We have not pursued this question because heat treatment of nuclear extract inactivates Sp1 before TFIIID (5) and because a reconstituted system for transcription of the CAD promoter has not yet been developed.

The Honk palindrome between –30 and –13 bound another activator protein, but had no apparent role in start site selection. Our results suggested that only the downstream half of the Honk palindrome was required for activation and protein binding. First, the 28-bp footprint over the Honk site is centered at –17 rather than at the middle of the palindrome at –21 (2). Second, mutating the region from –21 to –13 reduced the activity of the minimal CAD promoter in vivo to one-half, whereas mutating the region from –30 to –22 had no effect. In a manner similar to the 2-fold activation of the minimal CAD promoter by a single Honk palindrome, a pair of Honk palindromes upstream of a synthetic promoter activated transcription 4-fold. This 4-fold activation was modest compared to the almost 200-fold activation by the 21-bp repeats from simian virus 40; it is possible, however, that the arrangement of the Honk sites in the synthetic promoter construct did not allow optimal interaction between the Honk protein and the basal transcription machinery or that Honk sites activate strongly only under particular physiological conditions. The latter case would be similar to the activation of transcription by the transcription factor E2F, which binds to the mouse dihydrofolate reductase promoter between –8 and +1; this E2F site is necessary and sufficient for growth regulation of transcription, but is not required for basal transcription (9, 19). To determine the identity and the function of the Honk protein, we now need to define its exact binding site.

**DISCUSSION**

RNA polymerase II initiates transcription of the cellular CAD gene at a single major site (2, 3). Our analysis demonstrated that start site selection at the hamster CAD promoter was not controlled by the most common positioning element, a TATA box around –30. First, the ability of a sequence to recognize the DNA-binding subunit of TFIIID and to direct initiation 30 bp downstream correlates with its similarity to the TATA box consensus (4); the profile of sequence similarity to the TATA box consensus showed no significant matches in the minimal CAD promoter, in particular not around –30. Second, mutating the region between –30 and –22 did not affect transcription in vivo. The fortuitous creation of a better TATA box in plasmid pCM–31 (78% similarity to the consensus) increased promoter activity, but only in vitro. In vivo, the activity of this mutant may have been identical to that of the wild-type promoter because the DNA was less accessible and other steps than binding of TFIIID to the CAD promoter were limiting in the cell. Initiation at the CAD promoter, as at other TATA-less promoters (4), probably still requires TFIIID. We have not pursued this question because heat treatment of nuclear extract inactivates Sp1 before TFIIID (5) and because a reconstituted system for transcription of the CAD promoter has not yet been developed.

The Honk palindrome between –30 and –13 bound another activator protein, but had no apparent role in start site selection. Our results suggested that only the downstream half of the Honk palindrome was required for activation and protein binding. First, the 28-bp footprint over the Honk site is centered at –17 rather than at the middle of the palindrome at –21 (2). Second, mutating the region from –21 to –13 reduced the activity of the minimal CAD promoter in vivo to one-half, whereas mutating the region from –30 to –22 had no effect. In a manner similar to the 2-fold activation of the minimal CAD promoter by a single Honk palindrome, a pair of Honk palindromes upstream of a synthetic promoter activated transcription 4-fold. This 4-fold activation was modest compared to the almost 200-fold activation by the 21-bp repeats from simian virus 40; it is possible, however, that the arrangement of the Honk sites in the synthetic promoter construct did not allow optimal interaction between the Honk protein and the basal transcription machinery or that Honk sites activate strongly only under particular physiological conditions. The latter case would be similar to the activation of transcription by the transcription factor E2F, which binds to the mouse dihydrofolate reductase promoter between –8 and +1; this E2F site is necessary and sufficient for growth regulation of transcription, but is not required for basal transcription (9, 19). To determine the identity and the function of the Honk protein, we now need to define its exact binding site.

**The primary positioning element.** Bottom, the spacing between upstream and downstream sequences of the minimal CAD promoter was altered by replacing 26 bp from –12 to +14 with a shorter, 6-bp linker (–13–20(+15)) or by inserting a 15-bp linker between –45 and –44 (–45/+15); the proximal Sp1 site was replaced with two different linker sequences as shown in Fig. 1A (–55/+44 and –55/BssHII/+44). COS cells were transfected with plasmids pGL2BasicSV (vector), pCM–13[–21 (78% similarity to the consensus) increased promoter activity, but only under particular physiological conditions. The latter case would be similar to the activation of transcription by the transcription factor E2F, which binds to the mouse dihydrofolate reductase promoter between –8 and +1; this E2F site is necessary and sufficient for growth regulation of transcription, but is not required for basal transcription (9, 19). To determine the identity and the function of the Honk protein, we now need to define its exact binding site.

**DISCUSSION**

RNA polymerase II initiates transcription of the cellular CAD gene at a single major site (2, 3). Our analysis demonstrated that start site selection at the hamster CAD promoter was not controlled by the most common positioning element, a TATA box around –30. First, the ability of a sequence to recognize the DNA-binding subunit of TFIIID and to direct initiation 30 bp downstream correlates with its similarity to the TATA box consensus (4); the profile of sequence similarity to the TATA box consensus showed no significant matches in the minimal CAD promoter, in particular not around –30. Second, mutating the region between –30 and –22 did not affect transcription in vivo. The fortuitous creation of a better TATA box in plasmid pCM–31 (78% similarity to the consensus) increased promoter activity, but only in vitro. In vivo, the activity of this mutant may have been identical to that of the wild-type promoter because the DNA was less accessible and other steps than binding of TFIIID to the CAD promoter were limiting in the cell. Initiation at the CAD promoter, as at other TATA-less promoters (4), probably still requires TFIIID. We have not pursued this question because heat treatment of nuclear extract inactivates Sp1 before TFIIID (5) and because a reconstituted system for transcription of the CAD promoter has not yet been developed.

The Honk palindrome between –30 and –13 bound another activator protein, but had no apparent role in start site selection. Our results suggested that only the downstream half of the Honk palindrome was required for activation and protein binding. First, the 28-bp footprint over the Honk site is centered at –17 rather than at the middle of the palindrome at –21 (2). Second, mutating the region from –21 to –13 reduced the activity of the minimal CAD promoter in vivo to one-half, whereas mutating the region from –30 to –22 had no effect. In a manner similar to the 2-fold activation of the minimal CAD promoter by a single Honk palindrome, a pair of Honk palindromes upstream of a synthetic promoter activated transcription 4-fold. This 4-fold activation was modest compared to the almost 200-fold activation by the 21-bp repeats from simian virus 40; it is possible, however, that the arrangement of the Honk sites in the synthetic promoter construct did not allow optimal interaction between the Honk protein and the basal transcription machinery or that Honk sites activate strongly only under particular physiological conditions. The latter case would be similar to the activation of transcription by the transcription factor E2F, which binds to the mouse dihydrofolate reductase promoter between –8 and +1; this E2F site is necessary and sufficient for growth regulation of transcription, but is not required for basal transcription (9, 19). To determine the identity and the function of the Honk protein, we now need to define its exact binding site.

**CONSENSUS**

<table>
<thead>
<tr>
<th>Initiator</th>
<th>posición +1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus</td>
<td>KCABHYBY</td>
</tr>
<tr>
<td>Wild Type</td>
<td>TCCAGCCTAGTACGATTAGG</td>
</tr>
<tr>
<td>A</td>
<td>TCCAGCCTAGATCTTATTAGG</td>
</tr>
<tr>
<td>C</td>
<td>TCCAGCCTAGACTATTAGG</td>
</tr>
<tr>
<td>D</td>
<td>TCCAGCCTAGATGATTAGG</td>
</tr>
<tr>
<td>Mutants</td>
<td>TCCAGCCTAGTACGATTAGG</td>
</tr>
</tbody>
</table>

**FIG. 5. THE CAD INITIATOR.** The sequence between –2 and +6 in the minimal CAD promoter was replaced with random sequences. COS cells were transfected with plasmid pC–81/+26GLSV (wild type) or with plasmids pCM678GLSV, pCM692GLSV, pCM697GLSV, and pCM703GLSV (mutants); start sites of transcription were mapped by primer extension of cytoplasmic RNA (lanes 5–9). Lanes 1–4, sequencing reactions of the wild-type plasmid with the same primer; arrowhead, position +1. Bottom, the initiator consensus (18) and the sequences of the four mutants aligned to the wild-type sequence from –8 to +12. L arrowhead, base pairs that differ from the wild-type. The similarity to the initiator consensus in percent is indicated above each position that served as a start site. Mutants 697 and 703 had better matches to the consensus at +7 (65%) and at –2 (68%), respectively.
We have previously shown that Sp1 stabilizes preinitiation complexes at the CAD promoter (5). Here we describe observations that are consistent with the hypothesis that the proximal Sp1 site at -49 serves as the minimal TATA box as the primary positioning element. Transcription started at a fixed distance of about 50 bp downstream of this Sp1 site when sequences were inserted into or deleted from the minimal CAD promoter; when this Sp1 site was mutated, initiation at +1 was abolished. It is possible that in the mutant -451+15[-44 it was a primary positioning element. Transcription started at a fixed distance to the TATA promoter; when this Spl site was mutated, initiation at a candidate for such a positioning element, was 60% or less and therefore not significant. The start site of transcription in this insertion mutant shifted to a poor match to the initiator consensus at -19 (38% similarity) and not to the excellent match at -16 (94% similarity). This start was further upstream than the expected 15 bp; Honk protein bound to this region may have prevented initiation downstream of -19. We also observed initiation about 50 bp downstream from the distal Sp1 site at -70 in the wild-type promoter in vivo and after mutating the proximal Sp1 site. An Sp1 site is observed in transcription from the cellular CAD promoter; perhaps the six Sp1 sites contained in the origin of replication from simian virus 40 that was inserted into the plasmids used in these experiments, enhanced the ability of the distal Sp1 site in the minimal CAD promoter to direct initiation of transcription. While it is not yet generally accepted that an activator binding site can act as a positioning element, several such sites in other TATA-less promoters have been reported to direct initiation at similar distances: the proximal Sp1 site in the hamster dihydrofolate reductase promoter about 45 bp downstream (20); four binding sites for the glucocorticoid receptor about 45 to 55 bp downstream, preferentially at consensus initiator sequences (21, 22); the proximal sequence element in the toad U1 small nuclear RNA promoter about 57 bp downstream, at consensus initiator sequences within a small window (23). Two more lines of evidence suggest that Sp1 acts at such a close range. Although the distance to the start site is more variable for an Spl site than for TATA upstream of the start site (18). The activation domain of Sp1 fused to the DNA binding domain of the transcription factor GAL4 from yeast stimulates transcription only from a proximal promoter position, but not from a remote enhancer position (24). We therefore propose that in the hamster CAD promoter, Sp1 not only activates transcription, with its proximal binding site, Sp1 also performs the positioning function of the basal transcription factor TFIIID and the TATA box. This may, however, not be true for all promoters that contain Spl binding sites. In contrast, yeast and fruit fly do not contain Sp1 protein; the promoters of their CAD homologs, the URA2 gene (25) and the rudimentary gene (26), contain a TATA box. The similarity of the sequence between -2 and +6 to the initiator consensus suggested a role of this region in initiation. In support of this hypothesis, the start site pattern in one mutant of this region changed and the transcriptional activity in two other mutants was reduced as predicted from the consensus. Furthermore, the sequence of the CAD promoter between -2 and +6 could substitute for the initiator of the mouse terminal deoxynucleotidyl transferase gene (8) in synthetic promoters with Sp1 binding sites or Sp1 binding sites and a TATA box and direct efficient transcription. However, substitution of the sequence between -2 and +6 in the minimal CAD promoter with another good match to the consensus did not always preserve transcriptional activity (Fig. 5, mutant 692), nor did substitution with a poor match always reduce transcription activity (Fig. 1A, -3[2]+3 in vivo). These results indicate that the sequences around the start site contribute to the activity of the CAD promoter, but that the presence of a consensus initiator in this region is neither necessary nor sufficient for transcription. Perhaps the consensus in its present form does not describe the sequence requirements for an initiator completely, so that some sequences with a high similarity score do not function as well as predicted; analyzing more promoters may yield a better consensus. Alternatively, there may be different initiator families that require the presence of different activators (27), but a statistical analysis of base frequencies would not distinguish between them; this hypothesis could be tested by interchanging the initiators in different promoters.

In summary, we demonstrated that the sequences of the hamster CAD promoter between -81 and +26 were necessary and sufficient for accurate and efficient initiation of transcription. Two Sp1 binding sites, a binding site between -21 and -13 for a ubiquitous activator protein, and the sequence around +1 contributed to the activity of the CAD promoter; the proximal Sp1 binding site played a major role in determining the start site. We can now focus on the promoter elements and cellular factors that regulate CAD transcription at the boundary between G1 and phase of the cell cycle.

Acknowledgments—We thank Chris Bartley, Cathryn Lundberg, and Stephanie McMahon for technical assistance; Stephen Smale for plasmids IV and VII; Janet Mertz for plasmid pGEMHindIII and COS-M6 cells; Petes Bell for rat liver extract; Jennifer Martin for T cells; and David Brown, Nancy Thompson, Paul Lamb, and members of the Farnham laboratory for comments on the manuscript.

REFERENCES