Transcriptional Initiation Is Controlled by Upstream GC-Box Interactions in a TATAA-Less Promoter

MICHAEL C. BLAKE,[1,2] ROBERT C. JAMBOU,[1] ANDREW G. SWICK,[1]† JEANNE W. KAHN,[1]
AND JANE CLIFFORD AZIZKHAN[1,2,3]*

*Lineberger Cancer Research Center,[1] Departments of Pediatrics and Pharmacology,[3] and Curriculum in Genetics,[2]
University of North Carolina, Chapel Hill, North Carolina 27599-7295*

Numerous genes contain TATAA-less promoters, and the control of transcriptional initiation in this important promoter class is not understood. We have determined that protein-DNA interactions at three of the four proximal GC box sequence elements in one such promoter, that of the hamster dihydrofolate reductase gene, control initiation and relative use of the major and minor start sites. Our results indicate that although the GC boxes are apparently equivalent with respect to factor binding, they are not equivalent with respect to function. At least two properly positioned GC boxes were required for initiation of transcription. Abolishment of DNA-protein interaction by site-specific mutation of the most proximal GC box (box I) resulted in a fivefold decrease in transcription from the major initiation site and a threefold increase in heterogeneous transcripts initiating from the vicinity of the minor start site in vitro and in vivo. Mutations that separately abolished interactions at GC boxes II and III while leaving GC box I intact affected the relative utilization of both the major and minor initiation sites as well as transcriptional efficiency of the promoter template in in vitro transcription and transient expression assays. Interaction at GC box IV when the three proximal boxes were in a wild-type configuration had no effect on transcription of the dihydrofolate reductase gene promoter. Thus, GC box interactions not only are required for efficient transcription but also regulate start site utilization in this TATAA-less promoter.

Transcription of eucaryotic promoters by RNA polymerase II involves multiple sequence elements and protein factors that associate with these sequences. Certain DNA-protein interactions regulate the efficiency of transcriptional initiation, while others have the additional role of specifying the transcriptional initiation site. In many class II gene promoters, a TATAA sequence element is located approximately 25 to 30 bp upstream of the transcription start site (5, 17); interaction of a factor(s) with this sequence specifies the site of initiation in many of these promoters (24, 41). However, in other eucaryotic promoters (7, 25, 47), interaction with TATAA appears not to specify the start site but rather to control the efficiency of transcription from a downstream initiation site. Another common important control element, CCAAT, is the target of factors that regulate the efficiency of transcription (13, 24).

A large subclass of polymerase II promoters lacks both TATAA and CCAAT sequence motifs but contains multiple GC boxes. This promoter class includes several housekeeping genes (e.g., the genes encoding dihydrofolate reductase [DHFR] [reference 2 and references therein], hydroxymethylglutaryl coenzyme A reductase [39], hypoxanthine guanine phosphoribosyltransferase [33], and adenosine deaminase [46]) as well as nonhousekeeping genes (e.g., the transforming growth factor alpha [9, 23], rat malic enzyme [36], human c-Ha-ras [21], epidermal growth factor receptor [22], and nerve growth factor receptor [42] genes). The sequences and factors that specify the transcription start sites in these promoters have not been identified. Several factors have been found to interact with GC boxes, most notably Sp1 (for

a review, see reference 31); more recently, the factors LSF (28), ETF (26), GCF-1 (27), and AP2 (35) have also been shown to interact with GC boxes. Determining the functional role of multiple GC boxes in the absence of TATAA and CCAAT motifs is crucial to the understanding of transcriptional regulation of this important class of promoters.

It has been shown that GC boxes are required for efficient promoter activity in the genes in which they have been analyzed. Whether more than one GC element is absolutely required in the absence of the TATAA and CCAAT motifs is not clear, yet in all examined cases, more than one GC box is present. Many of these promoters have unique transcription start sites, while others display multiple but specific start sites. The role of spacing and positioning of the GC boxes within these promoters, both in relation to each other and with respect to the initiation site(s), is not entirely clear, although this study begins to address some of the important features.

The hamster DHFR promoter contains neither TATAA nor CCAAT sequence elements (2, 34). There are four GC boxes in the 210 bp 5' to the DHFR-coding sequence and two binding sites for the transcription factor E2F immediately 3' to the major transcription start site (8). The DHFR major start site in the hamster gene has been mapped to nucleotide position −63 relative to ATG (34). A minor transcriptional start site that accounts for 15 to 20% of DHFR transcription is located at position −107 relative to ATG (34). The center of the first GC box is positioned 45 bp upstream of the major transcription start site, while the center of the second GC box is 45 bp upstream of the minor start site. Interestingly, the minor start site is located in the center of the first GC box. This spatial arrangement is conserved among mammalian DHFR genes, suggesting that the relative distance of the initiation site to the center of each upstream GC box is

---

important for specifying the start site or regulating the efficiency of transcription.

Using the hamster DHFR gene as a model for TATAA-less promoters, we previously demonstrated that a GC box-binding factor is required for transcription and that a truncated promoter containing one GC box is transcriptionally inactive (44). In the study presented here, site-directed mutations in individual GC boxes within the DHFR promoter were used to carefully examine the function of these control elements. Our study demonstrates that the DNA-protein interactions occurring at the GC boxes in the DHFR promoter are functionally distinct and that factors binding to the GC boxes must interact in a position-dependent manner.

(This work was performed by M. C. Blake in partial fulfillment of the requirements for the Ph.D. degree from the Curriculum in Genetics, University of North Carolina.)

MATERIALS AND METHODS

Preparation of nuclear extracts. HeLa-S3 cells were grown in suspension culture and maintained at a density of 4×10^5 to 8×10^5 cells per ml in Joklik modified essential medium supplemented with 10% defined calf serum (HyClone) and penicillin-streptomycin. Nuclear extracts were prepared from these cells as described by Dignam et al. (15), with the following modifications. After homogenization of cells in hypotonic medium containing spermine and spermidine, nuclei were recovered by centrifugation at 30,000 × g for 30 s. Nuclei were then extracted in buffer C (20 mM N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid [HEPES; pH 7.9], 0.2 mM EDTA, 0.2 mM EGTA, 2 mM dithiothreitol [DTT], 25% glycerol, 0.15 mM spermine, 0.75 mM spermidine, 1 mM phenylmethylsulfonyl fluoride, 0.4 M NaCl). The extract was recovered by centrifugation at 150,000 × g for 90 min, followed by dialysis against modified buffer D (20 mM HEPES [pH 7.9], 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.2 mM EGTA, 2 mM DTT, 1.0 mM phenylmethylsulfonyl fluoride, 12.5 mM MgCl₂). The extract was stored in aliquots at −70°C.

DNase I footprinting. DNase I footprinting was carried out by a modification of the technique of Galas and Schmitz (16). Typically, 5 to 10 ng of end-labeled DNA was incubated with various amounts of nuclear extract in the presence of 15 μg of poly(dI-dC) in a 120-μl reaction volume containing 6.1% glycerol, 0.07 mM EDTA, 0.07 mM EGTA, 7.2 mM HEPES (pH 7.9), 39 mM KCl, and 7.5 mM MgCl₂. The binding reaction mixtures were incubated at room temperature for 30 min and adjusted to 2 mM final CaCl₂. Freshly diluted DNase I (Bethesda Research Laboratories, Inc.) was added to the reaction mixtures, and digestion was allowed to proceed for 4 min at room temperature. Digestion was stopped by adding 2 volumes of a mix containing 100 mM Tris hydrochloride (pH 8), 100 mM NaCl, 20 mM EDTA (pH 8), 0.1% sodium dodecyl sulfate, 100 μg of proteinase K per ml, and of 100 μg of tRNA per ml. Following incubation at 37°C for 20 min, the samples were extracted with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1) and ethanol precipitated. The samples were resuspended in 7 M urea dye mix and electrophoresed through 8% polyacrylamide gels (19:1 acrylamide/bisacrylamide) in 0.5× TBE buffer. The gels were exposed to X-ray film overnight at −70°C with an intensifying screen.

Oligonucleotide-directed mutagenesis. Mutations in the DHFR GC boxes were constructed by oligonucleotide-directed mutagenesis (30). To construct these mutants, the −210 DHFR promoter deletion construct was subcloned into

M13mp18, and the resulting phage was passaged through *Escherichia coli* CJ236 (obtained from T. Kunkel, National Institute of Environmental Health Sciences), which is deficient in uracil N-glycosylase and dUTPase activities. Uracil-containing single-stranded DNA was then purified and used as a template for second-strand synthesis, using mutagenic oligonucleotides with the following sequences (substitutions from the wild-type sequences shown in Fig. 1 are underlined):

for GC box I: 5'-GGAGGCCACTTTCTCAACAAG-3'
for GC box II: 5'-CAGAGGCCCTTTCTCCACGAG-3'
for GC box III: 5'-GTCAGACTCTTTCTCCACCAG-3'
for GC box IV: 5'-GGCCGGCCCTTTCCAGTCGCG-3'

These oligonucleotides bear a 3-bp mismatch to the wild-type sequences in the core hexanucleotide of each GC box. Each mutagenic primer (12.5 ng) was annealed to 200 ng of −210 template DNA. Second-strand synthesis was initiated by the addition of T4 DNA ligase (Boehringer Mannheim) and modified T7 DNA polymerase (Sequenase; U.S. Biochemical Corp.) to the reaction in a volume of 20 μl (final buffer conditions: 20 mM Tris [pH 7.5], 10 mM MgCl₂, 50 mM NaCl, 6.7 mM DTT, 1 mM ATP, 0.4 mM each dATP, dGTP, dCTP, and dTTP). Transformation of the synthesis products into *E. coli* DH5αF' (Bethesda Research Laboratories) produced plaques which were screened for the desired mutations by DNA sequencing by the chain termination method (40).

In vitro transcription assays. In vitro transcription reactions were performed in a final volume of 25 μl containing 400 μM each ATP, CTP, and UTP, 0.5 μM [α-³²P]GTP (800 Ci/mmol; 10 μCi per reaction), 1 mM creatine phosphate, and 140 μM EDTA. The transcription reactions were initiated by the addition of linearized plasmid template DNA and 15 μl of nuclear extract (150 μg of protein). The reaction mixtures were incubated at 30°C for 60 min, and the reaction was stopped by the addition of a mixture containing 8 M urea, 0.5% sodium dodecyl sulfate, 10 mM EDTA, and 10 mM Tris (pH 8.0). A 494-base Sp6 transcript was included in the stop mixture as a control for sample recovery during processing. The samples were phenol-chloroform extracted twice, ethanol precipitated, and electrophoresed on 4% polyacrylamide–7 M urea denaturing gels in 0.5× TBE. Autoradiography was performed as described above.

RNA analysis. Primer extension analysis of in vitro-transcribed and transfected DHFR RNAs was accomplished by using a synthetic 40-base oligonucleotide (5'-ACGATGCCA TTGGGATATATCAACGGTGGTATATCCAGTG-3') complementary to the bacterial chloramphenicol acetyltransferase (CAT) mRNA sequence, which hybridizes 116 nucleotides downstream of the major transcriptional start (−63 relative to ATG). Analysis of cellular poly(A)-selected RNA obtained from a methotrexate-resistant, DHFR-amplified CHO cell line (C400) was performed by using a second 40-mer oligonucleotide (5'-CCAGTACCTGAGCATTGGCC AGGGAAGGTCTCCGTTCTTG-3') that also hybridized 116 nucleotides downstream of the major transcription start site but was complementary to the DHFR mRNA.

RNA templates obtained from in vitro-transcribed plasmid constructs were synthesized and purified as described above but did not contain label. In most cases, enough template RNA was produced in a single transcription reaction to produce a detectable signal; GC box mutants required twice the amount of RNA template. RNA obtained from amplified cells was purified by the guanidinium isothiocyanate method

(12), followed by oligo(dT) selection as described by Aviv and Leder (1). Calcium phosphate-transfected HeLa cell RNA templates were purified and directly poly(A) selected as described by Badley et al. (3). Visualization of primer extension products required 5 μg of amplified poly(A)$^+$ cellular RNA and 10 μg of poly(A)$^+$ RNA from transfected cells.

The oligonucleotides were end labeled with [γ-^{32}P]ATP in the presence of T4 polynucleotide kinase and hybridized to the appropriate templates essentially as described previously (6) except that the formamide concentration was lowered to 50% and the hybridization mix was incubated at 42°C for 8 to 10 h before precipitation and extension. Extension reactions were performed as described previously (29), using avian myeloblastosis virus reverse transcriptase (Life Sciences). Primer extension products were resolved on 8% denaturing polyacrylamide gels. A chain termination sequence ladder (40) was run next to the primer extension products for precise mapping of the initiation site. This ladder was generated by using the CAT oligonucleotide described above.

Transient expression assays of DHFR promoter activity. HeLa cells were maintained in monolayer in Eagle minimal essential medium supplemented with 10% defined calf serum (HyClone) and antibiotics. The −210 wild-type and −210 mutant constructs were transfected into HeLa cells by the calcium phosphate coprecipitation method (20). Super-coiled, double cesium chloride-banded plasmid DNA (10 μg) was precipitated for 30 min in 500 μl of HEPES-buffered saline by the addition of $CaCl_2$ to 6.0 mM. The DNA precipitate was layered onto a washed, well-drained mono-layer containing ca. 10^6 cells per 10-cm-diameter dish. Complete medium was added after 30 min. After 4 h, the medium was aspirated and the cells were treated with 12.5% glycerol in HEPES-buffered saline for 4 min; the monolayer was then washed, and complete medium was added. The cells were incubated for 48 h prior to harvest. Freeze-thaw lysates were assayed for protein content by the Bradford method (10), and samples normalized for protein content were assayed for CAT activity, using [^{14}C]chloramphenicol as a substrate. Acetylated [^{14}C]chloramphenicol was separated from the nonacetylated forms by thin-layer chromatography, followed by autoradiography (19). Percentage conversion of chloramphenicol to its acetylated forms was determined by liquid scintillation counting of isolated spots from thin-layer chromatography plates.

RESULTS

The important features of the hamster DHFR promoter construct used as a template for site-directed mutagenesis are summarized in Fig. 1A. This promoter fragment, which has maximal activity in in vitro transcription and transient expression assays (44), contains four GC boxes that will be referred to as GC boxes I, II, III, and IV. The study described here addresses the functional role of GC boxes in controlling initiation of transcription and start site utilization in the DHFR promoter.

Generation of mutations that abolish specific GC box inter-actions in the DHFR promoter. To analyze the nature of the interactions occurring at each GC box in detail, we used site-specific mutagenesis to change each consensus GC hexanucleotide (5′-GGGCGG-3′) to one that would no longer conform to this consensus. This mutation (5′-GGAAAG-3′) was based on previous studies of the simian virus 40 early promoter which demonstrated that the GCG-
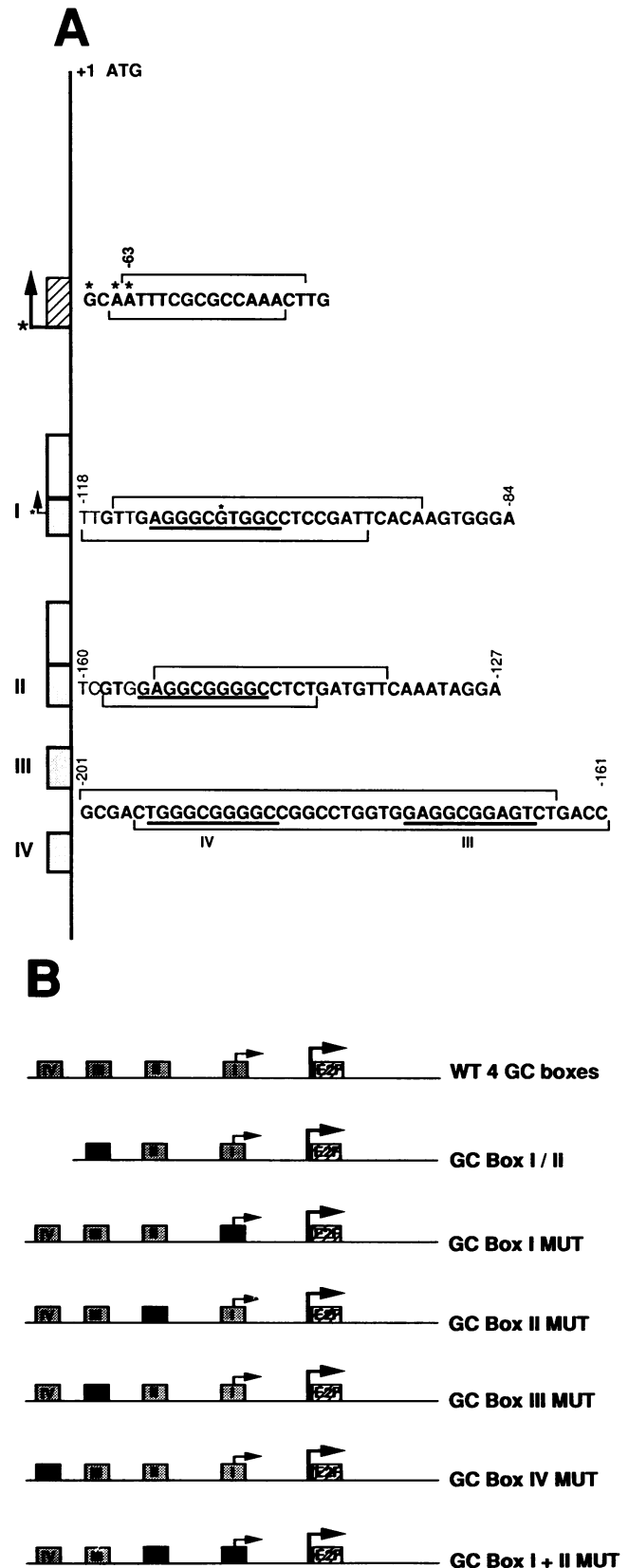
FIG. 1. (A) Diagrammatic representation of the hamster DHFR promoter. The major and minor start sites are shown by asterisks over the sequence. Symbols: ▨, pair of overlapping E2F-binding sites immediately 3' to the major transcription start site; ▢, GC boxes numbered IV, III, II, and I in a 5'-to-3' context; □, two copies of a conserved sequence element in the promoter (element 3). Brackets above and below the sequence indicate sequences protected from cleavage by DNase I in footprinting assays. (B) Diagrammatic representation of the mutated constructs used in the experiments. Rectangles I to IV represent functional (i.e., binding) GC boxes; black boxes indicate site-specifically mutated (nonbinding) GC boxes. All constructs containing four GC boxes originate at nucleotide −210 with respect to the start of translation; constructs containing three GC boxes originate at nucleotide −184. The wild-type construct is designated WT 4 GC boxes.

to-AAA mutation, when introduced into the simian virus 40 GC box consensus, abolished specific binding of Sp1 (18). The GC box mutations were made to examine the functional consequences of eliminating a DNA-protein interaction at a specific GC box in the DHFR promoter. The constructs analyzed are diagrammed in Fig. 1B. By deletion of box IV and site-directed mutagenesis of GC box III (see Materials and Methods), a promoter containing two functional GC boxes, I and II, was constructed. Mutations in each of the GC boxes were made. In all subsequent experiments, the DHFR promoter constructs shown in Fig. 1B were fused 5' to the CAT gene. This allows direct comparison of in vitro transcription and transient expression of CAT.

Each of the four constructs bearing a mutation in a single GC box was subjected to DNase I footprinting analysis in which HeLa nuclear extract was incubated with equal amounts of end-labeled promoter fragment from each of the DHFR promoter mutants to allow approximate quantitation of the amount of protein required to obtain a footprint and to assess the effect of a mutation on the other DNA-protein interactions. DNase I footprinting assays of the constructs bearing mutations in each of the four GC boxes demonstrated that mutation of a given GC box resulted in the abolishment of the corresponding DNA-protein interaction (i.e., footprint at the targeted site) (Fig. 2). There is no obvious cooperativity in binding to the GC boxes in this assay, since in each case the appearance of the footprint over the remaining GC boxes occurs at the same concentration of protein extract as in the control sequence and is not obviously changed in its boundaries. These mutations were then analyzed for DHFR promoter activity as described below.

Transcription of DHFR GC box mutants in vitro: change in start site utilization. In vitro runoff transcription of each of the mutants is shown in Fig. 3. Wild-type and mutant DHFR constructs were linearized at a ScaI site in the CAT-coding sequence, and the transcripts generated from these constructs in the presence of HeLa nuclear extract were compared by autoradiography of electrophoretically separated transcripts. The control construct, designated WT 4 GC boxes, produced two runoff products; the lower band corresponds to transcription initiation at the major start site (previously mapped to position −63 [34]), and the upper band corresponds to the minor transcription start site at position −107. Constructs containing three (Δ 3) or two (Δ 2) GC boxes were also tested in this experiment. The major transcript from the construct containing only GC boxes I and II was decreased twofold in comparison with the wild-type level; in addition, transcription arising from the minor start site was barely detectable with this construct. Since the

two-GC box construct (Δ 2 GC boxes) contains one GC box upstream of the minor start site and two GC boxes upstream of the major transcription start site, we reasoned that efficient utilization of each DHFR start site may require two upstream GC boxes.

The mutation of GC box I resulted in both a fivefold decrease in the major transcript and a concurrent threefold increase in transcription from the vicinity of the minor start site at −107; combined transcription was 80% of the wild-type level (Fig. 3; Table 1). The mutations abolishing GC box II or III resulted in decreased transcription from both the major and the minor start sites. In both of these mutations, the amount of transcription from the major start site was reduced 3-fold and the amount of transcription from the minor start site was reduced 10-fold (Fig. 3; Table 1). The box III mutant differed from the Δ 2 GC boxes construct in that a functional (i.e., binding) GC box IV was present in the box III mutant and was absent in the Δ 2 GC boxes construct. Ten percent of the transcripts initiated from the minor start site in the absence of box III when box IV was present, consistent with the hypothesis that two upstream GC boxes are required for initiation of transcription. The mutation in box IV did not cause a decrease in transcription from either site (and in some experiments actually resulted in increased transcription), a finding which was consistent with our earlier deletion analyses of this GC box (44). We also analyzed a double GC box mutant, Box I + II MUT. This construct contains two functional GC boxes (boxes III and IV) upstream of the minor start site. As anticipated from the effects of the individual box I and box II mutations, transcription from the major start site was nearly eliminated. Transcription arising from the vicinity of the minor start site, however, was as strong as that observed in the wild-type construct (Fig. 3, lanes 8 and 1). This result is somewhat puzzling in that one would have expected the Box I + II MUT construct to exhibit a decrease in minor start transcription consistent with the effects of the individual box II mutation (Fig. 3, lane 5). Thus, GC boxes I and II control utilization of the major start site, whereas utilization of the minor start site is controlled by any pairwise combination of GC boxes upstream of the minor start (e.g., boxes II and III or IV or boxes III and IV).

Mapping of the precise transcription start sites in vitro and in vivo. To carefully examine the specificity of the apparent reversion of DHFR transcriptional initiation to the minor start site in the GC box I mutant, primer extension assays were performed on the in vitro transcription reactions described above, using a primer complementary to the CAT mRNA sequence (Fig. 4). Dideoxy sequencing reactions were performed on the wild-type double-stranded template by using the same primer (Fig. 4, lanes G, A, T, and C). Three extension products were seen in the region corresponding to the DHFR major start site at positions −63, −64, and −66. A single extension product corresponding to the minor start site at −107 was seen. Lane 2 is a negative control in which the wild-type DHFR promoter without the CAT gene was used as a template; no extension products were seen or expected in this lane. The GC box I mutant differed from the wild type in that transcription from each of the start sites within the major cluster was decreased and transcription from the vicinity of the minor start site was increased. The primer extension analysis demonstrates that the increase in overall transcription seen in the Box I MUT runoff (Fig. 3, lane 4) was due to a new initiation site at position −105 in addition to the normal minor initiation site at −107 (Fig. 4, lane 4). It appears that additional weaker
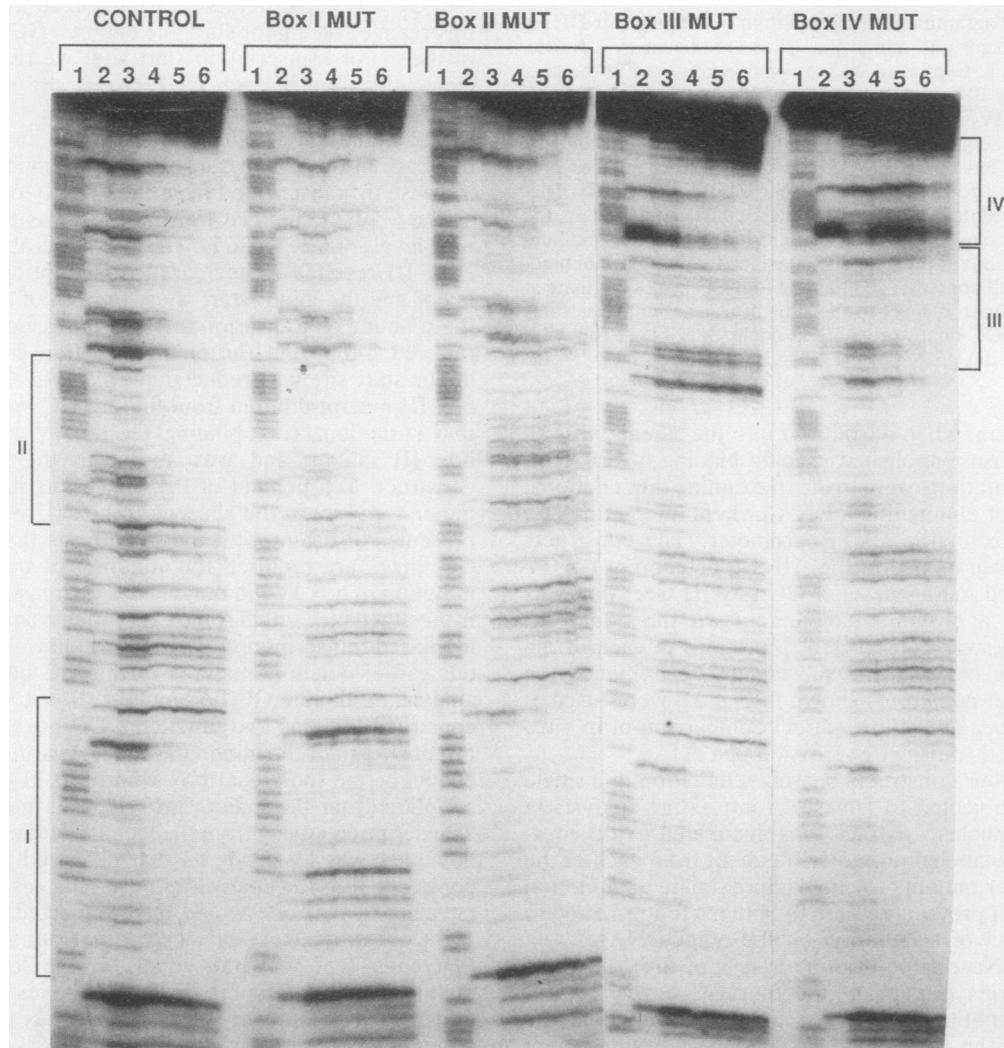
FIG. 2. DNase I footprinting of mutants in individual GC boxes. *Kpn*I-*Hin*dIII fragments containing bp -210 to -23 of the DHFR promoter were asymmetrically end labeled on the upper strand by a Klenow fill-in reaction, and 10^5 cpm of the resulting probe was incubated with 15 μg of poly(dI-dC) and various amounts of HeLa nuclear extract as described in Materials and Methods. Brackets mark the regions of the promoter protected by interactions at each GC box. Lane 1 in each series is a G+A ladder. Lanes 2 to 6 contained, respectively, 0, 30, 60, 120, and 240 μg of nuclear extract protein. DNase I concentrations used in lanes 1 to 6 were 0.01, 0.25, 0.5, 1, and 2 μg/ml, respectively.

initiation events may be occurring at nucleotide positions -108 and -104 in this mutation. Templates with mutations in GC box II or III utilized the same start sites as the wild-type construct but at decreased levels. The construct bearing mutations in both GC boxes I and II produced a pattern of transcription start sites similar to that of the GC box I mutation but at very low levels (Fig. 4, lane 6). Three additional transcripts originating 5' to nucleotide -107 were also evident in the GC Box I + II MUT construct; these additional transcripts were not detected with any of the other mutants. This increase in transcript heterogeneity occurring in the vicinity of the minor start site may explain the unexpected level of transcription observed in the Box I + II MUT runoff assay (Fig. 3, lane 8). It is possible that the decrease in transcription efficiency at nucleotide -107 as observed in the case of the individual mutations at box II or box III (Fig. 4, lanes 5 and 7) was offset by an increase in the number of transcription initiation sites as a result of the simultaneous mutation of box I and box II (Fig. 4, lane 6).

The pattern of transcription start sites obtained from the GC box IV mutation was similar to that seen in the wild-type construct. Lane 1 represents a control reaction using mRNA isolated from methotrexate-resistant CHO cells and a primer complementary to the DHFR mRNA sequence (see Materials and Methods). The purpose of this reaction was to confirm that the extension products seen in vitro corresponded with those seen with cellular mRNA. The wild-type construct (-210 CAT) had a pattern of extension products similar to that seen in vivo. Our results extend those of a previous report that mapped the major start site to position -63 (34).

Transient expression of the GC box mutants. Each GC box mutation was then examined for its ability to direct reporter gene (CAT) expression in vivo (Fig. 5A). Compared with the wild-type DHFR promoter construct, the GC box I mutation resulted in a 2.5-fold-lower level of CAT expression. Mutations in GC box II or III (Fig. 5A, lanes 3 and 4) both gave twofold-lower levels of CAT activity than did the wild-type
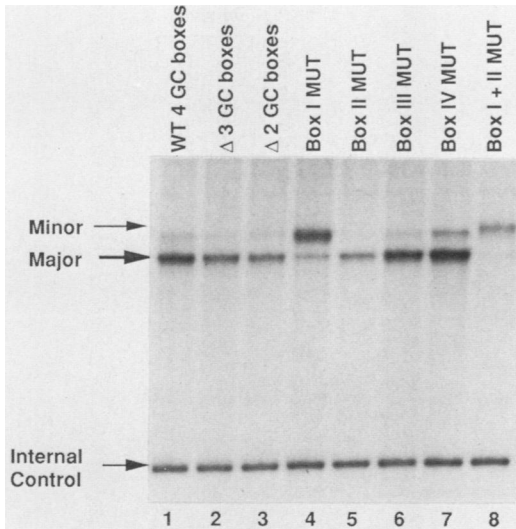
FIG. 3. In vitro transcription of DHFR promoter deletion constructs. Samples (0.5 μg each) of the wild-type, deletion constructs, and mutant −210 constructs, digested with ScaI, were used as templates for in vitro runoff transcription reactions. These templates were incubated with HeLa nuclear extract in the presence of ribonucleotides (including [α-^{32}P]GTP). RNA transcripts were processed and electrophoresed on 4% sequencing gels as described in Materials and Methods. Positions of the DHFR major and minor transcripts are indicated by arrows. A 494-base Sp6 transcript was included as an internal control for sample recovery. A diagrammatic representation of these mutants is shown in Fig. 1B.
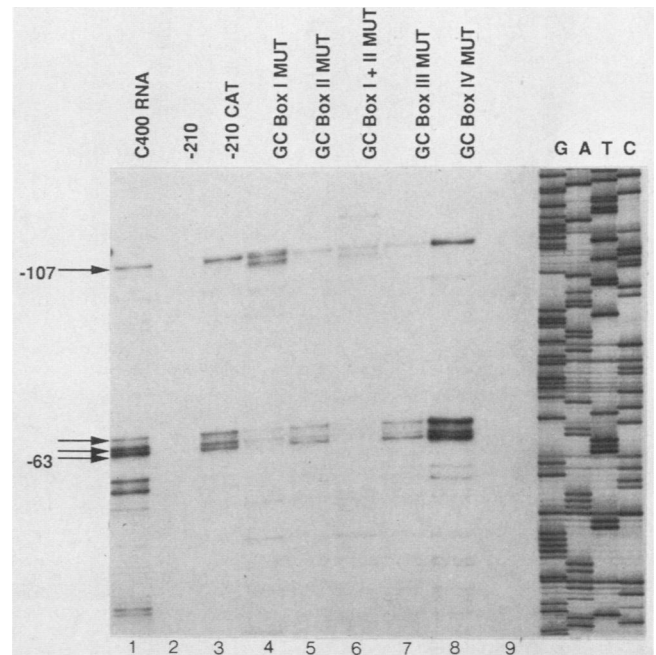

FIG. 4. Primer extension analysis of in vitro-transcribed RNA. RNA obtained from in vitro-transcribed plasmid constructs bearing various GC box mutations was used as a template for oligonucleotide-primed, reverse transcriptase-directed DNA synthesis. The oligonucleotide used to map the transcription initiation sites was complementary to the bacterial CAT mRNA. Precise localization of the transcription initiation sites was accomplished by electrophoresis of a colinear sequence ladder (lanes G, A, T, and C) using the same CAT primer. A hybridization control (−210) identical to the wild-type template but lacking the CAT sequence was included to demonstrate specific hybridization of the primer. For comparison, poly(A)-selected RNA purified from DHFR-amplified hamster cells (C400 RNA) was extended by using a second oligonucleotide complementary to the DHFR mRNA (see Materials and Methods). The wild-type DHFR template promoter construct is −210 CAT. Single GC box mutants are designated GC box n MUT, where n refers to a single box I, box II, box III, or box IV mutation. The double GC box mutant is shown as GC Box I + II MUT. Arrows indicate transcripts initiated at the major cluster (−63, −64, and −66) and at the minor start site (−107).

promoter. The GC box IV mutation did not result in lower levels of CAT activity (lane 5) in these assays. The transient expression data are summarized in Fig. 5B. Each bar represents a statistically distinct class as determined by an analysis of variance ($P < 0.05$). Thus, the effect of the GC box mutations on transient expression correlated more closely with the level of transcription initiation occurring at the major start site than with total transcription initiation from both the major and minor start sites. This result indicates that the abolishment of GC box I is more deleterious to promoter function in vivo than it is in vitro and that levels of CAT activity may be more consistent with a decrease of transcript initiation at the major start than with an increase of transcription at the minor start.

To determine whether the reversion of transcription to the vicinity of the minor start site seen in the GC box I mutation occurs in vivo as well as in vitro, primer extension assays

were used. These assays were identical to those using in vitro-transcribed RNA except that poly(A)$^{+}$-selected mRNA from transfected cells was used as the source of RNA template (Fig. 6). RNA from cells transfected with the GC box I-mutated CAT construct synthesized transcripts initiating at the same positions as seen by using in vitro-transcribed RNA, therefore confirming both the reversion of transcription to this region and microheterogeneity of start sites in vivo in the absence of an interaction at GC box I. Since transcripts initiating at the minor start site were abundant enough to permit primer extension, they are not extremely unstable.

Construction and analysis of mutations scanning a conserved region of sequence within the DHFR promoter. There is an additional conserved sequence element 3' to GC box I, the function of which has not been carefully studied. Since the GC box I mutation had a dramatic effect on start site utilization, it was important to ascertain whether this effect was mediated by the flanking sequence. This region (designated element 3) extends from position −84 to −102; GC box I extends from position −103 to −112 (Fig. 1). We wished to

TABLE 1. Summary of transcriptional data for DHFR GC box mutants

Construct	Total transcription ± SEM[a] (% of conrol)	Utilization ratio (major/minor transcripts)
Wild type	100	79/21
Box I mutant	82 ± 5.4	19/81
Box II mutant	28 ± 9.9	92/8
Box III mutant	30 ± 7.7	92/8

[a] Autoradiograms of electrophoretically separated in vitro transcription reaction products were scanned with an LKB Ultrascan densitometer, and the areas under the peaks were determined to quantitate transcript production. All areas are expressed as percentage of transcription from the wild-type promoter. Each value is the average of three separate experiments; the data in Fig. 3 represent one of these experiments. The major and minor transcripts were resolved and combined (column 2) or separately assessed (column 3).
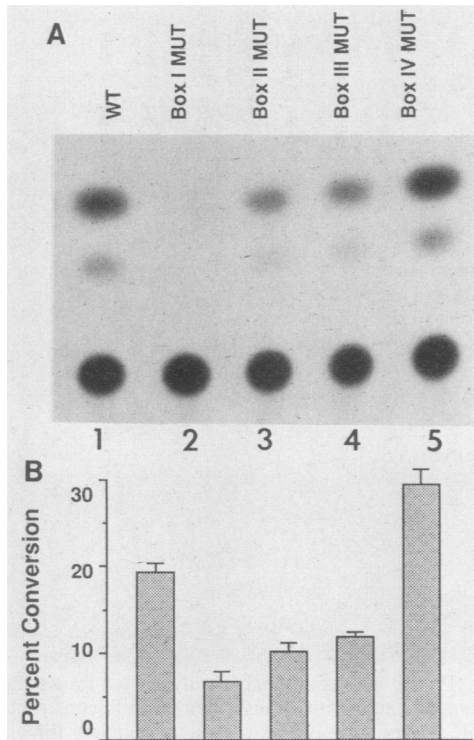
FIG. 5. (A) Transient expression assays of DHFR GC box mutants. HeLa cells in monolayer were transfected with the wild-type −210 construct (WT) or mutated −210 construct with the indicated GC box mutated (MUT). (B) Percent conversion of nonacetylated to acetylated [^{14}C]chloramphenicol over the three to six times that each DNA was tested. The bar represents the standard error of the mean.
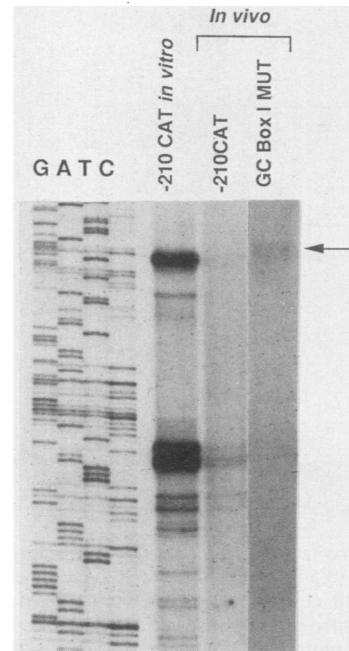

FIG. 6. Primer extension analysis of RNA obtained from transfected cells. RNA obtained from cells transfected with various GC box mutant promoter constructs fused to the bacterial CAT sequence was poly(A) selected and primer extended as described for Fig. 5. The primer extension products of the wild-type and mutated GC box RNA templates from transfected cells (in vivo) were compared to those obtained by in vitro transcription of the wild-type promoter construct. Lanes G, A, T, and C correspond to a sequence ladder generated with CAT primer used to prime RNA synthesis in the remaining lanes. Comparison is shown of in vitro RNA (−210 CAT) and RNA isolated from transfected cells in the lanes to the right. An arrow indicates the extension product resulting from the minor transcript, which is increased in cells transfected with the GC Box I MUT/CAT construct.

examine the functional importance of this sequence element, since it lies between GC box I and the major start site and might affect overall transcription or start site utilization. To analyze this sequence, we constructed mutations bearing triplet nucleotide substitutions throughout the region, from the 3' end of GC box I to the 3' end of the conserved sequence element (Fig. 7C). Mutation 1 is the same as GC Box I MUT (Fig. 1B) and is included to compare the effects of mutating the adjacent conserved sequence element.

In vitro transcription assays revealed consistent differences in the amounts of transcription obtained from three of these mutations (Fig. 7A). Mutations 3 and 5 transcribed at 60% of the wild-type level, while mutation 4 resulted in a twofold increase in transcription over the wild-type activity. These mutations were also subjected to DNase I footprinting assays, and the footprints were indistinguishable from those of the wild type; there were no detectable differences in the binding pattern over GC box I or elsewhere in the promoter (data not shown). In transient expression assays using these mutations, mutation 3 was consistently and significantly lower in expression level than the wild-type control ($P <$ 0.05) (Fig. 7B). Results for the other samples were not statistically different and exhibited variability over the six times the experiment was repeated. Thus, the effect of the GC box I mutation must be due to a factor interacting at the GC box, since utilization of start sites was not affected by mutating the conserved flanking sequence, although this flanking sequence element is clearly involved in controlling overall transcriptional efficiency.

DISCUSSION

A DHFR promoter fragment containing sequence from nucleotide positions −210 to −23 directs transcription from the same initiation sites in vitro and in vivo. The utilization of these sites is controlled so that 80% of the transcripts initiate at the major start and 20% initiate at the minor start (34). In this study, we used site-directed mutations to examine the role of GC boxes in the control of transcription initiation in the DHFR promoter. We found that two upstream GC boxes are required for efficient initiation of transcription and that mutation of the most proximal box results in decreased transcription from the major start site, which is located 45 bp 3' to it, and increased transcription from the immediate vicinity of the minor start site, which is located within the first GC box. Furthermore, mutation of GC box II or III affects both the relative utilization of the start sites and overall transcriptional efficiency. We also wish to note that the major start site actually consists of three tightly clustered initiation sites in the vicinity of the previously mapped start at nucleotide position −63 (34).

In promoters containing multiple GC boxes but lacking the TATAA box, transcription start sites may be single and specific, as observed in the nerve growth factor receptor gene (42) and the cellular retinol-binding protein gene (37), or there may be multiple heterogeneous start sites, such as those found in the c-myb (4), insulin receptor (45), and

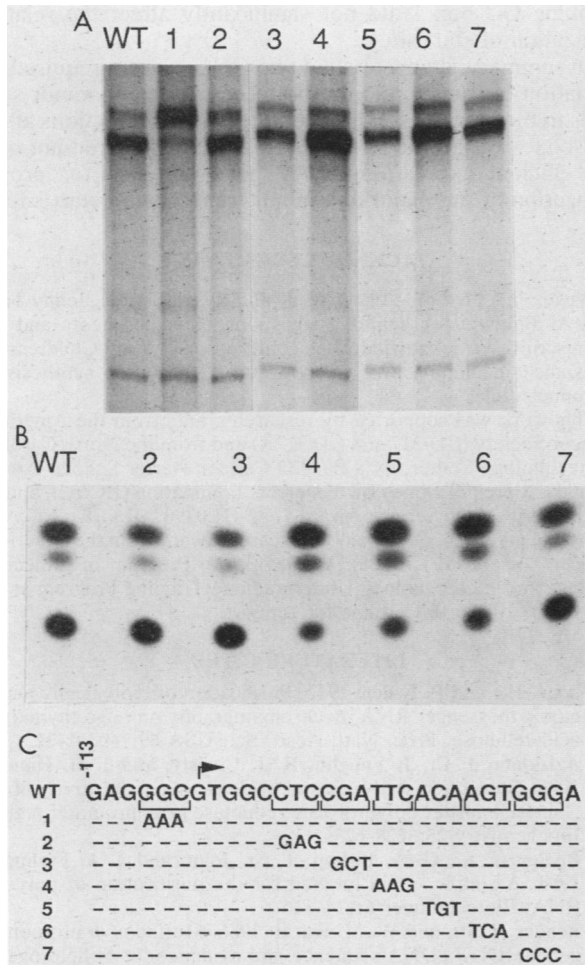TRANSCRIPTIONAL INITIATION IN A TATA-LESS PROMOTER

FIG. 7. Effects of mutations in a conserved sequence element within the DHFR promoter. In vitro transcription (A) and transient expression assays (B) were performed as described in the text with mutated constructs bearing substitutions throughout the region depicted in panel C.

c-Ha-*ras* (21) genes. Therefore, in contrast to those from TATAA-regulated genes, transcripts controlled by promoters in this class may be started at several different positions, the regulation or significance of which is not entirely clear.

GC boxes are responsible for directing transcription from the major and the minor start sites. One interpretation of the observed increase in transcription from the region of the minor start site in the absence of binding to GC box I is that occupancy of GC box I sterically interferes with initiation

from this site, as suggested in the model proposed in Fig. 8. Because the mutation of GC box I correlates with the loss of binding at that site and decreased transcription at the major start, it is tempting to predict that the interaction of a GC box-binding protein(s) with GC box I promotes transcription from the major start site.

In the murine terminal deoxynucleotidyltransferase promoter, the sequence immediately surrounding the site of initiation (Inr region) has been found to be the only sequence element required for accurate initiation of transcription (43). Placement of the simian virus 40 21-bp repeats (i.e., six GC boxes) upstream of the Inr element has been shown to dramatically increase transcription from this sequence element; optimal spacing of the GC boxes relative to the Inr was 42 to 50 bp upstream. The position of the major DHFR start site is 45 bp 3′ to the center of GC box I, and the position of the minor start site is 45 bp 3′ to the center of GC box II. Although the arrangement of GC boxes in other non-TATAA gene promoters is quite variable, they are usually found very close to the transcription start site(s). In this study, elimination of the interactions at GC box I, II, or III affected both overall transcriptional efficiency and the relative abundance of major and minor transcripts, implying that both the position of GC boxes and interaction between GC box-binding factors are important and that GC boxes in the absence of TATAA cannot function maximally at a distance. These data are consistent with the observations of Carey et al. (11) suggesting that weak activators cannot function at a distance.

Several lines of evidence suggest that at least two properly positioned upstream GC boxes are required for initiation from a start site in the hamster DHFR promoter. Transcription from the minor transcription start site is not seen in a clone with GC boxes I and II intact but missing GC boxes III and IV. Abolishing the interaction at GC box I results in decreased transcription from the major start site and increased transcription from and heterogeneity of initiation from the minor start site both in vitro and in vivo. Finally, abolishing the interaction at GC box II or III, leaving box I intact, reduces transcription from both the major and minor start sites and favors utilization of the major site (Table 1). All TATAA-less promoters have at least two GC boxes; this feature may be a functional requirement for inducing initiation in the absence of TATAA or an Inr-like sequence. Consistent with this idea, Courey et al. (14) have proposed that Sp1-Sp1 interactions may play an important and synergistic role in modulating promoter activity. It would be interesting to determine whether one GC box is sufficient to promote accurate transcription if placed 40 to 50 bp 5′ to the Inr sequence.

A recent study by Pugh and Tjian has demonstrated that the same general factors required for transcription of poly-
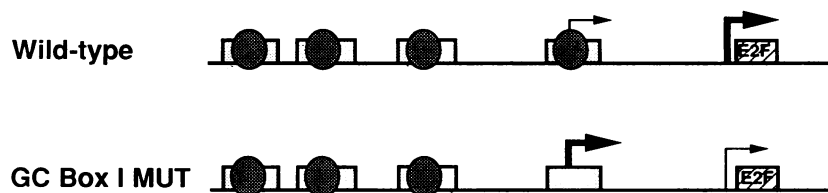

FIG. 8. Simple model for how the GC box I mutation could produce the observed transcriptional effects on the DHFR promoter. In the wild-type promoter, transcription from the minor start site would be sterically hindered by the presence of a GC box-binding factor. In turn, the interaction of this factor with GC box I promotes transcriptional initiation at the major start site. Abolishing this interaction would both decrease initiation at the major start site and increase availability of the minor start site to the transcriptional machinery.

merase II promoters containing a TATAA box are also
required for transcription of TATAA-less promoters con-
taining multiple GC boxes (38). One of the protein fractions
required for reconstitution of transcriptional activity in-
cludes the factor TFIID, which is known to interact with the
TATAA motif. Promoters that lack TATAA but contain GC
boxes require Sp1 in addition to other unpurified factors.
This fact is consistent with our earlier observation that Sp1
is required for transcription of the DHFR promoter (44).
Pugh and Tjian also suggest that Sp1 may be required to
sequester the transcriptional apparatus, presumably via pro-
tein-protein interactions. The data presented here favor this
notion, since we confirm that Sp1-binding sites clearly
regulate transcriptional initiation in a TATAA-less pro-
moter.

Specification of start sites in the DHFR promoter. The
results of the primer extension analysis indicate that start
site specification is not simply dependent on GC box inter-
actions, since most of the mutations that were analyzed
failed to significantly alter the positions of transcript initia-
tion sites. In the case of the GC box I mutation, some
initiation heterogeneity occurs in the immediate vicinity of
nucleotide −107. In addition, the double GC box mutant (GC
Box I + II MUT) produces a greater degree of transcript
heterogeneity in the same region. In some promoters, includ-
ing those of the human immunodeficiency virus type 1 and 2
(7, 25) and adenovirus E1B (47) genes, mutations in TATAA
deleterious to overall transcription and specific transcrip-
tional induction do not affect the site of initiation. In a recent
study, a sequence element immediately surrounding the
murine DHFR start site (5′-TTTCGCGCCAAACTTGACG-
3′) was found to specify initiation when placed into a
synthetic construct containing one GC box; this initiation
was independent of spacing between the GC box and this
element in the murine DHFR promoter (32). A binding
activity (designated HIP-1) was observed with this sequence
by DNase I footprinting. Studies in our laboratory had
previously demonstrated that the transcription factor E2F
interacts with the identical sequence in the hamster DHFR
promoter, which is also located at the major initiation site,
and that this interaction is required for efficient transcription
of the DHFR gene (8). We have observed that abolishing
binding to this sequence does not result in loss of specificity
of the major transcription start site; these results have been
confirmed by primer extension (unpublished data). We be-
lieve that the HIP-1 factor most likely is E2F. Therefore, it
is apparent that neither the E2F factor nor interactions at the
DHFR GC boxes alone are responsible for fixing the initia-
tion site, although we have now demonstrated that an
interaction at the first DHFR GC box dictates the relative
utilization of the major and minor start sites. These data,
taken together with the function of TATAA in some promot-
ers, suggest that the initiation site is fixed by complex
protein-protein interactions between upstream GC or
TATAA boxes and the initiator region.

**Function of additional conserved sequence flanking GC box
I.** To address whether the effect of abolishing the interaction
at GC box I was mediated by the GC box alone, we
examined the role of a conserved sequence element imme-
diately 3′ to GC box I by a series of site-directed mutations
scanning this region. This same element is found 3′ to box II,
although we have not examined its function at that location.
Within the element flanking box I, one mutation altered
transcription activity in vitro and in vivo but had no effect on
protein binding to GC box I. Mutation in the element

flanking GC box I did not significantly affect the relative
utilization of start sites.

In summary, it appears that two GC boxes are required for
initiation of transcription at both the major and minor start
sites in the hamster DHFR promoter. The interactions at the
GC boxes within the DHFR promoter are required not only
for efficient expression of the gene but also for proper
utilization of the major and minor transcription start sites.

LITERATURE CITED

1. **Aviv, H., and P. Leder.** 1972. Purification of biologically active
globin messenger RNA by chromatography on oligo thymidylic
acid-cellulose. Proc. Natl. Acad. Sci. USA **69:**1408–1412.
2. **Azizkhan, J. C., J. Vaughn, R. J. Christy, and J. H. Hamlin.**
1986. Nucleotide sequence and nuclease hypersensitivity of the
Chinese hamster dihydrofolate reductase gene promoter region.
Biochemistry **25:**6228–6236.
3. **Badley, J. E., G. A. Bishop, T. St. John, and J. A. Frelinger.**
1988. A simple, rapid method for the purification of poly A⁺
RNA. Biotechniques **6:**114–116.
4. **Bender, T. P., and W. M. Kuehl.** 1986. Murine *myb* protoonco-
gene mRNA: cDNA sequence and evidence for 5′ heterogene-
ity. Proc. Natl. Acad. Sci. USA **83:**3204–3208.
5. **Benoist, C., and P. Chambon.** 1981. In vivo sequence require-
ments of the SV40 early promoter region. Nature (London)
290:304–310.
6. **Berk, A. J., and P. A. Sharp.** 1978. Spliced early mRNAs of
simian virus 40. Proc. Natl. Acad. Sci. USA **75:**1274–1278.
7. **Bielinska, A., S. Krasnow, and G. J. Nabel.** 1989. NF-κB-
mediated activation of the human immunodeficiency virus en-
hancer: site of transcriptional initiation is independent of the
TATA box. J. Virol. **63:**4097–4100.
8. **Blake, M. C., and J. C. Azizkhan.** 1989. Transcription factor
E2F is required for efficient expression of the hamster dihydro-
folate reductase gene in vitro and in vivo. Mol. Cell. Biol.
9:4994–5002.
9. **Blasband, A. J., K. T. Rogers, X. Chen, J. C. Azizkhan, and
D. C. Lee.** 1990. Characterization of the rat transforming growth
factor α gene and identification of promoter sequences. Mol.
Cell. Biol. **10:**2111–2121.
10. **Bradford, M. M.** 1976. A rapid and sensitive method for the
quantitation of microgram quantities of protein utilizing the
principle of protein-dye binding. Anal. Biochem. **72:**248–254.
11. **Carey, M., J. Leatherwood, and M. Ptashne.** 1990. A potent
GAL4 derivative activates transcription at a distance *in vitro.*
Science **247:**710–712.
12. **Chirgwin, J. M., A. E. Przbyla, R. J. MacDonald, and W. J.
Rutter.** 1979. Isolation of biologically active ribonucleic acid
from sources enriched in ribonuclease. Biochemistry **18:**5294–
5299.
13. **Chodosh, L. A., A. S. Baldwin, R. W. Carthew, and P. A. Sharp.**
1988. Human CCAAT-binding proteins have heterologous sub-
units. Cell **53:**11–24.
14. **Courey, A. J., D. A. Holtzman, S. P. Jackson, and R. Tjian.**
1989. Synergistic activation by the glutamine-rich domains of

human transcription factor Sp1. Cell **59**:827–836.

15. **Dignam, J. D., R. L. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475–1489.

16. **Galas, D. J., and A. Schmitz.** 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5**:3158–3170.

17. **Ghosh, P. K., P. Lebowitz, F. J. Frisque, and Y. Gluzman.** 1981. Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. Proc. Natl. Acad. Sci. USA **78**:100–104.

18. **Gidoni, D., J. Kadonaga, H. Barrera-Saidana, K. Takahashi, P. Chambon, and R. Tjian.** 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science **230**:511–517.

19. **Gorman, C., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.

20. **Graham, F. L., and A. J. van der Eb.** 1973. Transformation of rat cells by DNA of human adenovirus 5. Virology **52**:456–467.

21. **Ishii, S., G. T. Merlino, and I. Pastan.** 1985. Promoter region of the human Harvey *ras* proto-oncogene: similarity to the EGF receptor proto-oncogene promoter. Science **230**:1378–1381.

22. **Ishii, S., Y.-H. Xu, R. T. Stratton, B. A. Roe, G. T. Merlino, and I. Pastan.** 1985. Characterization and sequence of the promoter region of the human epidermal growth factor receptor gene. Proc. Natl. Acad. Sci. USA **82**:4920–4924.

23. **Jakobovits, H. S., E. B. Schlokat, U. Vannice, R. Derynck, and A. D. Levinson.** 1988. The human transforming growth factor alpha directs transcription initiation from a single site in the absence of a TATA sequence. Mol. Cell. Biol. **8**:5549–5554.

24. **Jones, K. A., J. T. Kadonaga, P. J. Rosenfeld, T. J. Kelly, and R. Tjian.** 1987. A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. Cell **48**:79–89.

25. **Jones, K. A., P. A. Luciw, and N. Duchange.** 1988. Structural arrangements of transcription control domains within the 5'-untranslated leader regions of HIV-1 and HIV-2 promoters. Genes Dev. **2**:1102–1116.

26. **Kageyama, R., G. Merlino, and I. Pastan.** 1989. Nuclear factor ETF specifically stimulates transcription from promoters without a TATA box. J. Biol. Chem. **264**:15508–15514.

27. **Kageyama, R., and I. Pastan.** 1989. Molecular cloning and characterization of a human DNA binding factor that represses transcription. Cell **59**:815–825.

28. **Kim, C. H., C. Heath, A. Bertruch, and U. Hansen.** 1987. Specific stimulation of simian virus 40 late transcription *in vitro* by a cellular factor binding the simian virus 40 21 base pair repeat promoter element. Proc. Natl. Acad. Sci. USA **84**:6025–6029.

29. **Krug, M. S., and S. L. Berger.** 1987. First-strand cDNA synthesis primed with oligo(dT). Methods Enzymol. **152**:316–325.

30. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1985. Mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.

31. **McKnight, S., and R. Tjian.** 1986. Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795–805.

32. **Means, A. L., and P. J. Farnham.** 1990. Transcriptional initiation from the dihydrofolate reductase promoter is positioned by HIP-1 binding at the initiation site. Mol. Cell. Biol. **10**:653–661.

33. **Melton, D. W., D. S. Koneckl, J. Brennand, and C. T. Caskey.** 1984. Structure, repression, and mutation of the hypoxanthine phosphoribosyl transferase gene. Proc. Natl. Acad. Sci. USA **81**:2147–2151.

34. **Mitchell, P., A. M. Carothers, J. H. Han, J. D. Harding, E. Kas, L. Venolia, and L. A. Chasin.** 1986. Multiple transcription start sites, DNase I-hypersensitive sites, and an opposite-strand exon in the 5' region of the CHO *dhfr* gene. Mol. Cell. Biol. **6**:425–440.

35. **Mitchell, P. J., C. Wang, and R. Tjian.** 1987. Positive and negative regulation of transcription in vitro: enhancer binding protein AP2 is inhibited by SV40 T antigen. Cell **50**:847–861.

36. **Morioka, H., G. E. Tennyson, and V. M. Nikodem.** 1988. Structural and functional analysis of rat malic enzyme gene promoter. Mol. Cell. Biol. **8**:3542–3545.

37. **Nilsson, M. H., N. K. Spurr, J. Lundvall, L. Rask, and P. A. Peterson.** 1988. Human cellular retinol-binding protein gene organization and chromosomal location. Eur. J. Biochem. **173**:35–44.

38. **Pugh, B. F., and R. Tjian.** 1990. Mechanism of transcriptional activation by Sp1: evidence for coactivators. Cell **61**:1187–1197.

39. **Reynolds, G. A., S. K. Basu, T. F. Osborne, D. J. Chin, G. Gil, M. S. Brown, J. L. Goldstein, and K. L. Luskey.** 1984. HMG CoA reductase: a negatively regulated gene with unusual promoter and 5' untranslated regions. Cell **38**:275–285.

40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

41. **Schmidt, M. C., C. C. Kao, R. Pei, and A. J. Berk.** 1989. Yeast TATA-box transcription factor gene. Proc. Natl. Acad. Sci. USA **86**:7785–7789.

42. **Sehgal, A., N. Patil, and M. Chao.** 1988. A constitutive promoter directs expression of the nerve growth factor receptor gene. Mol. Cell. Biol. **8**:3160–3167.

43. **Smale, S. T., and D. Baltimore.** 1989. The "initiator" as a transcription control element. Cell **57**:103–113.

44. **Swick, A. G., M. C. Blake, J. W. Kahn, and J. C. Azizkhan.** 1989. Functional analysis of GC element binding and transcription in the hamster dihydrofolate reductase gene promoter. Nucleic Acids Res. **17**:9291–9304.

45. **Tewari, D. S., D. M. Cook, and R. Taub.** 1989. Characterization of the promoter region and 3' end of the human insulin receptor gene. J. Biol. Chem. **264**:16238–16345.

46. **Valerio, D., M. G. C. Duyvesteyn, B. M. M. Dekker, G. Weeda, T. M. Berkvens, L. VanderVoorn, H. van Ormondt, and A. J. van der Eb.** 1985. Adenosine deaminase: characterization and expression of a gene with a remarkable promoter. EMBO J. **4**:437–443.

47. **Wu, L., D. S. E. Rossner, M. C. Schmidt, and A. Berk.** 1987. TATA box implicated in transcriptional activation of a simple adenovirus 2 promoter. Nature (London) **326**:512–515.