Alternate Promoters and Developmental Modulation of Expression of the Chicken GATA-2 Gene in Hematopoietic Progenitor Cells*

Pascale Nony, Robert Hannon‡, Hannah Gould‡, and Gary Felsenfeld§

From the Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892-0540 and the Biomolecular Sciences Division, King's College, London WC2B 5RL, United Kingdom

The GATA family of transcription factors plays a wide range of roles in development. Three members of the family are absolutely required for normal hematopoietic development in the mouse: GATA-1 (1), GATA-2 (2), and GATA-3 (3). The hallmarks of these DNA-binding proteins, of which six have been identified so far in vertebrates, is a highly conserved (C4) zinc finger domain and a recognition of the consensus motif (T/A)GATA(A/G) in the DNA (4). However, binding sites which differ from the strict consensus originally defined are also recognized in vivo by GATA proteins (5, 6), as was already suspected from in vitro analysis (7–10).

The GATA transcription factors exhibit a distinct, although partially overlapping, tissue distribution and developmental expression profile. GATA-2 is expressed in a wide variety of tissues, which include hematopoietic progenitors, erythroid cells, mast cells, megakaryocytes, endothelial cells, the central nervous system, and the giant cells of the trophoblast (11–20). Several observations suggest that GATA-2 plays a fundamental role in hematopoietic development. In the mouse, disruption of the gene by gene targeting via homologous recombination in mouse embryonic stem cells leads to the death of embryos homozygous for the deletion. They die approximately at embryonic day 10–11 with severe anemia. In adult chimeric mice, GATA-2-deficient embryonic stem cells do not give rise to cells of any hematopoietic lineages (2). During erythroid maturation, GATA-2 mRNA is down-regulated (11, 21), whereas ectopic overexpression in chicken erythroid progenitors of GATA-2, but not GATA-1 and GATA-3, promotes their proliferation at the expense of their differentiation (22). In Xenopus and zebrafish, GATA-2 is expressed within the presumptive blood island of the embryo (16, 18, 23). Taken together, these data indicate that GATA-2 has a critical role in early hematopoietic cells, possibly influencing the maintenance or the proliferation of the progenitors.

Transcriptional regulation of the GATA-2 gene has been studied in Xenopus (18, 24, 25), the zebrafish (26), and the mouse (27). The genes have some regulatory features in common, but they also differ in important respects. In our laboratory, we have been interested in the regulation of erythroid-specific genes in the chicken, and particularly in the differential control of and activity of individual GATA family members. Because the solutions to regulatory problems found in the chicken often differ in illuminating ways from those found in other organisms, and because we study many GATA-dependent mechanisms in chicken, we cloned the chicken GATA-2 (cGATA-2)1 gene.

Earlier studies of cGATA-2 cDNA had suggested that the gene was transcribed from a single promoter far upstream of the coding region. We find, however, that cGATA-2 expression is controlled by both a proximal and a distal promoters. We show that only a minor transcript initiates from the upstream promoter, the proximal one accounting for most of the cGATA-2 transcription. The arrangement of the promoters is similar to that recently described for mouse GATA-2, and the proximal promoter also resembles the only promoter so far described in Xenopus. Nonetheless, the roles of the promoters in development and tissue-specific expression are quite different in these organisms, most strikingly in the mouse, which assigns developmental roles to its proximal and distal promoters that are quite different from those in the chicken. We suggest that although the overall organization may remain the same, the role assigned to each promoter varies among organisms. We identify distant upstream regulatory elements in the cGATA-2 gene that modulate expression from the proximal promoter and that may be responsible for this variation.

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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AF038592.

§ To whom correspondence should be addressed: Laboratory of Molecular Biology, Bldg. 5, NIDDK, National Institutes of Health, Bethesda, MD 20892-0540.
that plays a role in the affinity of the CCAAT protein for its binding site. Thus similar motifs are employed for quite different developmental tasks in chicken than they are in mouse. We show that upstream elements in cGATA-2 inhibit expression in a cell type-specific manner and suggest that they may contribute to this difference.

EXPERIMENTAL PROCEDURES

Isolation and Characterization of the Chicken GATA-2 Gene—A \( \lambda \)EMBL3 genomic library derived from chicken erythrocyte DNA (29) was screened with a cGATA-1 cDNA fragment spanning the finger region, using low-stringency washing conditions (19, 40), and 120 positive clones were purified to homogeneity through plaque hybridization screening. However, the inserts of these two clones hybridized only to 3' fragments of the cGATA-2 cDNA (11). As further screens of this library using oligo probes failed to identify any clones likely to contain the 5' missing sequences, a partial chicken genomic library was constructed in \( \lambda \)EMBL4 (Stratagene), with chicken erythrocyte DNA completely digested with EcoRI. Indeed, this restriction enzyme was shown by hybridization of a Southern blot with cGATA-2 cDNA-specific probe to generate a 17-kb genomic fragment containing the 5' part of the cDNA missing in the two phages purified from the first screening. The screening of this library allowed us to isolate several phages containing the expected 5' region. One of them, called cG2.5 was characterized in greater detail. It overlapped with 6.5 kb of pG2.11, the largest of the two clones isolated from the first screening. The position of each exon was assigned using data obtained from restriction mapping of the cDNA and the genomic clones and from exon-specific oligonucleotide hybridization. Delineation of exon-intron sequence boundaries was accomplished by DNA sequencing. The lengths of the intervening sequences were ascertained from the sizes of polymerase chain reaction products obtained using appropriate primers located in adjacent exons. Together, the clones cG2.6 and cG2.5 contained all the sequences present in the chicken GATA-2 cDNA. In order to clone more genomic DNA upstream of exon 1a, which was originally supposed to contain the regulatory sequences of the gene, a new partial genomic library was constructed in the plasmid pBluescript SK+ (Stratagene), after a double digestion of the genomic DNA with XhoI and EcoRV. These enzymes generate a 3.6-kb fragment which overlaps over 1 kb with the 17-kb EcoRI fragment. Several clones containing the expected insert were isolated, and one of them, pG2.2, was analyzed in more detail.

DNA sequencing was performed using the Thermo Sequenase system (Amersham Pharmacia Biotech). A 17-kb fragment encompassing nucleotides 1743–2315 of the cGATA-2 cDNA (11), which is located in exon 6, was inserted in pBluescript SK+ (Stratagene) to generate the exon 6 probe. Radiolabeled antisense RNA was generated using \( \alpha \)-\( ^{32} \)P\( \)CTP. A 175-bp fragment encompassing nucleotides 1743–1918 of the cGATA-2 cDNA, which map to exon 1a, was cloned in \( \alpha \)CRII plasmid (Invitrogen) to generate the exon 1 probe. Radiolabeled antisense transcripts were generated using SP6 polymerase (Stratagene) and \( \alpha \)-\( ^{32} \)P\( \)CTP. A 329-bp fragment spanning nucleotides –170 to +160 (see Fig. 5a) was cloned in pBluescript SK+ to generate the exon 1b probe. Radiolabeled antisense RNA was generated using T7 polymerase (Stratagene) and \( \alpha \)-\( ^{32} \)P\( \)CTP. All RNA preparations were purified prior to hybridization. 30 \( \mu \)g of RNA were denatured for 6 min in boiling water prior to hybridization to the exon 1 or exon 1b probes, and hybridizations were performed overnight at 65 °C. 15 \( \mu \)g of RNA were used with the exon 6 probe. In this case, hybridization was performed at 45 °C. In order to make the band intensities (shown in Figs. 2 and 3) approximately comparable, exposure times for the two autoradiograms were chosen to compensate for differences in probe size and specific activity.

Characterization of the DNase I Hypersensitive Sites—The nuclei preparations were performed on ice until DNase I digestion. Approximately 1 \( \times \)10\(^6\) cells were washed once in cold phosphate-buffered saline. Cells were lysed in 5 ml of Buffer A (10 mM Tris-HCl, pH 7.5, 10 mM MgCl2, 0.1 mM EDTA, 0.5 mg/ml 10 mM NaCl) for 5 min. Nuclei were pelleted for 5 min at 1000 \( \times \)g and washed once in Buffer A without Nonidet P-40. The pellet was resuspended in 1 ml of Buffer B (10 mM Tris-HCl, pH 7.5, 10 mM NaCl, 3 mM MgCl2, 1 mM CaCl2) at room temperature, and 100 \( \mu \)l of the nuclei preparation were used for each DNase I digestion. Approximately 0.5 \( \mu \)g of pBR322 plasmid DNA was used as an internal control (49). A 0.5 mg/ml of Proteinase K was added and incubation was performed overnight at 55 °C. The samples were next extracted with phenol-chloroform and chloroform and precipitated. For mapping of the hypersensitive (HS) sites, 1 site, 10 \( \mu \)g of DNase I treated DNAs were digested with HindIII and EcoRI, and the resulting DNA fragments were separated by 1.6% agarose. Southern blots were prepared on GeneScreen Plus membrane (NEN Life Science Products). Hybridization was performed using Quickhyb solution (Stratagene). The probe used was a SalI-BamHI restriction of the cGATA-2 gene, labeled by random priming. For the mapping of HS2, 10 \( \mu \)g of DNase I treated DNA were digested with BglII and SspI and then subjected to electrophoresis through 1.2% agarose. Southern blots were stained with Polybrene and hybridized as above, using in this case a BglII-XbaI fragment as a probe.

Cell Lines—HD24 cells are chicken multipotent erythroid-myeloid cells derived from the E26 virus. They cannot differentiate as efficiently as primary transfarments and express some markers for early erythroid progenitors. They were grown in blastodermic media (35). 6C2 and HD37 are CFU-E stage erythroid precursor cells. The HD37 line was generated by infection of a two day embryonic blastodermic with a mutant of E26 (35), and the HD24 line was obtained by transformation of bone marrow with wild type avian erythroblastosis virus. HD37 cells were grown in blastodermic media (35). 6C2 cells were grown in a minimum Eagle's medium supplemented with 10% fetal bovine serum, 2% chicken serum, 1 ml Hepes, 50 \( \mu \)M \( \beta \)-mercaptoethanol, and a standard complement of antibiotics. DT40 cells, purchased through American Type Culture Collection (Manassas, VA), were grown in Dulbecco's modified Eagle's medium supplemented with 50 \( \mu \)M \( \beta \)-mercaptoethanol, 2 ml glutamine, 10% fetal bovine serum, 5% chicken serum, 10% trypto phosphate broth, and antibiotics. All of the cells were maintained at 37 °C in 5% CO2.

Transfections—Approximately 3.2 \( \times \)10\(^6\) cells per sample were washed twice in phosphate-buffered saline and resuspended in 2.1 ml of Opti-MEM I (Life Technologies, Inc.). 300 \( \mu \)l of Opti-MEM I containing 1000 \( \mu \)g of LipofectAMINE (Life Technologies, Inc.) and 20 \( \mu \)g of plasmid DNA were added to the cells. 1 \( \mu \)g of RSV-CAT and 3 \( \mu \)g of test plasmid (or an equivalent copy number) containing the luciferase reporter gene were used in each transfection. Cells were incubated at 37 °C for 5 h in the presence of the transfection mix, returned to normal media, and incubated for 48 h. For assays, cells were harvested and washed twice in phosphate-buffered saline and resuspended in 150 \( \mu \)l of reporter lysis buffer (Promega). The supernatants were assayed for luciferase activity using the Promega luciferase assay system according to the manufacturer's instructions, and for chloramphenicol acetyltransferase (CAT) activity, using a liquid scintillation method (50). The luciferase activity was normalized to the CAT activity detected in each samples. The values presented are the mean of at least three independent experiments performed each time in duplicate, using different preparations of plasmid DNA.

Construction of GATA-2 Gene Derivatives—A 5.3-kb XhoI-Sall insert derived from the genomic phase \( \alpha \)G2.5 DNA fragment and containing the 5'-end of exon 2, the entire exon 1b, and the sequence upstream of it was subcloned into the unique XhoI site of the pBluescript SK+ plasmid. This fragment was then deleted using Bal-31 nucleases (New England Biolabs) from a unique HindIII site located in intron 1b. After digestion with Mung bean nuclease (New England Biolabs) diga- nization of an adapter to recreate a HindIII restriction site, plasmids were recircularized and sequenced. One resulting cGATA-2 fragment retaining the first 118 bp of exon 1b and the intact 4.9 kb of 5'-flanking sequence was selected. It was inserted between the XhoI and HindIII sites of the pG3L-basic (Promega) to give the –9400 LUC construct. The EcoRV, BamHI, and Smal restriction sites contained in the fragment were also used to generate respectively the constructs –1900 LUC,
2580 LUC, and 2160 LUC, using Smal plus BglII plus HindIII sites in pGL3-basic for the cloning. The 21900/2580 fragment was obtained by an EcoRV-BamHI digestion of the parental fragment, which was then cloned in the pGL3-basic in Smal-BglII.

The mCCAAT LUC, m9 LUC, and mCCAAT-9 LUC mutants were obtained by replacing the Smal-BssHII fragment (position 2160 to 277 in the proximal promoter (see Fig. 5A)) of the 2160 LUC construct, by a double-strand oligonucleotide with the correct restriction sites at its extremities and the sequence of interest mutated. The same mutations of the CCAAT box and the conserved region as the one used in electrophoresis mobility shift assay were introduced in these oligonucleotides.

Protein Extracts—Nuclear extracts were prepared as described (48). Leupeptin (0.5 µg/ml), pepstatin A (0.7 µg/ml), and phenylmethylsulfonyl fluoride (0.1 µM) were added in all the buffers. The integrity of the extracts was tested by mobility shift assays with oligonucleotides containing an Sp1 or a GATA-1 binding site (data not shown).

Gel Retardation Assay—All of the oligonucleotides used in gel retardation assay were synthesized on an Applied Biosystems Synthesizer and gel purified. Single-strand oligonucleotides were then labeled with γ-32P]ATP and then annealed to their complement present in the cDNA. Numbers above the boundaries of introns 2–5 refer to the codons in the cGATA-2 protein. The alternatively spliced variant (GATA-2') results in a protein with 11 extra amino acids located immediately upstream of the N-terminal finger. Thus, the glycine in position 278 in GATA-2 is found in position 289 in GATA-2' (both shown in boldface). The boundary conservation is extended to all the coding exons for the four GATA-2 genes that have been cloned (Refs. 17, 24, 27, and 33 and this paper).

Alternative splice

S C F H L E G289 R

TCC TGT TTC CAT TTA GAA GGC AGA

FIG. 1. Structure and organization of the chicken GATA-2 gene. A, two λ phage clones (JG2.611 and JG2.5) and one plasmid clone (pG2.2) were analyzed. Together, the clones JG2.611 and JG2.5 contained all of the sequences present in the chicken GATA-2 cDNA (11). The boxes represent the six exons identified (Ex1 to Ex6), based on the cDNA sequence. Ex4 and Ex5 (white boxes) encode the zinc fingers required for specific DNA binding. The symbols used for the restriction enzyme sites are as follows: B, BamHI; E, EcoRI; R, EcoRV; S, SalI; Sa, Sau3AI; X, XhoI. The EcoRV restriction sites have not been mapped in the JG2.611 clone, nor downstream of the SalI site in JG2.5. B, sequences of the exon/intron boundaries. The numbers above the intron 1 boundaries refer to the nucleotide positions relative to the first nucleotide identified in the published cDNA. Numbers above the boundaries of introns 2–5 refer to the codons in the cGATA-2 protein. The alternatively spliced variant (GATA-2') results in a protein with 11 extra amino acids located immediately upstream of the N-terminal finger. Thus, the glycine in position 278 in GATA-2 is found in position 289 in GATA-2' (both shown in boldface). The boundary conservation is extended to all the coding exons for the four GATA-2 genes that have been cloned (Refs. 17, 24, 27, and 33 and this paper).
Alternative Exon Usage in Chicken GATA-2

Structure of the Chicken GATA-2 Gene—We isolated two overlapping genomic clones (see under “Experimental Procedures” and Fig. 1A) together spanning the entire cGATA-2 cDNA previously isolated by Yamamoto et al. (11). A third clone (pG2.2) suspected to contain some of the regulatory sequences of the gene was also analyzed. The organization of the gene appeared at first very similar to that of the human and the Xenopus GATA-2 genes (17, 24) and therefore to that of the GATA-1 and GATA-3 genes cloned from different species (13, 29–32). As shown in Fig. 1A, the cGATA-2 gene is composed of six exons distributed over 16 kb of DNA. All exon-intron boundaries conform to the GT-AG rule (Fig. 1B). The first exon is noncoding, and the methionine codon for translational start is found in exon 2. The two highly conserved zinc finger domains are encoded separately in exons 4 and 5, the intron/exon boundaries of which are highly conserved among all characterized family members (13, 17, 24, 29–32). In the case of the chicken gene, exons 1 and 2 are separated by a relatively large intron (6.5 kb) compared with what is found in the human (0.8 kb) (17).

RESULTS

Exon 1 Is Differentially Transcribed—We next identified the promoter of the cGATA-2 gene in chicken hematopoietic progenitor cells. We first attempted to delineate the 5′-end of the cGATA-2 mRNA. Comparison of the length of the published cDNA with the mRNA detected by Northern blot (11) indicated that about 1.5 kb of transcribed sequences remained to be identified. These yet unknown sequences could be located 5′ and/or 3′ of the mRNA, as no consensus polyadenylation signal is present in the cDNA. We used an RNase protection assay to analyze this message. Two radiolabeled probes were generated: one of 395 bp containing part of exon 6 and one of 180 bp containing the entire known sequence of the first exon (Fig. 1B).

A GATA-2 cDNA variant, GATA-2′, containing an additional in-frame 33 nucleotides immediately 5′ to the N-terminal zinc finger motif, has been described in both chicken and Xenopus (14, 21, 24, 34). This alternatively spliced form results in a protein with 11 extra amino acids, which are highly conserved between the two species. The additional sequence within GATA-2′ represents a 5′ extension of exon 4 that encodes the N-terminal zinc finger (Fig. 1B). Thus the differential use of two splice acceptor sites upstream of exon 4 leads to the expression of GATA-2 and GATA-2′.

Identification of an alternative first exon in the cGATA-2 gene—An RNase protection experiment was performed using a 359-bp riboprobe containing 329 bp of the cGATA-2 gene and 30 bp of plasmid DNA. The riboprobe contains the sequence starting immediately 3′ of intron 1b and spans the 329 bp upstream. The same RNAs as described previously were used in this experiment. A major transcript giving rise to a protected fragment of around 160 bp was detected in the samples; we called it exon 1b. Larger protected fragments were also detected.

As shown in Fig. 2, the exon 6 probe protected a fragment of
Alternative Exon Usage in Chicken GATA-2

Identification of an Alternative First Exon for the cGATA-2 Gene—We first took advantage of a 5'-RACE clone (provided by Dr. Todd Evans) that had been made using poly(A)⁺ RNA isolated from primitive red blood cells and a cGATA-2-specific oligonucleotide hybridizing in exon 2. This clone contained not only the entire expected sequence of exon 2 but also a previously unidentified sequence that matched a portion of our genomic sequence (in χQ2.5). Although this 5'-RACE clone did not contain the entire message sequence, it allowed us to identify the 3'-end of a new exon, which we called exon 1b, lying 403 bp upstream of exon 2 (Fig. 3; see also Fig. 5A). The first exon present in the cDNA cloned by Yamamoto et al. (11) was therefore renamed exon 1a. The 403 bp interval represents intron 1b. This newly identified exon-intron boundary conforms to the GT-AG rule. Exon 1b, like exon 1a, is noncoding.

In order to locate the 5'-end of this new exon, as well as to analyze the tissue distribution and the abundance of the exon 1b-containing mRNA, we used a 329-bp probe, including the entire 3'-end of exon 1b and the sequence located upstream, and the same RNAs as described previously, in an RNase protection assay (Fig. 3). Several probe fragments were protected with all the RNA tested, with the exception, as expected, of the DT40 RNA. A major transcript giving rise to a fragment of about 160 bp was observed in erythroid precursor cells.

To our surprise, the results obtained with the upstream exon 1 probe were clearly different from those obtained with the exon 6 probe. Indeed, a protected fragment of a size consistent with presence of exon 1 sequences was detected only in 6C2, HD37, primitive red blood cells, and 10-day brain, and then at much lower abundance than found for exon 6 (see legend of Fig. 2 and under “Experimental Procedures”). No exon 1-containing mRNA whatever was detected in HD24 and 10-day definitive red blood cells, although these cells clearly express the cGATA-2 gene.

These data show that the cGATA-2 mRNA corresponding to the cloned cDNA is not the major form transcribed in the cells and tissues expressing GATA-2 that we tested. Most importantly, transcripts containing this exon are specifically restricted to only some of these cells and tissues and are not well represented in the total GATA-2 transcript population in erythroid cells. This signaled the existence of cGATA-2 mRNA with an alternative first exon and, thus, the existence of more than one promoter driving the transcription of the chicken gene. We focused our attention on identifying this alternative first exon.

The abundance of the fragment was similar to that detected with the exon 6 probe (Fig. 2, see under “Experimental Procedures”), suggesting that most of the cGATA-2 mRNA tran-
Alternative Exon Usage in Chicken GATA-2

In order to analyze the transcriptional activity of the sequence located upstream of the exon 1b, we constructed chimeric reporter plasmids containing this sequence fused to the luciferase (LUC) reporter gene (Fig. 6) and transfected them into 6C2 and HD24 cells, both of which transcribe exon 1b (Fig. 3). We carried out successive 5′-RACE to determine the position of the 3′-polyadenylation signal of cGATA-2 in this cell (data not shown). We found that it is located about 1400 bp downstream of the end of the published cDNA clone. Taking into consideration the small difference between exon 1a and exon 1b, we account for about 1240 bp of the unexplained sequence, consistent within our limits of error with the observed discrepancy of 1.5 kb.

Small amounts of larger fragments corresponding to the fully protected probe were also observed, showing that minor amounts of transcription can be initiated upstream of the major initiation site corresponding to the 160-bp exon 1b. However, these mRNAs were low in abundance in progenitor cells and could not contribute appreciably to the population that contained the coding region represented by exon 6. For these reasons, we focused our attention on identifying the promoter sequence that directs the transcription of the most abundant cGATA-2 mRNA in the hematopoietic progenitor cells, the exon 1b-containing mRNA.

**DNase I Hypersensitive Site Mapping**—The promoters and enhancers of active genes are typically associated with nuclear HS sites (36, 37). To help locate such elements near exon 1b, we examined the DNase I sensitivity of chromatin of the HD24 cell line, which expresses abundant mRNA containing exon 1b but not exon 1a. We used as a control the DT40 cell line, which does not express any cGATA-2 mRNA. We obtained similar results with a variety of restriction enzyme and probes. Two representative sets of HS mapping data obtained using the same preparation of DNase I-treated nuclei are presented in Fig. 4. A prominent HS site, HS1 (Fig. 4A), was found in HD24 cells, located at approximately the transcription start site of exon 1b. Because of its position, it seemed likely that HS1 might mark the promoter required for the transcription of this exon, indicative of its specificity for chromatin structure. It was also detected in 10-day red blood cells and in DT40 cells, although reproducibly weaker, but not in naked DNA (data not shown). Another HS site, HS2 (Fig. 4B), which actually appears as a doublet, was observed in HD24 but also clearly in the DT40 cell line, in 10-day red blood cells and brain but not in naked DNA (data not shown). It is located 3 kb upstream of the 5′-limit of exon 1a. Interestingly, the sequence downstream of HS2 and extending until at least exon 2 is very CG rich and the ratio CpG/GpC is close to 1, as expected in the case of a CpG island (38), whereas upstream of HS2, the CpG content drops to the expected ratio for a non-CpG island sequence. Using 6C2 cells in which some exon 1a is transcribed, we were not able to detect any HS site closer than HS2 to the 5′-limit of exon 1a.

**DNA Sequence of the cGATA-2 Promoter Region Upstream of Exon 1b**—The sequence of the putative promoter located upstream of exon 1b (GenBank accession no. AF038592), which is clearly hypersensitive in cells expressing this exon (Fig. 4A) is presented in Fig. 5A. As in the other GATA genes described (13, 24, 28–32), no canonical TATA box or consensus to the transcription initiator element (39) is present. No homology to the binding site of the putative housekeeping initiator protein 1 (HIP1) described in the case of the GATA-3 genes (40) is apparent, nor is there the downstream promoter element motif observed in some TATA-less promoters (41). The putative promoter region lacks any strict consensus GATA site, and none has been detected in the 1800 bp upstream of exon 1b. However, a CCAAT box in inverted position is located in the putative promoter region lacks any strict consensus GATA site, and none has been detected in the 1800 bp upstream of exon 1b. HD24

HD24
ments important for the transcription of exon 1b in both cell types. In this respect, the cGATA-2 proximal promoter is quite similar to that of mouse, with strong stimulatory activity localized to a region between 100 and 200 bp upstream of the transcription start site.

The profiles of activity as a function of the size of the 5' deletion were somewhat different in HD24 compared with 6C2 cells. Deletion between -4900 and -1900 resulted in an increase in activity in each type of cell, suggesting that repressive elements might be present in this region. However, this deletion had a much greater effect in HD24 than in 6C2, and in fact, all of the upstream region 5' of -160 had a strong inhibitory effect in HD24. We note that a fragment containing the 5' sequence -1900 to -580, when linked to the LUC gene, displayed considerable promoter activity in HD24, suggesting that the proximal elements necessary for high level expression in 6C2 and HD24 cells, we further explored its role in the transcription of the cGATA-2 gene. We noticed that not only the CCAAT element itself was conserved; in the region 3' of the element, several additional nucleotides are perfectly conserved among the chicken, Xenopus, human, and mouse genes (Fig. 5B).

We first explored the central CCAAT motif. A double-stranded oligonucleotide spanning the sequence -115 to -89, which contains the CCAAT box, and the entire conserved region was used in electrophoretic mobility shift assays with protein extracts prepared from HD24, 6C2, DT40, and QT6 cells (Fig. 7A). QT6 is a quail fibroblast line (42) in which we were not able to detect GATA-2 message by Northern blot analysis. Each extract gave rise to a single complex, and the bands showed identical mobilities. This complex is competed by a 100-fold molar excess of the unlabeled probe but not by the downstream conserved sequence. When labeled, this oligonucleotide, which contains only 4 bp upstream of the CCAAT box, also does not compete the complex, even though it contains the adjacent conserved sequence (data not shown), suggesting that the adjacent conserved sequence is not itself recognized by any specific proteins.

Although the central CCAAT motif is essential, it is also not sufficient. Mutation of the nine nucleotides between -97 and -89 in the context of the probe used to detect the CCAAT binding factor (oligonucleotide m9) reduced the affinity of binding of the specific complex observed. This is clearly shown by the range of competition presented in Fig. 7C: the mutated
oligonucleotide m9 does not compete as well as the nonlabeled wild type sequence. This suggests that the affinity of the CCAAT factor that we detected for its target sequence is affected by the surrounding sequence, a conclusion also suggested by results obtained in direct binding assays comparing the wild type and m9 probes (Fig. 7D). Thus, the extended sequence homology that we have identified among GATA-2 CCAAT motifs appears to be important for binding at that site.
We have tested the function of the CCAAT element by introducing mutations at the central residue of the CCAAT motif (CCGAT, the mutant mCCAAT LUC). We also measured the effect of the mutation of the nine conserved nucleotides we had tested in vitro (m9 LUC) in the same context, as well as the simultaneous mutation of the CCAAT box and the conserved nucleotides (mCCAAT + m9 LUC). Transient transfection of the mCCAAT LUC construct into both HD24 and 6C2 cells revealed a 3-fold decrease in activity relative to the wild type promoter (data not shown). Mutations in the 3'-flanking region of the site had a similar effect, consistent with our gel shift data, and showing the importance of the extended motif.

**DISCUSSION**

We have found that the chicken GATA-2 gene, like other members of the GATA family, has a complex series of regulatory elements controlling alternate first exons and can thus give rise to distinct transcripts in different cell types. The cGATA-2 gene has two distinct transcription start sites separated by about 6.2 kb, and we observed two transcripts carrying different untranslated first exons. Of those GATA-2 genes that have been cloned, the only other so far discovered to have this arrangement is that of the mouse, quite recently reported. However, although the mouse also uses two different first exons, they are employed in quite different ways in the two organisms. In the chicken, the predominant transcript in all cells and tissues, and most notably in erythroid cells, derives from the proximal promoter, whereas in the mouse, hematopoietic-specific expression is delegated to the distal promoter. The two regulatory systems are thus remarkably distinct.

The alternative exon structure of cGATA-2 might easily have escaped notice, because the cDNA earlier isolated from a total 10-day embryonic cDNA library (11) corresponded only to transcripts from the distal promoter, thus leading to the assumption that there was a single transcript in all cells. We were able to determine by examination of our genomic clone that this transcript initiated at a site about 6.7 kb upstream from the first coding exon. However, when we examined the RNA population in a variety of cells and tissues it soon became apparent that this was not the principal transcript in any of them, and was not detected at all in HD24 and 10-day definitive red blood cells, indicating that its expression is subject to a developmental-specific, cell-specific, or inducible regulation. Instead we found that the GATA-2 gene can be transcribed from an alternative first exon (exon 1b), which is located 403 bp upstream of exon 2 in the genomic DNA. The resulting mRNA is the most abundant form detected in all of the cells and tissues we tested that express GATA-2. It is possible that the under-representation of this mRNA in the cDNA sample analyzed by Yamamoto et al. (11) arises from difficulty in reverse transcription of GC rich sequences containing exon 1b, which we have observed.

The existence of alternative first exons has not been reported for the Xenopus (24) or the human (17, 28) GATA-2 gene. It has been described in the mouse, and as noted above, in that organism the upstream promoter appears to be utilized primarily in hematopoietic cells, whereas the promoter proximal to the gene is used in all tissues. The existence of alternative first exons has also been observed for other members of the GATA family (30, 44, 45). For example, the mouse GATA-1 gene is transcribed from alternative first exons, at least in mouse erythroleukemia and MC/9 mast cells (30). In the mouse testis, the gene is expressed from another first exon, which is different from the cell-specific ones (44). The chicken GATA-5 gene has also been reported to be transcribed from two alternative first exons (45).

The more proximal first exon of the cGATA-2 gene, exon 1b, appears analogous to the first exon described for the Xenopus and the human GATA-2 genes, suggesting that an alternative first exon located further upstream may exist for these two genes as well. Indeed, the reported first introns of the Xenopus (250 bp) (24) or human (800 bp) (17) gene are closer in size to the 403-bp chicken intron located between exon 1b and exon 2 than they are to the chicken intron between exon 1a and exon 2 (6.5 kb). Furthermore, an inverted CCAAT box positioned upstream of a highly conserved sequence is located in the minimal promoter of the Xenopus gene and just upstream of the chicken exon 1b. This motif is also present in the human gene but its function has not been yet analyzed. In Xenopus, the CCAAT box is required for the activation of the zygotic gene at the onset of gastrulation during embryogenesis. In this study, we have shown that this element is also necessary for the full activity of the chicken minimal promoter driving the transcription of exon 1b-containing mRNA in hematopoietic progenitor cells. This minimal promoter is contained between the sequences –160 and –40 and maps to a strong hypersensitive site in vivo. Thus, promoter organization appears to be conserved across species, although the particular developmental tasks assigned to each promoter vary.

Our identification of the activation properties of the CCAAT box show that the conservation between species is not restricted to the overall organization of the GATA-2 gene but extends also to some aspects of the regulatory mechanism. The CCAAT box appears to function regardless of the point in development or the cell type in which GATA-2 is expressed. It is possible that the particular CCAAT factor employed may vary. A multiplicity of CCAAT box binding activities have been identified, some of which are tissue-restricted, whereas others are expressed ubiquitously (47). However, our electrophoretic mobility shift assay experiments do not reveal the presence in different cell types of other CCAAT factors binding to the CCAAT motif of the GATA-2 gene. We have shown that the affinity of binding of this protein to its target depends on the highly conserved nucleotides located 3' of the CCAAT box, suggesting that there may be additional factors that bind to the extended site. Our data show that if such a factor exists, it binds only when the CCAAT factor already occupies its own adjacent site.

The CCAAT factor, which binds in vitro to the CCAAT motif located in the cGATA-2 gene, is also detected in DT40, a chicken lymphoblastoid cell line in which the transcription of the gene is, as expected, not detected. This leads us to conclude that the CCAAT site is not sufficient for the restricted expression of the gene. This is consistent with results seen with mouse GATA-2, in which the region containing the CCAAT motif induces severalfold stimulation of reporter expression, even in cells that do not express GATA-2 (27). Previous work performed on the Xenopus embryo (18) showed that during gastrulation, the initial expression of the zygotic gene occurs in a broad domain, throughout the ventral and lateral regions of the embryo. In embryo explants, accumulation of xGATA-2 mRNA can be inhibited by co-culturing with either activin, dorsal marginal zones, or the dorsalizing and neural inducer noggin, suggesting that the localization of xGATA-2 expression to the ventral region is a consequence of negative control during dorsalization and neural induction (18).

These results suggest that in all GATA-2 genes, transcription of the exon 1b-containing mRNA may be controlled by ubiquitous transcription factors, including the CCAAT factor, and that the absence of transcription of the gene in some of the cells of the hematopoietic system, such as the lymphocytes, or in other tissues, is the consequence of negative control. We suggest that tissue-specific suppression of GATA-2 expression might therefore arise from inhibitory signals elsewhere in the
neighborhood of the gene, but some distance away. Our transient transfection experiments are consistent with that point of view. They indicate that sequences located upstream of the minimal promoter have a strong inhibitory effect in HD24, a chicken multipotent erythroid-myeloid cell line, but not in 6C2 or HD37, two distinct CFU-E stage precursor cells. This is entirely consistent with the relative endogenous expression of GATA-2 in these cells, suggesting that the upstream elements confer developmentally specific regulation.

Such a model also may explain the plasticity of the GATA-2 regulatory pattern, which allows similar promoters to be used in quite different ways in different organisms. The isolated primordial regulatory pattern, which allows similar promoters to be used in quite different ways in different organisms.
Alternate Promoters and Developmental Modulation of Expression of the Chicken
GATA-2 Gene in Hematopoietic Progenitor Cells
Pascale Nony, Robert Hannon, Hannah Gould and Gary Felsenfeld

J. Biol. Chem. 1998, 273:32910-32919.
doi: 10.1074/jbc.273.49.32910

Access the most updated version of this article at http://www.jbc.org/content/273/49/32910

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 49 references, 28 of which can be accessed free at
http://www.jbc.org/content/273/49/32910.full.html#ref-list-1