Interaction of the erythroid transcription factor cGATA-1 with a critical auto-regulatory element

Gary Schwartzbauer, Kim Schlesinger and Todd Evans*

Department of Biological Sciences, Langley Hall, Room 518, University of Pittsburgh, Pittsburgh, PA 15260, USA

Received June 23, 1992; Revised and Accepted August 6, 1992

ABSTRACT

We have performed a mutational analysis of the promoter for the chicken erythroid-specific GATA-1 transcription factor, and have investigated in detail the interaction of the factor with an upstream auto-regulatory element (ARE). We find that a single proximal GATA binding site of the ARE is required for promoter activity in primary erythroid cells; however, this minimal promoter is inappropriately active in fibroblasts. At least two molecules of GATA-1 can interact with the ARE, and sequences outside of the consensus site appear critical for the transcriptional activity of the bound protein. Finally, we provide evidence for complex protein/DNA interactions at the ARE, including the ability of GATA-1 to bend DNA.

INTRODUCTION

The GATA-1 transcription factor is a DNA-binding protein thought to be involved in the regulation of many if not all erythroid-specific genes ((1,2), for review). The cDNA and genomic clones encoding the factor have been isolated from several vertebrate species (3–6) allowing its function and regulation to be analyzed at the molecular level. In the chick and mouse systems, certain erythroid genes have been demonstrated by transient transfection assays to be direct targets of GATA-1 trans-activation (7–9), while disruption of the normal mouse GATA-1 locus results in embryonic blastomeres incapable of generating cells of erythroid lineage (10). These results have implicated the GATA-1 protein as a critical component of the molecular mechanism which initiates and maintains the erythroid program during development and cellular differentiation.

Recent studies of the GATA-1 promoter from chick, mouse, and human have revealed that one of the targets for GATA-1 activation is likely to be the GATA-1 gene itself (11–13). Such an auto-regulatory role is consistent with the hypothesized function of this transcription factor in establishing the fate of a pre-determined erythroblast precursor cell by a self-perpetrating regulatory loop. Auto-regulation of so called 'switch' proteins has now been documented in several other systems (14–17). Critical components of the mammalian GATA-1 promoter include

two CACCC motifs 200 bp upstream of the transcription start sites, in addition to a 'complex' GATA binding site around –680 (containing two adjacent GATA motifs with opposite orientations).

The chick promoter has certain features in common with the mammalian gene, including GATA motifs and the lack of a TATA box. However, the structure of the cGATA-1 gene and upstream region is quite unusual and it may not function by an entirely analogous mechanism. Multiple GATA sites are present, including a similar complex site containing three overlapping GATA motifs. Compared to the mammalian genes, these sites are relatively close (around –150) to a cluster of predominant start sites. An additional distal GATA site is located further upstream around –425. The region including these GATA sites was shown to function as an extremely efficient promoter when used to direct expression of a reporter gene in primary erythroid cells (12).

In addition to GATA motifs, binding sites for other proteins from erythroid or non-erythroid cells were mapped to presumptive upstream regulatory elements. For example, the promoter contains multiple stretches of G residues known to bind the chicken protein BGP1, which also binds upstream of the chick β -globin gene (18). However, the functional significance of any cis-elements in regulating the cGATA-1 promoter has not been tested. We provide evidence that the upstream GATA binding sites are crucial for the function of this gene. Because mutations in and around the most proximal motif of the –150 complex site abolish promoter activity, we have designated this region an auto-regulatory element (ARE), and investigate the sequence requirements for its function.

GATA-1 is a member of a multi-gene family, whose members are conserved in the DNA-binding domain which recognizes the GATA motif (19–22). The binding sites are sometimes found in over-lapping arrangements that appear to confer a higher affinity (11,23,24). In addition, binding sites may be present in proximal or distal locations, and in forward or reverse orientations (23). Taken together, there is the potential for considerable regulation to be dictated through the particular sequence and or position of the GATA binding site. With this in mind, we have undertaken a detailed analysis of the protein/DNA interactions that occur at the cGATA-1 ARE.

* To whom correspondence should be addressed

MATERIALS AND METHODS

Plasmid Constructions

All plasmids used in this study are derived from pGATP1 (described in ref. (12)), which consists of a 642 bp NcoI genomic cGATA-1 fragment inserted upstream of a reporter CAT gene (pCATBASIC vector, Promega). We refer to this as the full-length or wild-type promoter and have designated it p-422 for consistency with other plasmids used in this study. This full-length promoter contains a single HindIII site located at the 5' end. We made use of convenient restriction enzyme sites to generate the deletion mutations. Reporters shown in Fig. 1 were made by first employing the following enzymes: p-415, TaqI; p-386, SmaI; p-349, FokI; p-337, AluI; p-147, StuI; p-124, BstNI. In some cases partial digestions were necessary; the end-point of the deletion (corresponding to the name of the reporter) was blunted using mung bean nuclease or Klenow enzyme as required. Plasmids were then digested with PstI, which cuts once in the 3' end of the genomic insert. The progressively deleted fragments were gel-purified before sub-cloning back into HindIII (blunted)/PstI digested p-442 vector. The pD147/124 reporter (deleting the single proximal GATA site) was constructed by purifying the HindIII to StuI fragment of p-442, and inserting it upstream of the BstNI site of p-124 using a HindIII linker.

To generate the reporters illustrated in Fig. 3, the proximal promoter region from the -124 BstNI site (blunted using mung bean nuclease to remove 1 bp) to a downstream XbaI site was first transferred into XbaI/SmaI digested pBluescript IISK (Stratagene) vector. The insert was cut out using HindIII and XbaI and ligated back into p-442/HindIII/XbaI. The result of this manipulation (designated pGPB) was to change the BstNI site into a SmaI site, and generate polylinker sequences upstream of -124. The pARE reporter and all derivatives were generated by ligating synthetic oligomers into HindIII/SmaI (XmaI) digested pGPB. The -136 and -138 mutations were discovered by sequencing random clones. For the +us reporter, upstream sequences from HindIII to StuI (-442 to -147) were inserted into HindIII digested plasmid using a HindIII linker. For the DNA bending studies, the oligomer used to generate pARE was inserted into the XbaI site of pBEND2 (25) using synthetic linkers. Each of the plasmid constructs was verified by sequencing.

Cell culture and DNA transfections

Erythroid cells were harvested from 10 day chick embryos and transfected by hypotonic shock as previously described (7). Each 0.5 ml reaction contained 1 µg of reporter, 0.1 µg β-galactosidase internal control plasmid (regulated by the Rous sarcoma virus LTR), and 100 µg DEAE-dextran. Cells were harvested after incubation for 40–48 hr at 37°C. Quail fibroblast (QT6) cells were transfected, cell lysates were harvested, and CAT assays were performed exactly as described (7). The CAT assays were quantified using the AMBIS Radioanalytic Imaging System; the β-galactosidase activities were measured by fluorescent spectrophotometry.

DNA-binding assays

Nuclear extracts were prepared from 10 day embryonic erythroid cells or QT6 fibroblasts and used in gel mobility-shift experiments as previously described (7). For the off-rate analysis, a similar binding reaction was scaled up seven-fold and incubated on ice for 30 min. After addition of a vast excess of unlabeled competitor

DNA (-147* oligomer) aliquots were loaded onto a continuously running mobility-shift gel. The half-life of the DNA/protein complex was determined as described (26). Methylation interference assays were performed by scaling up a standard mobility-shift reaction five-fold, using partially methylated end-labeled probe (27). Bound and free probe was electro-eluted from preparative gels onto NA45 membrane (Sleicher and Schuell), rinsed off using 1 M NaCl, ethanol precipitated, and cleaved by piperidine treatment (27) prior to denaturing gel electrophoresis. For DNA-bending assays, after appropriate restriction digestions the inserts were end-labeled and gel-purified (27). Standard conditions were used for the mobility-shift assays (7), except that the cross-linking of the 8% gel was reduced to 72:1, reactions included 4 µg of non-specific poly(dIC) DNA, and gels were run at 4°C.

RESULTS

Identification of a minimal cGATA-1 promoter active in erythroid cells

By performing DNaseI footprinting experiments, a number of binding sites for factors known to be present in erythroid extracts (eg. GATA-1, BGP1, and Sp1) were mapped *in vitro*, while several other consensus binding sites (eg. for the proto-oncogenes ets and myb) did not show obvious protections (12). We analyzed the functional significance of these binding and putative binding sites by constructing deletion mutations from the 5' end of this promoter region and assaying the effect on expression of a reporter CAT gene as sites were progressively removed (see Fig. 1).

The full-length cGATA-1 promoter contains 442 bp of upstream flanking sequence relative to the most 5' start site of the first non-coding exon. Each of the plasmids tested in this study maintain downstream sequences including the entire 161 bp first intron, all clustered RNA initiation sites, and at least 124 bp of upstream flanking sequence. To determine promoter activity, plasmids were co-transfected with a control β-galactosidase expressing plasmid into primary embryonic chick erythroid cells. Cell lysates derived from transfected cells were assayed for CAT activity, measured relative to the accumulation of β-gal activity. The activity of the mutated plasmids is given in Fig. 1 (RBC), relative to the activity of the full-length ('wild-type') promoter region containing sequences to -442.

Deletion to -415 removes a single distal GATA-1 binding site and has a significant effect on promoter activity; the -415 plasmid is 27% as active as the full promoter. Interestingly, further deletion all the way to -147 shows little additional effect, although certain deletions do show some fluctuations in activity. For example, deletion to -386, which removes the putative c-ets binding site, results in a two-fold increase in activity, relative to -415. Deletion to -147 removes binding sites for BGP1, the putative c-myb binding site, and 2 overlapping GATA-1 sites and results in a 2-fold drop in activity, relative to -386. While the significance of these minor alterations in activity is unclear, deletion of the most proximal GATA-1 binding site by removing sequences to -124 results in a complete loss of promoter activity.

As will be discussed below, the GATA sites located around -150 represent a complex element consisting of three potential binding sites, two of which overlap on opposite strands (the two deleted in the -147 plasmid). We specifically deleted the most proximal site while maintaining the remaining upstream sequences

Figure 1. Relative activity of cGATA-1 promoter mutations in erythroid or fibroblast cells. The solid line at top represents 442 bp upstream of the most distal major transcription initiation site. The endpoint of each deletion is given on the left of each line representing the extent of promoter present in each reporter. On each line is also shown binding or putative binding sites for known transcription factors which are removed by the next progressive deletion. The bottom line represents a reporter ($\Delta 147/124$) in which the proximal GATA motif was selectively deleted. Activity is reported relative to that obtained using the full-length reporter in primary chick erythroid cells (RBC), avian fibroblasts (QT6), or QT6 cells co-transfected with a cGATA-1 expression vector (GATA-1/QT6). The extent of trans-activation by GATA-1 in QT6 cells is indicated under TA, determined as the ratio of cGATA-1 promoter activity in the presence or absence of co-transfected cGATA-1 expression plasmid. Each number represents the average of at least 4 experiments standardized using a β -galactosidase expression vector as internal control. Standard deviations were in each case about 20% of the indicated value.

in their entirety ($p\Delta 147/124$). This promoter is only 14% active relative to the wild-type plasmid, even though the proximal GATA site is 'replaced' by the two upstream adjacent sites. At this time we can not distinguish between two possible interpretations: 1. upstream 'enhancer' elements (eg. including the GATA site at -440) only function with the natural proximal GATA site, or 2. the enhancer is active, but the promoter only functions well in the context of the natural proximal site. Regardless, this result confirms the crucial function of the most proximal GATA element.

We can make the following conclusion with regard to the results of our transient assay in primary cells: while the host of upstream binding activities may contribute to the precise level of promoter activity, a critical element for maintaining a high level of expression is the single proximal GATA-1 binding site located at -140 to -145.

Negative regulatory elements may be required in non-erythroid cells

Because the cGATA-1 gene is restricted in expression to hematopoietic cells, we next tested whether the deletion mutants could be used to identify elements involved in repressing the activity of the promoter in non-erythroid cells. Therefore, in a similar set of experiments, cGATA-1 promoter reporters were transfected either in the absence or presence of cGATA-1 expression plasmid into QT6 avian fibroblasts.

The results of these experiments are also shown in Fig. 1 (QT6). Compared to the endogenous gene (which is completely

inactive), it appears that the transfected promoter is inappropriately expressed at a high level in the fibroblasts. Only in the case of the full-length reporter do we detect a significant level of trans-activation (10-fold). This might be explained by a cumulative effect of negative elements which keep the background expression low (relative to the deleted reporters), and the presence of multiple GATA sites, including the most distal site at -425. While these results confirm the functional role of the distal GATA site for GATA-1 dependent promoter activity, the deletions do not identify a specific element involved in the negative regulation of the promoter. The high basal expression of all the reporters (relative to the CAT vector) indicates either that positive-acting factors may fortuitously function on the downstream proximal promoter, or that other negative regulatory elements may exist for the cGATA-1 gene outside of the region we are studying.

We have considered one alternative explanation for the modest levels of trans-activation in the QT6 cells: that the expressed GATA-1 may not interact well with the ARE in a non-erythroid environment. We tested *in vitro* the ability of this protein to bind to the critical GATA of the -147 reporter (site 3 of the ARE; see Fig. 4, below) by gel mobility-shift assays using extracts either from erythroid cells containing endogenous GATA-1, or from QT6 cells transfected with the GATA-1 expression vector. In both cases, efficient binding of GATA-1 to the site was detected; as shown in Fig. 2 (a and b), the off-rate of the protein/DNA complex was determined to be equivalent in the two cases.

Figure 2. A.) Determination of the off-rate for the critical GATA promoter motif with GATA-1. Gel mobility-shift assays were used to determine the stability of the protein/DNA complex in the presence of a vast excess of cold probe, added after pre-incubation of a complete reaction. Samples included erythroid cell nuclear extract (RBC), or nuclear extract from QT6 cells transfected with a cGATA-1 expression vector (QT6). Unbound labeled probe (a 25 bp duplex containing the single proximal GATA binding site from -127 to -147) is indicated by the P. The arrowhead shows the position of the GATA-1/DNA complex (the minor complex which migrates faster is due to a protein degradation product). For each set, aliquots were removed and loaded at the following times (left to right in min) after addition of 400-fold molar excess unlabeled probe as competitor: 0, 1, 2, 4, 8, 16, 38.5. B.) The graph plots the log of the per cent bound probe relative to probe bound at time 0 (at the time of addition of competitor), as a function of time. Open squares are using erythroid cell extract, closed squares are using QT6 lysate. The half-life of the complexes was determined by the slope of the curves (26) to be approximately 25 min in both cases. C.) The graph plots a similar off-rate analysis performed using erythroid nuclear extract and either an oligomer containing the wild-type site 3 (open circles) or an oligomer containing the -138 point mutation 3' of site 3 (filled circles). The mutation decreases the half-life of the complex in this experiment from 37 to 27 min. Time points after addition of wild-type site 3 competitor were 0, 2, 4, 16, and 32 min. D.) Off-rates were measured as above using erythroid extracts and the following probes: squares, wild-type ARE in a single-bound complex; triangles, wild-type ARE in a double-bound complex; circles, ARE-2 (mutated at site 2) in a single-bound complex; diamonds, ARE-2 in a double-bound complex. Time points after addition of cold competitor (wild-type ARE) were: 0, 1, 2, 4, 32, and 45 min. Note that the mutation only effects the single-bound complex (reducing the half-life in this experiment from 27 to 19 min).

The precise sequence context around the proximal GATA is critical for promoter function

In order to more conveniently assay the function of putative binding sites in modulating the activity of the minimal cGATA-1 promoter, we engineered a new series of plasmids (pARE and derivatives). Using synthetic oligomers, the sequence of the ARE from -162 to -134 was cloned into the -124 plasmid, such that restriction enzymes could now be used to insert sequences upstream (HindIII) or downstream (SmaI) of the ARE. In constructing this reporter, a BstnI site at -124 was altered to form the SmaI site, and 9 bp between the new SmaI site and the ARE cassette were deleted. The sequence of this region for the wild-type and mutated plasmids is shown in Fig. 3.

We expected that the pARE plasmid should be at least as active as the closely related -147 reporter, and perhaps more so, owing to the presence of an intact ARE containing all three GATA sites. To our surprise, the pARE promoter was found to be only 3% active in transfected erythroid cells, relative to the wild-type promoter (14% of the -147 plasmid). The only differences in pARE compared to plasmid -147 are the additional GATA sites of the ARE and the 9 bp deletion of immediate downstream flanking sequence. We focused on these regions to determine the cis-elements responsible for the difference in activity (see Fig. 3).

Insertion of a 9 bp linker sequence downstream of the SmaI site (pARE/link) restores the proper length of the promoter and

	-160	-150	-140	-130	-120	Relative Activity
Wild-type	CCC	GCTCGCTATCAGATAAGGCCCTTATCAGTGCCTACAGCCCTGGGGC				100%
					BstNI	
-147.....						22%
-124.....						1%
pARE.....				*****		3%
pARE/link.....				GGGGCCGCT		4%
pARE/-1.....		CAG		*****		6%
pARE/-2.....		CTG		*****		4%
pARE/-3.....			CAG	*****		4%
p-147*.....					*	19%
p-147*+us.....					*	90%
pMt-136.....					*	11%
pMt-138.....				*	*	6%

Figure 3. Mutational analysis of the ARE. The sequence of the wild-type promoter between -120 and -168 is shown on the top line. Also indicated on this sequence are the three GATA motifs (arrows denote orientation) and the BstNI site used for sub-cloning purposes. Changes in nucleotide sequence and relative promoter activity are shown for the reporters listed below; dotted lines indicate plasmid vector sequence, solid lines indicate identity to the wild-type sequence and asterisks indicate deleted base pairs. The filled rectangle (+us reporter) represents the addition of sequences between -148 and -442. Therefore, the p-147*+us reporter is identical to the wild-type promoter plasmid, except for the deletion of a single T residue of the former BstNI site. Transfections, assays, and standard deviations were as described in Fig. 1. Data for p-147 and p-124 are taken from Fig. 1.

the spacing between the GATA sites and downstream elements, but does not restore activity to the level of the -147 plasmid. Although it seemed unlikely that the additional GATA motifs were inhibiting activity of the promoter, we introduced mutations into individual sites. These also failed to restore full activity (pARE-1, pARE-2, pARE-3). Activity equal to the -147 reporter is only regained when the precise sequence downstream of the proximal GATA site is restored (except for modification of the BstNI site for cloning purposes; p-147*). When upstream elements are included on this plasmid (p-147*+us), nearly full wild-type activity is restored.

A possible problem with interpreting data from such deletion clones may arise due to interference (either positive or negative) by adjacent vector sequences brought close to the point of the deletion. As a control for this type of artifact, we inserted random bacteriophage sequences upstream of plasmids p-147* and pARE/link to remove the vector DNA away from the promoter sequences. In both cases, the activity of the plasmids was not effected (not shown), demonstrating that vector sequences were not responsible for the high (p-147*) or low (pARE/link) activity of these reporters.

The 3' sequence adjacent to the proximal GATA site is not bound by nuclear extracts *in vitro* (12) and unpublished data). Likewise, we do not find that any of these sequences are critical for GATA-1 binding (see below). Nevertheless, our data strongly implicate the sequences between -125 and -140 as important for promoter function. To confirm this interpretation, we analyzed several point mutations and/or deletions in this region. Indeed, even a single base deletion or alteration displays some negative effect on promoter activity (pMt-136 and pMt-138). We find the deletion at -138 to result in a binding site with a 50% increase in the dissociation rate of GATA-1 with this element, relative to the wild-type sequence (shown in Fig. 2c), which may explain the effect on promoter activity.

Multiple GATA-1 molecules may interact at the ARE

We have undertaken a detailed analysis of the protein-DNA binding interactions at the -150 ARE containing the three GATA

motifs. A representative gel mobility-shift assay using a wild-type ARE oligomer (probe ARE) containing all three potential GATA binding sites and nuclear extract from primary embryonic erythroid cells is shown in Fig. 4a. Also shown are identical assays using oligomers in which individual GATA motifs are mutated (probes ARE-1, ARE-2, and ARE-3; the sequence of the probes are given in Fig. 4c). As a control, we have included an additional assay using a well characterized element from the chicken β -globin enhancer which binds independently two molecules of GATA-1 (probe DH, (23)). Indicated in Fig. 4a is the position that the probes migrate when complexed with a single GATA-1 molecule (filled arrows). As additional protein is included in the binding reaction, some of the complex is driven into a slower mobility species (open arrows) representing probe bound simultaneously to two GATA-1 molecules. Note that this double complex forms readily using any of the probes except ARE-3.

It is apparent from this analysis that the -150 ARE is capable of binding at least two molecules of GATA-1; as in the case of the enhancer probe DH, the binding appears to be additive and not cooperative. Probe ARE-3 readily forms a single complex but is not driven into the double complex (an additional complex with a mobility similar to the double-bound probes does not increase in abundance with added protein). We conclude that the region binds one molecule of GATA-1 at the proximal minimal promoter site 3, and an additional single molecule at the adjacent upstream site containing two overlapping GATA motifs (sites 1 and 2). It is possible that a minor population of the probe may bind additional molecules (see Fig. 5, below), but we do not readily detect it in this assay.

To determine precisely where the GATA-1 binds, and the sequences critical for binding one or two molecules of GATA-1 within the ARE, probes were first partially methylated using DMS. Following separation of the free probe from bound complexes as in Fig. 4a, the DNA was purified from preparative gels, cleaved in the presence of piperidine, and the free or bound fractions compared by denaturing gel electrophoresis as shown in Fig. 4b. A critical binding determinant in the DNA sequence

Figure 4. Interaction of GATA-1 with the ARE. A.) Gel mobility-shift assays demonstrate binding of at least two GATA-1 molecules to the ARE. The migration of unbound probes are marked by a P (the length of the DH probe, shown previously to bind independently two molecules of GATA-1 (23) is 54 bp, the others are 30 bp). The migration of complexes containing 1 or 2 molecules of GATA-1 are indicated by the filled and open arrows, respectively. B.) Methylation interference assays for single or double complexes. The partially methylated probes were labeled on the top strand (left) or bottom strand (right) and used in mobility-shift assays as in (A). After purification of the unbound free probe (F), monomer bound complex (B1) or double bound complex (B2), DNA was cleaved by piperidine and equal counts run on a denaturing 15% polyacrylamide gel. The positions of GATA sites 1–3 are indicated on the left of each panel, as are the G residues (arrow heads) at –148 (top strand) and –147 (bottom strand). C.) Summary of the interference data. The sequence of each probe is given; numbers correlate with those shown in Fig. 3. On the wild-type probe (WT), the positions of GATA motifs 1–3 are indicated; the mutated sites of the other three probes are boxed. Autoradiographs were scanned by densitometry and band intensity was normalized (if necessary) to bands outside of the GATA motifs. The height of each bar represents the relative intensity difference for a given band between the free and bound fraction (open bar, single complex; filled bar, double complex).

is revealed by a lack of a band in the bound fraction, indicating that methylation at that residue interferes with productive binding.

The methylation interference results support the evidence that multiple GATA-1 molecules interact with the ARE. In the case

of the wild-type ARE probe, methylation at the site 2 residues is most deleterious to formation of the single-bound complex. However, in the double-bound complex, when two sites must be filled, methylation at either site 2 or site 3 equally inhibits

Figure 5. GATA-1 bends DNA. A.) Gel mobility-shift assays were performed using erythroid nuclear extract and probes containing a circularly permuted wild-type ARE binding site (WT in Fig. 4c). Lane 0 is an identical assay in which the ARE was omitted; lanes 1–10 correspond to the probes described in (C). The migration of complexes containing 1, 2, and possibly 3 molecules of GATA-1 are indicated. The P marks the migration of unbound free probe. B.) The experiment shown in (A) was repeated but run longer on the gel (the free probe has just run off) to calculate the bending angles, as described (25). C.) The polylinker of the pBEND2 vector is illustrated, showing restriction sites used to generate the permuted probes (1–10). The open rectangle represents the wild-type ARE which was cloned into the XbaI site.

formation of the complex. The results of analyzing single and double-bound complexes for the various probes is presented schematically in Fig. 4c.

The interference pattern indicates that GATA-1 lies along one face of the helix on the wild-type oligomer and methylation in either the major (G residues) or minor (A residues) groove affects binding. Close inspection of the patterns of interference reveals that binding of an additional molecule of GATA-1 can influence the interference pattern. This might indicate that bound GATA-1 molecules can interact with each other. For example, the G residue at –148 is a critical determinant for a single molecule bound at site 2, but is no longer important for binding when a second molecule is bound at site 3. In addition, we note that alteration of one GATA motif may affect the interference pattern detected at a different site. Critical binding determinants are detected primarily at sites 2 and 3; site 1 might only be bound when site 2 is mutated (ARE-2). However, data concerning site 1 is difficult to analyze as the site is located very close to the end of the probe.

Because mutations 3' of site 3 influence the off-rate of GATA-1 at that site (Fig. 2c), we analyzed whether mutations upstream (at site 2) have a similar effect. The results (shown in Fig. 2d) indicate this is the case, as the off-rate of a single-bound complex is somewhat higher using the site 2 mutant probe, relative to the wild-type ARE. The off-rates for the double-bound complexes is however not effected by the site 2 mutation. This result confirms the potential influence of flanking sequence on binding at site 3, although it fails to provide evidence for stabilizing interactions between GATA-1 bound at both sites 2 and 3.

Bending of DNA by GATA-1

Close inspection of the mobility-shift assay shown in Fig. 4a reveals that the ARE-3 probe migrates as a single complex slightly further than the single complexes formed using the other probes. Therefore, to further investigate the interaction of GATA-1 with the ARE, we systematically analyzed the ability of bound GATA-1 to introduce a bend into the target DNA. Bending is

measured by the gel mobility-shift assay, using DNA fragments in which the position of the ARE is located at various locations relative to the ends of the probe (25). The mobility of protein/DNA complexes is greater when a bend is present at the end of the DNA fragment than when it is located in the center.

Gel mobility-shift assays are shown in Fig. 5, using circularly permuted probes containing the wild-type ARE consisting of all three GATA motifs. The results indicate that GATA-1 does bend DNA, and the extent of the bend differs depending on the complex formed. In this case we can clearly detect three distinct complexes when the autoradiograph is over-exposed. At this time we can not determine whether the slowest mobility complex is due to a triple-bound complex (insufficient material could be collected for methylation interference analysis), although in the absence of the ARE no specific complexes form on these probes. We calculate the maximum degree of bending for the three complexes to be 51°, 64°, and 109°, (complexes 1–3, respectively). These are significant bending angles, comparable to those generated by prokaryotic DNA-binding proteins such as CRP, LacI, and Cro (25).

DISCUSSION

We find that for function in erythroid cells the cGATA-1 promoter is completely dependent on a single GATA-binding site located 140 bp upstream of the mRNA start sites. In contrast, the mammalian gene appears to require a complex distal GATA site in addition to proximal CACCC motifs (11). By our deletion analysis we defined a minimal GATA-1 promoter and have designated the complex GATA-binding region an ARE, as this gene is clearly activated by its own product through this element. An additional distal GATA site is also required for full promoter activity. In our limited analysis we find that other upstream binding sites do not contribute significantly to expression, including potential sites for interaction with c-myc, c-ets, BGP1, or CACCC factors.

It is unlikely that the ARE and distal GATA site function alone for promoter activity. We have not yet analyzed elements downstream of the ARE, due to technical difficulties in working with this sequence (for example, complete lack of restriction sites and long runs of poly dG). A good candidate for an element which may cooperate with the ARE is a proximal Sp1 binding site, located just upstream of the predominant start sites. We have previously shown GATA-1 to be sufficient for trans-activation of a minimal α -globin promoter containing single GATA and Sp1 binding sites (7). The regulation of the GATA-1 promoter may be similar to this α -globin gene, in that both are expressed in erythroid cells of all lineages and therefore do not require developmentally stage-specific regulators.

We have analyzed the function of the cGATA-1 promoter in primary erythroid cells containing abundant GATA-1. Therefore, our results bear on the maintenance of gene expression in the differentiated cell-type, rather than the activation of the 'switch' factor. Obviously, other factors (positive and/or negative acting) must function to regulate this gene in early development prior to the accumulation of GATA-1 protein. This could include other GATA factors which might function during specific stages of erythropoiesis.

We find that minor sequence changes in the region near the ARE result in significant effects on promoter activity; at least some of these may be explained by subtle alterations in GATA binding. Evidence has been accumulating that a precise DNA/protein interaction is required for the full activity of the GATA-1 transcription factor to be realized. For example, we find mutations of the GATA-1 DNA-binding domain which show very minor alterations in binding (slightly altered methylation interference patterns) yet do not trans-activate (Yang, H.-Y. and T.E., submitted). Data consistent with this notion is derived from studies of a natural HPFH, in which a mutated GATA-1 binding site leads to a very subtle alteration in the DNA-protein interaction, but results in inappropriately high levels of globin gene expression (28).

Finally, we have provided evidence that GATA-1 may bend DNA upon interaction with the GATA motif, and that this effect may be enhanced by binding of multiple molecules at a complex site such as the -140 ARE. We do not yet know if any functional significance can be ascribed to this structural alteration, nor can we rule out that the effect is due to a writhing phenomenon rather than bending per se. However, bending by DNA-binding eukaryotic transcription factors has recently been noted in a number of systems (29,30), and seems at least theoretically capable of facilitating distant protein/protein interactions. Because of the evidence that flanking sequences not directly implicated as binding determinants seem to influence GATA-1 activity, it is also interesting to speculate what role bending could play in determining interactions of GATA-1 with DNA. For example, it might be necessary to consider that certain mutations or bound proteins outside of the GATA motif could alter the binding and/or transcriptional activity of GATA-1 indirectly by influencing this bending.

ACKNOWLEDGEMENTS

We thank S. Adyha (NIH) for the gift of pBEND2. T.E. is a Searle Scholar and gratefully acknowledges support from the Chicago Community Trust. In addition, parts of this work were aided by Basil O'Connor Starter Scholar Research Award No. 5-91-0521 and by the National Institutes of Health, PHS grant No. DK44167 (both to T.E.).

REFERENCES

1. Evans, T., Felsenfeld, G. and Reitman, M. (1990) *Annu. Rev. Cell Biol.* **6**, 95–124.
2. Orkin, S.H. (1990) *Cell*, **63**, 665–672.
3. Evans, T. and Felsenfeld, G. (1989) *Cell*, **58**, 877–885.
4. Tsai, S.-F., Martin, D.I.K., Zon, L.I., D'Andrea, A.D., Wong, G.G. and Orkin, S.H. (1989) *Nature*, **339**, 446–451.
5. Trainor, C.D., Evans, T., Felsenfeld, G. and Boguski, M.S. (1990) *Nature*, **343**, 92–96.
6. Zon, L.I., Mather, C., Burgess, S., Bolce, M.E., Harland, R.M. and Orkin, S.H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10642–10646.
7. Evans, T. and Felsenfeld, G. (1991) *Mol. Cell. Biol.*, **11**, 843–853.
8. Zon, L.I., Youssoufian, H., Mather, C., Lodish, H.F. and Orkin, S.H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 10638–10641.
9. Chiba, T., Ikawa, Y. and Todokoro, K. (1991) *Nucl. Acids Res.*, **19**, 3843–3848.
10. Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S.-F., D'Agati, V., Orkin, S.H. and Costantini, F. (1991) *Nature*, **349**, 257–260.
11. Tsai, S.-F., Strauss, E. and Orkin, S.H. (1991) *Genes Dev.*, **5**, 919–931.
12. Hannon, R., Evans, T., Felsenfeld, G. and Gould, H. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 3004–3008.
13. Nicolis, S., Chiara, B., Ronchi, A., Crotta, S., Lanfranco, L., Moroni, E., Giglioni, B. and Ottolenghi, S. (1992) *Nucl. Acids Res.*, **19**, 5285–5291.
14. Bergson, C. and McGinnis, W. (1992) *EMBO J.*, **9**, 4287–4297.
15. Chen, R., Ingraham, H.A., Treacy, M.N., Albert, V.R., Wilson, L. and Rosenfeld, M.G. (1990) *Nature*, **346**, 583–586.
16. Thayer, M.J., Tapscott, S.J., Davis, R.L., Wright, W.E., Lassar, A.B. and Weintraub, H. (1989) *Cell*, **58**, 241–248.
17. Schier, A.F. and Gehring, W.J. (1992) *Nature*, **356**, 804–807.
18. Lewis, C.D., Clark, S.P., Felsenfeld, G. and Gould, H. (1988) *Genes Dev.*, **2**, 863–873.
19. Cunningham, T.S. and Cooper, T.G. (1991) *Mol. Cell. Biol.*, **11**, 6205.
20. Fu, Y.-H. and Marzluf, G.A. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 5331–5335.
21. Yamamoto, M., Ko, L.J., Leonard, M.W., Beug, H., Orkin, S.H. and Engel, J.D. (1990) *Genes Dev.*, **4**, 1650–1662.
22. Kudla, B., Caddick, M.X., Langdon, T., Martinez-Rossi, N.M., Bennett, C.F., Sibley, S., Davies, R.W. and Arst, H.N., Jr. (1990) *EMBO J.*, **9**, 1355–1364.
23. Evans, T., Reitman, M. and Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5976–5980.
24. Martin, D.I.K. and Orkin, S.H. (1990) *Genes Dev.*, **4**, 1886–1898.
25. Kim, J., Zwieb, C., Wu, C. and Adhya, S. (1989) *Gene*, **85**, 15–23.
26. Li, R., Knight, J., Bream, G., Stenlund, A. and Botchan, M. (1989) *Genes Dev.*, **3**, 510–526.
27. Maxam, A.M. and Gilbert, W. (1980) *Meth. Enzymol.*, **65**, 499–560.
28. Martin, D.I.K., Tsai, S.-F. and Orkin, S.H. (1989) *Nature*, **338**, 435–438.
29. Verrijzer, C.P., Van Oosterhout, J.A.W.M., Van Weperen, W.W. and Van der Vliet, P.C. (1991) *EMBO J.*, **10**, 3007–3014.
30. Kerppola, T.K. and Curran, T. (1991) *Cell*, **66**, 317–326.