Nuclear Factor Binding Sites in Human β Globin IVS2*

(Received for publication, August 9, 1995)

Christine E. J. Jackson‡§, David O’Neill‡¶, and Arthur Bank‡**‡‡

From the Departments of ‡Genetics and Development, **Medicine, and ¶Pathology, College of Physicians and Surgeons, Columbia University, New York, New York 10032

The second intron of the human β globin gene (β IVS2) has been previously identified as a region required for proper expression of β globin. To further characterize this region, we have footprinted the entire β IVS2 and have analyzed regions of interest by electrophoretic mobility shift assay. Through these studies we have identified four utilized binding sites for the erythroid regulatory factor GATA-1, two sites bound by general transcription factor Oct-1, and a site bound by the nuclear matrix attachment DNA binding protein SATB1. The binding of this latter protein to IVS2 suggests that IVS2 leads to a substantial change in the expression of β globin and renders the gene uninducible (18).

β IVS2 has also been shown to be a nuclear matrix attachment region (MAR), one of 10 found in the β globin complex (19, 20). These MARs may be involved in transcription, splicing, and replication of DNA (21–23). In order to identify DNA binding proteins that may play a role in the expression of human β globin, we have characterized the entire human β IVS2 region by DNase I footprint analysis. Based on the footprint pattern and transcription factor binding analysis utilizing a computer-generated map, sites were chosen to be analyzed by EMSA. We report here evidence that three of nine potential GATA-1 binding sites are utilized, as well as a fourth degenerate GATA-1 site. Additional proteins seen to bind IVS2 are Oct-1, a ubiquitous factor binding to a homeobox consensus site, several stage or tissue-specific factors, and the nuclear matrix binding protein SATB1. The binding of this latter protein to β IVS2 suggests that β IVS2 and nuclear matrix attachment may be involved in the regulation of β globin transcription.

MATERIALS AND METHODS

Subcloning and Footprint Analysis—β IVS2 was cleaved into three smaller pieces as depicted (see Fig. 1). The parent vector for β IVS2 subcloning was pSP72p 3.4, which contains a 3.4-kb ClaI/SphI fragment of the human β globin gene subclone into the ClaI and SphI sites of vector pSP72 (Promega). The BamHI–EcoRI fragment of the gene, including all of β IVS2, was analyzed further. To produce subclone I, pSP72p3.4 was sequentially digested with restriction enzymes BamHI and Muni, and the 200-bp insert was ligated into the EcoRI and BamHI sites of the Bluescript SK+ vector (Stratagene). For subclone II, pSP72p3.4 was sequentially digested with MunI and Dral, and the 329-bp insert was subcloned into the EcoRI and Smal sites of plasmid pGEM T7Z+ (Promega). For subclone III, pSP72p3.4 was sequentially digested with EcoRI and Dral, and the resultant 400-bp insert was subcloned into the EcoRI and Smal sites of pGEM T7Z+.

All three subclones were entirely footprinted at least twice each in both orientations, using nuclear extracts from CEM (24), HEL-92 (25), HeLa (26), K562 (27), and MEL (28) cells. CEM is a human T-lymphocyte cell line; HEL-92 and K562 are human fetal-embryonic hematopoietic lines; and MEL is a murine adult hematopoietic line. HeLa cells are human cervical carcinoma cells and are not hematopoietic. Linearized subclones were dephosphorylated and end-labeled with [γ-32P]ATP using T4 polynucleotide kinase. Following a second restriction digest, specifically, in the region spanning the 3'-end of IVS2 and the beginning of exon 3 (13–15). Previous work has indicated that β IVS2 is required for β globin expression (16). Two DNase I hypersensitive sites have been identified within the β globin structural gene, a stronger site in exon 3, and a weaker site in the center of IVS2 (17). The β IVS2 intronic enhancer region has a utilized binding site for erythroid transcription factor GATA-1.2 We have previously shown in murine erythroleukemia (MEL) cells that the IVS2 sequences from the δ globin gene will not substitute for β IVS2 in human β globin expression. The replacement of β IVS2 with δ IVS2 leads to a substantial decline in the expression of β globin and renders the gene uninducible (18).

The adult human β globin gene has been shown to include enhancers 3' to the structural gene and within the gene itself.

* This work was supported in part by Public Health Service Grants from the National Institutes of Health DK-25274, HL-28381, and HL-48345 from the National Institutes of Health. 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™/EMBL Data Bank with accession number(s) J 00179.

‡ Supported by a National Institutes of Health Hematology Training Grant DK-07373.

¶ Supported by a National Institutes of Health Clinical Investigator Award DK-02260.

§ To whom correspondence should be addressed: Columbia University College of Physicians and Surgeons, 701 W. 168th St., HSSC 1602, New York, NY 10032. Tel.: 212-305-4186; Fax: 212-923-2090.

The abbreviations used are: kb, kilobase pair(s); MEL, murine erythroleukemia; MAR, matrix attachment region; LCR, locus control region; IVS2, intervening sequence 2; SATB1, special A-T-rich binding protein 1; bp, base pair(s); EMSA, electrophoretic mobility shift assay.

2 M. Flamm, submitted for publication.
probes were polyacrylamide gel electrophoresis-purified on a 5% non-denaturing polyacrylamide gel (29). Some of each probe was subjected to Maxam-Gilbert sequencing (29) to provide a sequence ladder for footprint gels. Labeled probe was footprinted by DNase I digestion as described previously (30) using 50 μg of nuclear extract and 160 ng of DNase I/reaction, except where indicated.

Nuclear Extracts—All nuclear extracts were prepared from cells as described previously (31) or by a small scale adaptation of this method (32).

Electrophoretic Mobility Shift Assays—After annealing, oligonucleotide probes were end-labeled using T4 polynucleotide kinase and [γ-32P]ATP. In general, 20,000 cpm of probe/reaction were used. EMSA were run on 3.5, 4, or 5% nondenaturing polyacrylamide gels as described previously (30). Reactions included 10 μg of nuclear extract and 5 μg of double-stranded poly(dI-dC) as nonspecific competitor each. EMSA buffer was 20 mM Tris-HCl, pH 7.6, 10% glycerol, 0.2 mM EDTA, 2.5 mM MgCl2, and 60 mM KCl, with 1 mM diithothreitol and 0.2 mM phenylmethylsulfonyl fluoride. Reactions were incubated for 20 min on ice. For competition assays, 10-, 100-, or 1000-fold molar excess of unlabeled competitor oligonucleotides was added to the reaction. For supershift assays, reactions were incubated as usual for 20 min on ice, and then 1 μg of antibody was added and the reactions were further incubated for 45 min to 1 h on ice. All antibodies were Supershift grade antibodies from Santa Cruz Biochemicals, except for the anti-SATB1 antibody, which was the generous gift of Dr. T. Kohwi-Shigematsu (La Jolla Cancer Research Foundation, La Jolla, CA).

Oligonucleotides—The following oligonucleotides (sense strands listed), which were synthesized and polyacrylamide gel electrophoresis-purified by Genosys Biotechnologies, Inc., were used in these analyses: site 2, 5′-GAATGGGAAACAGACGAATGATTGCATCAGTGTGGAAG-

| Footprint number | Footprint coordinates and extracts | Gel shift analysis | Putative factor binding sitesa |
|------------------|-----------------------------------|-------------------|-------------------------------|
| 1                | 60–78 s, hs                        | ND                |                               |
| 2                | 110–130 s                          | 42-bp oligo binds GATA-1 in erythroid lines. | GATA-1, oct-1, oct-2 |
| 3                | 100–109 a, hs only erythroid       |                                                               |
| 4                | 155–176 s                          | ND                |                               |
| 5                | 216–251 a, hs                      | 60-bp oligo. Complex gel shift pattern includes two bands present only in HEL-92, K562, and weakly in CEM. Also, a band unique to MEL, and a band only in CEM and K562. A ubiquitous band is competed by an oct-1 consensus oligo and supershifted by an anti-oct-1 antibody. Does not bind GATA-1 on gel shift. | GATA-1, oct-1, oct-2 |
| 6                | 331–380 all extracts               | 41-bp oligo. Complex gel shift pattern includes a band present only in HEL-92 and K562. A ubiquitous band is competed by an Antennapedia consensus oligo, as well as an oct-1 consensus oligo. Binds nuclear matrix binding protein SATB1 in CEM. Does not bind GATA-1 on gel shift. | GATA-1, oct-1, oct-2 |
| 7                | 397–428 all extracts               | 60-bp oligo. Complex gel shift pattern includes two bands present only in HEL-92, K562, and weakly in CEM. Also, a band unique to MEL, and a band only in CEM and K562. A ubiquitous band is competed by an oct-1 consensus oligo and supershifted by an anti-oct-1 antibody. Does not bind GATA-1 on gel shift. | GATA-1, oct-1, oct-2 |
| 8                | 531–569 s                          | 40-bp oligo binds GATA-1 in erythroid lines. This oligo spans two GATA-1 binding consensus sequences. Binds nuclear matrix binding protein SATB1 in CEM and K562. A faint band in CEM and K562 is competed by an oct-1 consensus oligo. | GATA-1, oct-1, oct-2 |
| 9                | 604–591 all extracts               | 20-bp oligo. No binding to double or single stranded probe. | GATA-1, oct-1, oct-2 |
| 10               | 604–629 all extracts               | 29-bp oligo binds GATA-1 in erythroid lines. | GATA-1, oct-1, oct-2 |
| 11               | 649–689 s                          | 28-bp oligo. Complex gel shift pattern of which one band is seen only in CEM, HEL-92, K562, and NIH 3T3. | GATA-1, oct-1, oct-2 |
| 12               | 719–733 s                          | 28-bp oligo. Complex gel shift pattern of which one band is seen only in CEM, HEL-92, K562, and NIH 3T3. | GATA-1, oct-1, oct-2 |
| 13               | 808–822 s                          | ND                |                               |
| 14               | 832–842 s                          | Binds GATA-1c     | GATA-1, oct-1, oct-2 |

* Recognition sites for certain DNA binding factors of interest, as determined by the tfsites portion of the University of Wisconsin GCG program.

b All GATA-1 binding sites conform to GCG tfsites consensus sequences with the exception of footprint 2.

c M. Flamm, submitted for publication.
IVS2 was subcloned in three pieces into pGEM 7 or Blue Vector. TATTTAGAATCTTTAATTTCTAATATATTTAGAA-3' (21); and Santa AATTGGCGGCTGGT-3' (22). Site 6, 5'-GTTTACATTATTTCAATTTCTAATATATTTAGAA-3'. Site 5, 5'-CATGCTCCTTGGCAGCATT-3'. Site 10, 5'-ATTCTAAAGAATAATGATACATAATCATTATACATATTTATGGG-3'. Site 9, 5'-ATTGATACATAATCATTATACATATTTATGGG-3'. Site 7, 5'-CAGCTACAAT-3'. Site 12, 5'-CAGTGATAATTTCTG-3'. Site 11-850, with accession number J00179.

RESULTS

Footprinting Analysis of Human β IVS2—The 850-bp human β IVS2 was subcloned in three pieces into pGEM 7 or Blue Vector. Sk-1 restriction map was analyzed to determine which sequences to characterize by EMSA.領域を追加した表現には、以下のフィールドが利用できます。以下に示すフィールドは、特定の要素を示すために使用されます。

Fig. 1. β IVS2 subclones for footprint analysis. Restriction map of the three subclones (I, II, and III) used in footprinting β IVS2, as described under "Materials and Methods." β IVS2 has been numbered 1–850, with 1 corresponding to the start of IVS2 (base number 62684 in the GenBank human globin gene complex sequence, accession number J 00179).

Fig. 2. Characterization of nuclear factor binding sites in β IVS2. All experiments were performed using CEM, HEL-92, HeLa, K562, and MEL nuclear extracts. β IVS2 has been numbered 1–850, with 1 corresponding to the start of IVS2 (base number 62684 in GenBank β globin complex sequence). Included as reference points are the BamHI site in exon 2 and the EcoRI site in exon 3. A, sites of footprints observed by DNase I digestion of end-labeled probe. Thicker bands are footprints observed on both DNA strands; thinner bands are footprints observed on one strand. Sites have been labeled as to which extracts generated the footprint and are numbered 1–14 for easy reference. s, sense; a, antisense; C, CEM; E, erythroid (HEL-92, K562, and MEL); A, all extracts; EF, embryonic-fetal (HEL-92 and K562); hs, hypersensitive site. B, consensus sites for binding of GATA-1 (33, 34). C, sites found experimentally to bind GATA-1.

Fig. 3. EMSA of site 2 with the 42-bp site 2 oligonucleotide. Gels are 5% polyacrylamide. A, EMSA with nuclear extracts as marked. Lane 1, control with no nuclear extract. Bands 1a and 1b are seen only in erythroid cell lines HEL-92, K562, HeLa, and MEL. Lane 3, competitor with no nuclear extract. Bands 2a and 2b are seen only in erythroid cell lines HEL-92, K562, and MEL. Lane 5, competition EMSA with site 2; competitor oligonucleotides as marked. GATA1m is the mutant GATA oligonucleotide. Lanes 2, 5, and 7 include 10 mol excess of unlabeled competitor probe; lanes 3, 6, and 8 include 10 mol excess of unlabeled competitor probe.
and 10. Each of these three probes was specifically competed by an excess of unlabeled competitor GATA-1 consensus oligonucleotide, but not with a GATA-1 mutant consensus oligonucleotide (Fig. 3B and Fig. 8B). Site 14 has been previously characterized in this laboratory, and was found to bind GATA-1.

Site 5 Analysis—Footprint site 5, a broad footprint seen with all nuclear extracts tested (Fig. 4), contains consensus sequences for transcription factors Oct-1 and Oct-2 (ATTTG-CAT), and GATA-1 (ATAATCTC) (Table I). EMSA done with a 60-bp oligonucleotide containing the ubiquitous footprint 5 sequence revealed a complex pattern (Fig. 5A). Within this site are two possibly stage-specific bands seen only with HEL-92, K562 (bands 2, a and b), and weakly with CEM nuclear extracts in some gels (not shown). There is also a different sized band seen only with MEL nuclear extract (band 4), a unique band found only with CEM and K562 (band 5), and a ubiquitous band (band 1). Each of these bands was competed by an excess of unlabeled site 5 probe (Fig. 5B, lanes 2, 8, and 9) but not by a nonspecific probe for the general transcription factor AP-1 (not shown). The large ubiquitous band (Fig. 5A, band 1) was competed by an Oct-1 consensus oligonucleotide (Fig. 5B, lanes 3, 10, and 11) and was also supershifted by an anti-Oct-1 antibody (lane 5). Additionally, the GATA-1 consensus oligonucleotide did not compete any bands even though a GATA-1 consensus is present in the site 5 oligonucleotide (lane 4). In an attempt to identify the nuclear factors responsible for bands 3, a and b, 4, and 5, supershift analyses were performed using antibodies against c-Fos, c-j un/AP-1, Oct-2 (lane 6), Fli-1, Pu.1, Ets 1/Ets 2, GATA-1, NFE-2, YY1, and SATB1. All of these were negative by supershift assay with the site 5 oligonucleotide.

Site 6 Analysis—Footprint site 6 is also seen with all nuclear extracts (Fig. 4) and includes consensus sequences for the homeobox protein engrailed (CAATTAAA) and GATA-1 (ATAATCAT). Fig. 6A shows an EMSA gel with an oligonucleotide synthesized to the site 6 footprint sequence. As occurred for site 5, the site 6 oligonucleotide EMSAs revealed a complex pattern buried within this ubiquitous footprint. This includes an embryonic-fetal erythroid specific band seen only with HEL-92 and K562 (band 2B). There is also a larger band seen with HEL-92, K562, and possibly MEL (band 2a), although the low mobility of this band makes resolution difficult. All bands are competed by an excess of unlabeled site 6 probe (Fig. 6B, lanes 2, 3, 6, and 7; Fig. 6C, lanes 2 and 9, Fig. 6D, lanes 2, 3, and 4), but not with the GATA-1 consensus oligonucleotide, although the gel shift oligonucleotide contains a GATA-1 consensus se-
One ubiquitous band is competed by an oligonucleotide containing the consensus sequence for the homeobox protein Antennapedia (CAATTAAA) (Fig. 6C, lane 4, and lanes 5–8) nuclear extracts, with competitor oligonucleotide or antibody (Ab) as marked. Lanes 2 and 6 include 100× molar excess of unlabeled competitor oligonucleotide; all other competition lanes include 1000× molar excess of unlabeled competitor oligonucleotide. A, control with no nuclear extract. Band 1, ubiquitous band; band 2a, HEL-92-, K562-, and MEL- (?) specific band; band 2b, HEL-92- and K562-specific band; band 3, SATB1; and band 4, ubiquitous “homeobox” band. Lower bands on the gel are likely proteolytic degradation products and are not reproducible.

B, competition and supershift EMSA of site 6 with CEM (lanes 1–4) and K562 (lanes 5–8) nuclear extracts, with competitor oligonucleotide or antibody (Ab) as marked. Lanes 2 and 6 include 100× molar excess of unlabeled competitor oligonucleotide; all other competition lanes include 1000× molar excess of unlabeled competitor oligonucleotide.

C, competition EMSA of site 6 with CEM (lanes 1–7) and K562 (lanes 8–13) nuclear extracts and competitor oligonucleotides as marked. GATAm and oct-1m are the GATA and Oct-1 mutant oligonucleotides, respectively. Lane 3 includes 100× molar excess of unlabeled competitor oligonucleotide; all other competition lanes include 1000× molar excess of unlabeled competitor oligonucleotide. D, competition EMSA of site 6 with HEL-92 nuclear extract. ANT is the Antennapedia consensus oligonucleotide. Lanes 2 and 5 include 100× molar excess of unlabeled competitor oligonucleotide; lanes 3 and 6, 100× molar excess of unlabeled competitor oligonucleotide; lanes 4 and 7, 1000× molar excess of unlabeled competitor oligonucleotide.

Interestingly, the nuclear matrix attachment DNA binding protein SATB1 was found to bind to the site 6 probe in CEM nuclear extract (Fig. 6A, band 3). β IVS2 has been described as one of nine sites in 90 kb of globin gene sequence studied to contain a MAR (19). Additionally, it has been noted that MARs from the human β globin gene can bind SATB1 (21). A SATB1 consensus oligonucleotide inhibits band 3 formation with the site 6 probe (Fig. 6C, lane 3), and the SATB1 band is supershifted by an anti-SATB1 antibody (Fig. 6D, lane 4). SATB1 is a 103-kDa protein (39), and the SATB1 runs very slowly on EMSA. It is also possible that the SATB1 band seen in Fig. 6B might be a complex of SATB1 and some other protein. SATB1 complexes have been suggested in a recent paper describing the binding of SATB1 to an αβγ globin regulatory region (20).

Site 7 Analysis—The sense strand footprint at site 7, which is seen with CEM, HEL-92, and K562 nuclear extracts (Fig. 7) has consensus sequences for the homeobox protein bicoid (CCTAATCTC) and GATA-1 (CCTAATCTC). This footprint is broader in HEL-92 and K562, and the region footprinted only in HEL-92 includes a GATA-1 consensus sequence. EMSA with a 40-bp oligonucleotide including the site 7 footprint showed a more complex pattern with erythroid nuclear extracts than other extracts (Fig. 8A). The site 7 probe used spanned two GATA-1 consensus sequences (CCTAATCTC and TTAATCTCA), and in fact proved to bind GATA-1 in the erythroid lines, generating a GATA-1 band which could be supershifted with an anti-GATA-1 antibody (Fig. 8A, lane 8; Fig. 8C, lane 7). This band was also specifically competed with a GATA-1 consensus oligonucleo-
that this region has several interesting structural and functional features; it contains a 3’-enhancer region (13–15), two DNase I hypersensitive sites (17), and is required for proper expression of the β globin gene (16). We have previously analyzed the expression in MEL cells of β constructs in which β IVS2 has been replaced by δ or γ globin IVS2, and have found that these globin IVSs are not interchangeable. When β IVS2 is replaced with δ IVS2, the base-line expression of β is greatly decreased, and the cells are not inducible with Me2SO (18). In addition, constructs in which β IVS2 has been replaced with γ IVS2 produce β transcripts that are improperly initiated in K562 cells (42). Comparison of δ, γ, δ, and β IVS2 using restriction maps and maps generated by the tfsites data base reveals no significant sequence conservation on the nucleotide level, and few conserved potential transcription factor binding sites, except for two GATA-1 binding sites that are conserved in position. The first is the second intronic GATA-1 site (Fig. 2), which is conserved in position between δ and β globin. The second is the seventh GATA-1 site, which is conserved in position between δ and β globin. Neither of these sites were found to bind GATA-1 in our experiments and are not apparently functionally important, at least in the expression of β globin.

The footprint pattern of β IVS2 and those areas studied by EMSA have revealed a very dense and complex pattern of protein binding (Table I). Previous studies in which only the DNase I hypersensitive site of murine β IVS2 was characterized also revealed a complex gel shift pattern (43). These gel shift analyses covered about one-third of the total sequence of murine β IVS2. Two proteins were identified as binding to murine β IVS2, GATA-1 and Spi-1/PU.1, an Ets family protein (44). These murine β IVS2 binding sites are not conserved in human β IVS2. Human β IVS2 does contain four potential Ets binding sites, but only one of these is footprinted in human β IVS2 (site 14), and this one site has been shown to bind GATA-1 only.2 The complexity of the binding pattern seen in IVS2 suggests that it is an area of complex regulatory function, and might be involved in the regulation of β expression in the adult, and perhaps in earlier stages of erythropoiesis. Certainly a regulatory function is supported by the extensive binding of erythroid transcription factor GATA-1 to this region. The redundancy of the GATA-1 consensus sequences alone (nine sites), unique among the globin genes, indicates that some function is likely. By comparison, the human δ globin gene has only three GATA-1 consensus sequences, and the human γ gene only two GATA-1 consensus sequences. Three of the nine consensus sites in β IVS2 are bound by GATA-1 as is a fourth related sequence. It is interesting that although binding sites for GATA-1 have been extensively characterized (34–36), one still cannot predict with certainty which sites are utilized in vivo or in vitro.

Several stage- or tissue-specific bands were seen on EMSA of β IVS2. The gel shift analyses on site 5 in particular revealed several bands of interest (Fig. 5, A and B), none of which could be supershifted by antibodies to general transcription factors or known factors important in the regulation of globin expression such as NFE-2, YY1, or GATA-1. Of particular interest is a band seen only with MEL (adult erythroid) cells (Fig. 5A, band 4), as this could represent a potential factor for positive expression of β globin. The two bands seen with HEL-92 and K562 (bands 3, a and b) and faintly in CEM and the band seen only with CEM and K562 nuclear extracts (band 5) are also intriguing. Perhaps these proteins are not seen in murine MEL nuclear extract due to the species difference, or possibly they play a role in embryonic-fetal erythropoiesis, or in lymphoid cells. The HEL-92, K562, and possibly MEL-specific band bound to the site 6 oligonucleotide (Fig. 6A, band 2a) can be approximately sized due to the presence of SATB1 (band 3) binding to

Discussion

Human β globin IVS2 has been entirely footprinted and further characterized by EMSA. Previous data have indicated that this region has several interesting structural and functional features; it contains a 3’-enhancer region (13–15), two DNase I hypersensitive sites (17), and is required for proper expression of the β globin gene (16). We have previously analyzed the expression in MEL cells of β constructs in which β IVS2 has been replaced by δ or γ globin IVS2, and have found that these globin IVSs are not interchangeable. When β IVS2 is replaced with δ IVS2, the base-line expression of β is greatly decreased, and the cells are not inducible with Me2SO (18). In addition, constructs in which β IVS2 has been replaced with γ IVS2 produce β transcripts that are improperly initiated in K562 cells (42). Comparison of δ, γ, δ, and β IVS2 using restriction maps and maps generated by the tfsites data base reveals no significant sequence conservation on the nucleotide level, and few conserved potential transcription factor binding sites, except for two GATA-1 binding sites that are conserved in position. The first is the second intronic GATA-1 site (Fig. 2), which is conserved in position between δ and β globin. The second is the seventh GATA-1 site, which is conserved in position between δ and β globin. Neither of these sites were found to bind GATA-1 in our experiments and are not apparently functionally important, at least in the expression of β globin.

The footprint pattern of β IVS2 and those areas studied by EMSA have revealed a very dense and complex pattern of protein binding (Table I). Previous studies in which only the DNase I hypersensitive site of murine β IVS2 was characterized also revealed a complex gel shift pattern (43). These gel shift analyses covered about one-third of the total sequence of murine β IVS2. Two proteins were identified as binding to murine β IVS2, GATA-1 and Spi-1/PU.1, an Ets family protein (44). These murine β IVS2 binding sites are not conserved in human β IVS2. Human β IVS2 does contain four potential Ets binding sites, but only one of these is footprinted in human β IVS2 (site 14), and this one site has been shown to bind GATA-1 only. The complexity of the binding pattern seen in IVS2 suggests that it is an area of complex regulatory function, and might be involved in the regulation of β expression in the adult, and perhaps in earlier stages of erythropoiesis. Certainly a regulatory function is supported by the extensive binding of erythroid transcription factor GATA-1 to this region. The redundancy of the GATA-1 consensus sequences alone (nine sites), unique among the globin genes, indicates that some function is likely. By comparison, the human δ globin gene has only three GATA-1 consensus sequences, and the human γ gene only two GATA-1 consensus sequences. Three of the nine consensus sites in β IVS2 are bound by GATA-1 as is a fourth related sequence. It is interesting that although binding sites for GATA-1 have been extensively characterized (34–36), one still cannot predict with certainty which sites are utilized in vivo or in vitro.

Several stage- or tissue-specific bands were seen on EMSA of β IVS2. The gel shift analyses on site 5 in particular revealed several bands of interest (Fig. 5, A and B), none of which could be supershifted by antibodies to general transcription factors or known factors important in the regulation of globin expression such as NFE-2, YY1, or GATA-1. Of particular interest is a band seen only with MEL (adult erythroid) cells (Fig. 5A, band 4), as this could represent a potential factor for positive expression of β globin. The two bands seen with HEL-92 and K562 (bands 3, a and b) and faintly in CEM and the band seen only with CEM and K562 nuclear extracts (band 5) are also intriguing. Perhaps these proteins are not seen in murine MEL nuclear extract due to the species difference, or possibly they play a role in embryonic-fetal erythropoiesis, or in lymphoid cells. The HEL-92, K562, and possibly MEL-specific band bound to the site 6 oligonucleotide (Fig. 6A, band 2a) can be approximately sized due to the presence of SATB1 (band 3) binding to

![Fig. 7. DNase I footprint of β IVS2 site 7. Lane 1, C lane of Maxam-Gilbert sequencing reaction; lane 2, control with 10 ng of DNase I; lane 3, control with 20 ng of DNase I. Reactions with nuclear extract include 50 μg of nuclear extract except as marked, and were treated with 160 ng of DNase I.](http://www.jbc.org/Doctor)
this oligonucleotide in CEM nuclear extract. The HEL-92 and K562 specific band runs more slowly than SATB1 which has a molecular mass of 103 kDa. This size would be larger than known erythroid regulatory proteins, with the exception of gPE.

gPE is a 108-kDa protein that binds to sites near human γ globin(45). Its broad pattern of tissue distribution argues against it being any of the uncharacterized proteins we have found. The particular pattern of expression of the HEL-92- and K562-specific band, i.e. only in embryonic-fetal erythroid cells, could be relevant to down-regulation of β globin expression early in development. This could be of particular importance if GATA-1 binding to β IVS2 is indeed important in positive regulation of β globin expression, as GATA-1 is certainly present in embryonic-fetal erythroid cells. Another differentially expressed band at site 12 (Fig. 9), seen with CEM, HEL-92, K562, and NIH 3T3 nuclear extracts, but not with the adult HeLa or MEL cells, is of unknown significance.

We have found one potential homeobox protein binding site in β IVS2 (site 6, Fig. 6A, band 4). Previous data have shown that homeobox proteins may be important in erythroid differentiation (46). Eight of nine genes in the HOX 2 cluster are expressed in erythroid cells, but rarely in B or T cells (46–49). There is also indirect evidence that HOX 3C may be necessary for adult hematopoiesis (50). A band found in all extracts at site 6 was competed by an oligonucleotide with the Oct-1 consensus sequence (ATTTGCAT) (Fig. 6C, band 4, lanes 6 and 12) and by an oligonucleotide with the Antennapedia consensus sequence (CAATTAAA) (Fig. 6D, band 4, lane 7). The Antennapedia
sequence is within the site 6 footprint and site 6 probe and is listed in the sites database as the engrailed consensus. This is the core consensus for many HOX proteins including HOX B6, which may have a role in erythroid differentiation (51). However, we do not see any erythroid cell specificity of the particular protein binding at this site. Although the site 7 footprinted region and oligonucleotide contain the consensus sequence for the homeobox protein bicaudal, all bands seen with K562 nuclear extracts could be competed by GATA-1, SATB1, or Oct-1 sequence oligonucleotides, and so all bands in erythroid cell extracts are accounted for. There seems to be no erythroid-related homeobox binding at this site in β IVS2.

We have found that the nuclear matrix-associated DNA binding protein SATB1 binds to site 6 of β IVS2 with CEM nuclear extract and more intensely to site 7 of β IVS2 with CEM and K562 nuclear extract (Fig. 6A, band 3; Fig. 8A, band 1). MARs are postulated to play an important role in the functional organization of chromatin loop domains. There is evidence that replication and transcription occur at the interface of DNA and the nuclear matrix and that the nuclear matrix is involved in RNA splicing (21–23). Recent reports have indicated that DNA binding of some transcription factors is associated with the nuclear matrix (52–55). MARs have a strong potential for extensive unpairing or unwinding. Although MARs often contain or reside close to enhancer sequences (21), their role is not clear as yet.

SATB1 is one of the characterized MAR binding proteins. It is a 103-kDa protein that binds as a monomer and is expressed primarily in thymus (21, 39). It binds selectively to MARs with well mixed ATC sequences (21, 39). β IVS2 has been previously characterized to be one of nine sites in the 90 kb of the human β globin gene locus to function as an MAR (19). MARs are regions of DNA at least 200 bp in length and are generally 70% AT-rich (21, 22). The areas binding SATB1 in β IVS2 are about 73% AT rich and do consist of a well mixed ATC sequence. Two distinct sites, in footprints 6 and 7, bind SATB1; two sites seem to be required for a strong SATB1 interaction to occur (21, 39). The site 6 and 7 oligonucleotides are, respectively, 83 and 75% AT-rich. The bands run very slowly on EMSA and may consist of a complex of SATB1, and some other protein as has been suggested (20).

Preliminary data show that SATB1 is a suppressor of transcription based on transient cotransfection assays with a reporter gene (21). One regulatory region to which SATB1 binds is the Igκ heavy chain intronic enhancer, which is flanked by MARs. In this context SATB1 may help to repress expression in non-B cells (56). SATB1 has also been observed to bind the β globin gene (21). Also, SATB1 has been recently reported to bind to the human Aγ 3′-regulatory region at sites I and IV (20).

These sites had been previously characterized as binding HOX protein 2.8 (2H) (57, 58). Besides being highly expressed in CEM cells, SATB1 was also found in heart, skeletal muscle, fetal liver, K562 cells, and B and T cells (20). It was proposed that the Aγ regulatory region might influence gene expression through interaction with the nuclear matrix. The regulatory region was also found to be an MAR, and this group speculated that promoter/enhancer interaction is mediated by SATB1 binding of MARs. However, they found no MAR near the Aγ promoter (20).

MARs and SATB1 binding in β IVS2 could have any of several functions. Previous data seem to indicate a correlation between MARs and enhancer regions (21, 22). Also, each β family gene (except εγ) harbors an MAR, while by comparison, no such sites exist in the large α globin gene complex (19). MARs might mediate an attachment between individual β globin family genes and the β globin LCR (which also contains MARs) (19), possibly having some role in globin switching. The β IVS2 MAR might increase expression mediated by the β IVS2 enhancer, as the β 3′-MAR (19), situated 500 bp downstream of the β 3′-enhancer, might facilitate expression mediated by this enhancer. Or, the β IVS2 MAR in combination with GATA-1 binding or binding of other factors might function as an independent enhancer in β IVS2. Possibly, MARs could mediate interaction between IVS2 and the β promoter or 3′-enhancer.

From the complexity of DNase I footprint and EMSA results we have obtained, it is clear that there are many interactions between human β IVS2 sequences and nuclear factors, both known factors and those yet to be characterized. The biological significance of the presence of protein-DNA interactions in β IVS2, and any interactions between β IVS2 and other regulatory sequences 5′ or 3′ to the human β gene, other β family genes or the LCR remain to be determined. The details of the relationships between DNA binding factors and globin gene switching also remain to be elucidated. Deletion analysis and site-directed mutagenesis of human β IVS2 transacting factor binding sites may provide new insights into the relationship between protein binding to this region and human β globin gene function.

REFERENCES
1. Blom van Assendelft, G., Hanscombe, O., Grosveld, F., and Greaves, D. R. (1989) Cell 56, 969–977.
2. Grosveld, F., Blom van Assendelft, G., Greaves, D., and Kollias, G. (1987) Cell 51, 975–985.
3. Talbot, D., Collins, P., Antoniou, M., Vidal, M., Grosveld, F., and Greaves, D. R. (1989) Nature 338, 352–355.
4. Orkin, S. H. (1990) Cell 63, 665–672.
5. Satamatopoulos, G., and Gelinas, R. (1997) in The Molecular Basis of Blood Diseases (Stamatoyannopoulos, G., and Papayannopoulou, T., eds), pp. 66–105, W. B. Saunders Company, Philadelphia, PA.
6. Tsai, S. F., Martin, D. I. K., Zon, L., D’Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) Nature 339, 446–451.
7. Evans, L. F., and Felsenfeld, G. (1989) Cell 58, 877–885.
8. Andrews, N. C., Erdjument-Bromage, H., Davidson, M. B., Tempst, P., and Orkin, S. H. (1993) Nature 362, 722–728.
9. Chan, J. Y., Han, X.-L., and Kan, Y. W. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 11382–11370.
10. Miller, I., and Bleier, J. J. (1993) Mol. Cell. Biol. 13, 2776–2786.
11. Caterina, J. J., Donze, D., Sun, C.-W., Ciavatta, D. J., and Townes, T. M. (1994) Nucleic Acids Res. 22, 2393–2391.
12. Mol, P., Chan, K., Asunis, I., Cao, A., and Kan, Y. W. (1994) Proc. Natl. Acad. Sci. U. S. A. 91, 9926–9930.
13. Antoniou, M., de Boer, E., Habets, G., and Grosveld, F. (1998) EMBO J. 7, 377–384.
14. Behringer, R. R., Hammer, R. E., Brinster, R. L., Palmiter, R. D., and Townes, T. M. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 7056–7060.
15. Trudel, M., and Gelinas, R. (1987) Biochemistry 26, 954–961.
16. Collins, P., Antoniou, M., and Grosveld, F. (1990) EMBO J. 9, 233–240.
17. Groudine, M., Kohwi-Shigematsu, T., Gelinas, R., Stamatoyannopoulos, G., and Papayannopoulou, T. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 7551–7555.
18. LaFlamme, S., Acuto, S., Markowitz, D., Vick, L., Landschulz, W., and Bank, A. (1987) J. Biol. Chem. 262, 4819–4826.
19. Jarman, A. P., and Higgs, D. R. (1988) EMBO J. 7, 3337–3344.
20. Cunningham, J. M., Purucker, M. E., Jane, S. M., Safer, B., Vanin, E. F., Ney,
Nuclear Factor Binding Sites in Human β Globin IVS2

28456

P. A., Lowrey, C. H., and Nienhuis, A. W. (1994) Blood 84, 1298–1308
21. Dickinson, L. A., Joh, T., Kohwi, Y., and Kohwi-Shigematsu, T. (1992) Cell 70, 631–645
22. Cockerill, P. N., and Garrard, W. T. (1986) Cell 44, 273–282
23. Sun, J.-M., Chen, H. Y., and Davie, J. R. (1994) J. Cell Biol. 55, 264–272
24. Foley, G. E., Lazarus, H., Farber, S., Uzman, B. G., Boone, B. A., and McCarthy, R. E. (1965) Cancer 18, 522–529
25. Martin, P., and Papayannopoulou, T. (1982) Science 216, 1233–1235
26. Gey, G. O., Coffman, W. D., and Kubicek, M. T. (1952) Cancer Research 12, 264–272
27. Lozzio, C. B., and Lozzio, B. B. (1975) Blood 45, 321–334
28. Friend, C., Scher, W. H., Holland, J. G., and Sato, T. (1971) Proc. Natl. Acad. Sci. U.S.A. 68, 378–382
29. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, pp. 6.46–6.48 and 13.88–13.94, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
30. O'Neill, D., Bornschlegel, K., Flamm, M., Castle, M., and Bank, A. (1991) Proc. Natl. Acad. Sci. U.S.A. 88, 8953–8957
31. Dignam, J. D., Martin, P. L., Shastry, B. S., and Roeder, R. G. (1983) Methods Enzymol. 101, 582–598
32. Schreiber, E., Matthias, P., Muller, M. M., and Schaffner, W. (1989) Nucleic Acids Res. 17, 6419–6420
33. Genetics Computer Group (1991) Program Manual for the GCG Package, version 7, Genetics Computer Group, Madison, WI
34. Wall, L., de Boer, E., and Grosveld, F. (1988) Genes & Dev. 1089–1100
35. Ko, L. J., and Engel, J. D. (1993) Mol. Cell. Biol. 13, 4011–4022
36. Merika, M., and Orkin, S. H. (1993) Mol. Cell. Biol. 13, 3999–4010
37. Vallerga, A., Shen, W.-F., Hauser, C., Largman, C., and Lawrence, H. J. (1991) Blood 78, 254 (Suppl.)
38. Alison, R. J., Lowrey, C. H., and Nienhuis, A. W. (1994) Blood 83, 1420–1427
40. Bamberidis, R. J., Swanson, M. S., and Dreyfuss, G. (1989) Genes & Dev. 3, 431–437
41. Jainchill, J. L., Aaronson, S. A., and Todaro, G. J. (1969) J. Virol. 4, 549–553
42. Donovon-Paluso, M., Acuto, S., Swanson, M., Dobkin, C., and Bank, A. (1987) J. Biol. Chem. 262, 17051–17057
43. Galson, D. L., and Housman, D. E. (1988) Mol. Cell. Biol. 8, 392–397
44. Galson, D. L., Hensold, J. O., Bishop, T. R., Schalling, M., D'Andrea, A. D., Jones, C., Auron, P. E., and Housman, D. E. (1993) Mol. Cell. Biol. 13, 2929–2941
45. Lloyd, J. A., Case, S. S., Ponce, E., and Lingrel, J. B. (1994) J. Biol. Chem. 269, 19385–19393
46. Lawrence, H. J., and Largman, C. (1992) Blood 80, 2445–2453
47. Shen, W.-F., Largman, C., and Lawrence, H. J. (1990) Adv. Exp. Med. Biol. 271, 211–219
48. Shen, W.-F., Detmer, K., Mathews, C. H. E., Hack, F. M., Morgan, D. A., and Lawrence, H. J. (1992) EMBO J. 11, 983–989
49. Lavelle, D., DeSimone, J., and Steigemann, P. (1994) Science 265, 1221–1225
50. Forrester, W. C., van Genderen, C., Jenuwein, T., and Grosschedl, R. (1994) Science 265, 1221–1225
