Regulatory Mechanism of Human Factor IX Gene: Protein Binding at the Leyden-Specific Region†

Sumiko Kurachi,‡ Midori Furukawa,‡,§ Jean-Philippe Salier,‡,∥ Ching-Tuan Wu,‡ Elizabeth J. Wilson,⊥ Frank S. French,⊥ and Kotoku Kurachi*,‡

Department of Human Genetics, University of Michigan Medical Center, Ann Arbor, Michigan 48109-0618, and Department of Pediatrics, University of North Carolina, Chapel Hill, North Carolina 27579

Received May 25, 1993; Revised Manuscript Received October 5, 1993®

ABSTRACT: Hemophilia B-Leyden is characterized by the gradual amelioration of bleeding after the onset of puberty. All Leyden phenotype mutations found to date lie within the Leyden-specific region, which spans roughly nt −40 to +20 in the 5′ end of the human factor IX gene. With HepG2 cell nuclear extracts, the Leyden-specific region and its immediate neighboring region of the normal factor IX gene showed five DNase I footprints: FP-I (nt +4 to +19), FP-II (nt −16 to −3), FP-III (nt −27 to −19), FP-IV (nt −67 to −49), and FP-V (nt −99 to −77). Protein binding affinities of short oligonucleotides containing sequences of FP-I, FP-II, or FP-III were substantially reduced in the presence of Leyden phenotype mutations in these areas, correlating well with the negative effects of these mutations on factor IX gene expression. A Leyden phenotype mutation at nt −20 (T to A) caused a loss of both footprints FP-III and FP-II but generated a new footprint, FP-III′ (nt −34 to −23), partially overlapping with FP-III, indicating mutation-dependent competitive protein binding at these sites. Although the FP-III′ area contains an androgen responsive element-like sequence, the nuclear protein that binds at FP-III′ is not androgen receptor. The protein was not recognized by anti-androgen receptor antibody and, furthermore, was present not only in liver but also in both androgen receptor-positive and androgen receptor-negative cells in electrophoretic mobility shift assays. The nuclear concentration of this protein increased significantly upon treatment of the HepG2 cells with testosterone. Its binding affinity to an oligonucleotide (−32sub) containing the FP-III′ sequence was greatly reduced in the presence of exogenous androgen receptor, suggesting a possible interaction of this protein with androgen receptor. The affinities of both this protein and a protein which binds to FP-III (presumably HNF-4) to −32sub with a mutation at nt −26 were grossly lowered. These findings suggest that the amelioration of hemophilia B-Leyden with a mutation at nt −20 after puberty involves binding of a specific non-androgen recepter nuclear protein at FP-III′ and it is able to substitute for the function of a protein bound at FP-III in the normal gene optimally through its elevated interaction with androgen receptor upon a surge of testosterone. The major transcriptional initiation site of the factor IX gene in human liver was determined to be at nt −176, localizing the entire Leyden-specific region to the 5′-untranslated region.

Factor IX (FIX)[1] plays a crucial role in the early phase and maintenance of blood coagulation (Hedner & Davie, 1989; Kurachi et al., 1993), and its deficiency from the systemic circulation results in hemophilia B (Briet et al., 1982). Like most other blood coagulation factors, the FIX gene is expressed with a high liver specificity (Salier et al., 1990). Its developmental regulation (Yao et al., 1991) and various structural elements responsible for its overall expression have been described (Salier et al., 1990; Jallat et al., 1990).

An intriguing hemophilia B phenotype, hemophilia B-Leyden, shows a gradual amelioration of abnormal bleeding due to an induction of FIX gene expression after the onset of puberty (Briet et al., 1982; Hirosawa et al., 1990; Kurachi et al., 1993). The Leyden-specific (LS) region, a small region arbitrarily defined as roughly nucleotides (nt) −40 to +20 of the 5′-end sequence, has been implicated in this unique gene regulation [Hirosawa et al., 1990; see Yoshitake et al. (1985) for the nucleotide numbering system]. All molecular defects in FIX–Leyden genes known to date are located in the LS-region (Giannelli et al., 1992; Thompson, 1991; Kurachi et al., 1992) and include single-base mutations at nt −21 (T to G) (Reijnen et al., 1993), −20 (T to A or C) (Reitsma et al., 1988; Giannelli et al., 1992), −6 (G to A or C) (Hirosawa et al., 1990; Crossley et al., 1990), −5 (A to T) (Picketts et al., 1992), +6 (T to A) (Freedenberg & Black, 1991), +8 (T to C) (Royle et al., 1991), or +13 (A to G, or deletion) (Reitsma et al., 1989). Recently, HNF-4 (LF-A1), a distant member of the steroid receptor family, was found to bind to the sequence element in the region of nt −20 (Reijnen et al., 1992; Crossley et al., 1992). This sequence element partially overlaps a possible androgen responsive element (ARE) spanning nt −36 to −22 of the region (Crossley et al., 1992). However, no information is currently available on the competition between HNF-4 and androgen receptor (AR) in their binding to these partially-overlapped sequence elements of the FIX gene.

† This work was supported in part by Research Grant HL38644 from the National Institutes of Health.

* To whom correspondence should be addressed at the Department of Human Genetics, 3712 Medical Science II Building, University of Michigan Medical Center, Ann Arbor, MI 48109-0618.

‡ University of Michigan Medical Center.

§ Present address: Institute of Gastroenterology, Tokyo Women's Medical College, Tokyo, Japan.

∥ Present address: INSERM Unite-78, 76233 Boisguillaume Cedex, France.

⊥ University of North Carolina.

® Abstract published in *Advance ACS Abstracts,* January 15, 1994.

[1] Abbreviations: AR, androgen receptor; ARE, androgen responsive element; bp, base pair; CAT, chloramphenicol acetyltransferase; ds, double-stranded; EMSA, electrophoretic mobility shift assay; FIX, factor IX; GR, glucocorticoid receptor; GRE, glucocorticoid responsive element; kb, kilobase; NEs, nuclear extracts; nt, nucleotide; PCR, polymerase chain reaction; rAR, recombinant androgen receptor; rGR, recombinant glucocorticoid receptor; RT, reverse transcription; RTase, reverse transcriptase; ss, single-stranded.

0006-2960/94/0433-1580$04.50/0

Table 1: Oligonucleotides Used in EMSA

subsequence oligonucleotide[a]	position (nt number)	sequence
-60sub	-65 to -47	GATGGACATTATTTCCCAG
-32sub	-40 to -20	ATACAGCTCAGCTTGTACTTT
-32Msub(-26:C)	-40 to -20	ATACAGCTCAGCTTCTACTTT
-20sub	-30 to -10	GCTTGTACTTTGGTACAACTA
-20Msub(-26:C)	-30 to -10	GCTTCTACTTTGGTACAACTA
-20Msub(-20:A)	-30 to -10	GCTTGTACTTAGGTACAACTA
-6sub	-18 to +1	CAGTACAACTAATCGACCTTATT[b]
-6Msub(-6:A)	-18 to +1	CAGTACAACTAATCAACCTTATT[b]
+13sub	+3 to +22	CACTTTCACAATCTGCTAGC
+13Msub(+13:G)	+3 to +22	CACTTTCACAGTCTGCTAGC
C3 ARE		AAGCTTAGTACGTGATGTTCTAAGCTT[c]

[a] Parentheses indicate mutant sequences with nt positions and base changes. All others are of the normal sequence. [b] The 5'- and 3'-end dinucleotide sequences (CA and TT, respectively) are spacer sequences unrelated to the FIX gene. [c] Underline indicates the ARE consensus sequence.

In order to delineate better the mechanism of action of the LS-region, we have carried out a systematic characterization of the region. Here we report DNA–protein and protein–protein interactions which explain in part the complex mechanisms underlying the action of the LS-region in regulation of the FIX gene. We also describe the establishment of the major transcription initiation site of the FIX gene in human liver.

EXPERIMENTAL PROCEDURES

Materials. Restriction enzymes and DNA modifying enzymes were obtained from Boehringer Mannheim or BRL. Reverse transcriptase (RTase) was obtained from Life Sciences, and Taq polymerase was purchased from BRL. Radioactive nucleotides ([α-^{32}P]dATP, [γ-^{32}P]ATP, and [α-^{35}S]dATP) were purchased from Amersham Corp. Double-stranded (ds) poly(dI-dC) and oligo(dT)–cellulose were from Pharmacia Biochemicals. Plasmids pCR1000 and pCRII were purchased from Invitrogen. The DNA sequencing kit and X-ray film (X-Omat AR) were obtained from United States Biochemicals and Kodak, respectively. Sterile plastic wares for tissue culture were from Corning. Dulbecco's modified Eagle's medium (DMEM) and RPMI, fetal bovine serum (FBS), antibiotics (penicillin–streptomycin), and L-glutamine were obtained from GIBCO. Testosterone (17β-hydroxyandrost-4-en-3-one), dihydrotestosterone [(5α,17β)-17-hydroxyandrostan-3-one], and charcoal-stripped FBS (F-6765) were from Sigma. Protein assay kit (no. 500-0006) was purchased from Bio-Rad. Purified recombinant glucocorticoid receptor (rGR) (Freedman et al., 1988) was obtained from Dr. Gelehrter with the permission of Dr. K. Yamamoto at the University of California, San Francisco. Recombinant androgen receptor (rAR) composed of the carboxyl-terminal half of AR (amino acids 460 through 902) containing the entire DNA and steroid binding domains was prepared from a bacterial expression system and purified as described before (Simental et al., 1991). Antibody against AR was raised to a synthetic peptide corresponding to the sequence spanning amino acids 527–541 which is located in the amino-terminal side flanking the sequence of the DNA binding domain, common to both human and rat ARs (Tan et al., 1988). The antibody was affinity-purified and successfully used as described in previous studies on androgen responsive genes (Tan et al., 1992). Anti-HNF-4 antibody was kindly provided by Dr. J. Darnell at The Rockefeller University. Preparation of human liver poly(A)$^+$ RNA was previously described (Salier et al., 1987). All other chemicals were of the highest quality commercially available.

Synthetic Oligonucleotides. Subsequence oligonucleotides employed in electrophoretic mobility shift assays (EMSAs) or polymerase chain reactions (PCRs) were synthesized by using an automated synthesizer (Applied Biosystems Model 380A) and purified by high-performance liquid chromatography at the Biomedical Research Core facility on this campus. The synthesized oligonucleotide subsequences with and without mutations used for EMSAs are listed in Table 1.

Preparation of Nuclear Protein Extracts. HepG2 cells were grown in DMEM as previously described (Salier et al., 1990; Hirosawa et al., 1990). Baby hamster kidney (BHK), African green monkey kidney (CV1), human prostate cancer (LNCaP), and human breast cancer (T47D) cells were grown in DMEM supplemented with 10% FBS, DMEM with 7% FBS, RPMI with 10% FBS, and DMEM with 10% FBS and 0.6 μg/mL insulin, respectively. All media were supplemented with penicillin–streptomycin.

Nuclear extracts (NEs) of cultured cells were prepared from about 1×10^8 cells by the method of Dignam et al. (1983) with slight modifications. Protein concentrations of NEs were quantitated by the Bradford method (Bio-Rad). NE preparations were frozen in aliquots at −80 °C until further use. To obtain NEs of cells treated with androgen, cells were cultured in media supplemented with 1 or 0.1 μM testosterone or dihydrotestosterone as stated, for 16 h prior to harvesting.

Rat liver NEs were prepared from 3–4-month-old male rats (Sprague-Dawley) according to Graves et al. (1986) with minor modifications.

DNase I Footprinting Analysis. Plasmids pCR1000FIX(−150/+29) and pCR1000FIX(+29/−150) were used for generating dsDNA fragments with a ^{32}P-labeled sense or antisense strand, respectively. These plasmids were constructed by subcloning DNA fragments, which were generated by PCR from the region spanning nt −150 to +29 of the FIX gene as the template, into pCR1000 plasmid in both orientations. These plasmids (20 μg of each) were cleaved at a unique *Eco*RI site in the vector sequence just outside its 3' end of the insert, and the newly-generated 3' termini were labeled in a fill-in reaction with [α-^{32}P]dATP and Klenow enzyme. The radiolabeled plasmid DNAs were then cleaved at a unique *Hin*dIII site located outside of the 5' end of the insert, producing fragments (236 bp in size) containing a sense or antisense strand with ^{32}P-labeled 3' ends. These fragments were separated by preparative agarose gel electrophoresis and recovered from gels by electroelution. The radiolabeled fragments with ~20 000 cpm were incubated on ice for 60 min with 100–200 μg of nuclear proteins prepared from HepG2 cells grown in the presence and absence of testosterone (100 nM) or from BHK cells, in a final volume of 50 μL of DNA binding buffer (10 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM DTT, 1 mM EDTA, 4% glycerol, and 100 μg/mL nuclease-free bovine serum albumin) containing 1 μg of poly-

(dI-dC). DNase I digestion and electrophoretic analysis of the digested products were carried out according to Jones et al. (1987) with minor modifications. Briefly, the DNA–protein complex mixtures were digested with 0.5 unit of DNase I in a total volume of 100 μL of DNase I buffer (10 mM $MgCl_2$ and 5 mM $CaCl_2$) for 3 min. The control DNA samples with no nuclear proteins were incubated with 0.02 unit of DNase I for 2 min. Digestion was stopped by adding DNase I stop buffer (200 mM NaCl, 20 mM EDTA, and 1% SDS), and the digested DNAs were extracted with a phenol–chloroform mixture, followed by precipitation with 70% ethanol. DNA fragments obtained were analyzed on a 6% polyacrylamide sequencing gel. Plasmid pCRII (−150/+29, −20:A), which contains the FIX sequence with a mutation (T to A) at nt −20, was constructed by ligating DNA fragments PCR-amplified from the FIX sequence with the mutation as a template (Hirosawa et al., 1990) into pCRII plasmid and was used for DNase I footprinting analysis in a similar manner as described above for the normal FIX gene sequence. All plasmids constructed in these experiments were subjected to sequence analyses before use to confirm their correct sequences. Each footprinting analysis was repeated 3–4 times to determine its conditions and reproducibility with separate NE preparations.

Electrophoretic Mobility Shift Assay (EMSA). Double-stranded (ds) oligonucleotides used in EMSAs were prepared by mixing equimolar amounts of sense and antisense strands of each corresponding synthetic oligonucleotide. These oligonucleotides were labeled at the 5′ ends with [^{32}P]ATP and T4 polynucleotide kinase to a specific activity of about 1×10^9 cpm/μg and used as probes in EMSAs as follows. Aliquots of nuclear proteins prepared from cultured cells or rat liver tissue were mixed with 1 μg of sonicated ds poly(dI-dC) in DNA binding buffer (same as used for footprinting) and added with ^{32}P-labeled ds oligonucleotides (0.05–0.1 ng with approximately 30 000 cpm) in a final volume of 20 μL. The reaction mixtures were incubated at room temperature for 20 min and electrophoresed on a nondenaturing 4% polyacrylamide gel in glycine buffer (50 mM Tris, 0.38 M glycine, and 2 mM EDTA, pH 8.5) for 4 h at 25 mA. Gels were then dried, and DNA–protein complexes were visualized as mobility-shifted bands by autoradiography. Bands were quantitated by scanning the dried gels on a Betascope 603 blot analyzer. All EMSAs were repeated at least twice with separate NE preparations to test the reproducibility.

Northern Blot Analysis. Northern blot analysis was carried out with poly(A)$^+$ RNA preparations of human liver as well as BHK cells which were previously described (Salier et al., 1987; Kusumoto et al., 1988). Briefly, poly(A)$^+$ RNAs of human liver (1 μg) and BHK cells (2.5 μg) were electrophoresed on a 1% agarose gel and blotted to a nylon membrane (GeneScreen Plus hybridization transfer membrane, NEN). The membrane was hybridized with the human FIX cDNA (Kurachi & Davie, 1982) which was ^{32}P-labeled by a random priming method to a specific activity of about 1×10^9 cpm/μg.

Primer Extension. Primer extension analysis was carried out as previously described (Salier et al., 1990). Synthetic oligonucleotide primer (30-mers), which contained a complementary strand FIX gene sequence spanning nt +71 to +51 attached with *Pst*I linker sequences, was radiolabeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. An aliquot (1×10^6 cpm) of the primer was mixed with 5 μg of human liver poly(A)$^+$ RNA. As a control, 10 μg of poly(A)$^+$ RNA prepared from BHK cells was used. Mixtures were then hybridized at 32 °C overnight, followed by primer extension reaction using unlabeled dNTPs and 50 units of AMV RTase at 42 °C for 90 min. The reaction mixture was treated with RNase H for 1 h at 37 °C and extracted with phenol–chloroform, and cDNA fragments generated were precipitated with 70% ethanol. The primer extension products were separated on a 6% acrylamide DNA sequencing gel and visualized by exposing to X-ray film.

Reverse Transcription and PCR (RT–PCR). Single-stranded cDNA templates for PCR were prepared by reverse transcription (RT) of poly(A)$^+$ RNA using a cDNA cycle kit from Invitrogen. Human liver poly(A)$^+$ RNA preparation (2 μg) was mixed with 0.2 μg of oligo $d(T)_{12-18}$ primer, 10 units of RNase inhibitor, 4 μL of 5 × RT buffer (0.5 M Tris, pH 8.3 at 42 °C, 0.2 M KCl, and 0.05 M $MgCl_2$), 1 μL of dNTPs (25 mM for each nucleotide), and 10 units of AMV RTase in a final volume of 20 μL and incubated at 42 °C for 1.5 h. The reaction volume was increased to 100 μL with 10 mM Tris-HCl, pH 7.5, containing 1 mM EDTA and extracted with phenol–chloroform, followed by precipitation with 70% ethanol. The first-strand cDNAs generated were resuspended and subjected to PCRs. Each PCR mixture contained 0.2 μM primers, 1.25 units of Taq polymerase, and dNTPs (200 μM for each nucleotide) in 50 μL of reaction buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, and 1.5 mM $MgCl_2$). Templates used for PCRs included the first-strand cDNA (200 ng) generated by reverse transcription of poly(A)$^+$ RNA without and with RNase A treatment, genomic DNA (200 ng) without and with RNase A treatment, or poly(A)$^+$ RNA similarly processed without adding RTase. PCR primer oligonucleotides used included the common 3′-side primer (complementary strand sequence) $^{+247}$CGCAGGTTGGTAAGTACTGGTTCTT^{+271} and one of the 5′-side primers (coding strand sequence) $^{-300}$CATTGCTCTCTGACAAAGATACGGTGGGTCC^{-270}, $^{-176}$ACAGTGGCAGAAGCCCACGAAATC^{-153}, $^{-150}$AGGTGAAATTTAATAATGACCAC^{-128}, $^{-40}$ATACAGCTCAGCTTGTACTTT^{-20}, or $^{+3}$CACTTTCACAATCTGCTAGC^{+22} (numbers indicate nucleotide positions). Amplifications were performed on a Perkin-Elmer Cetus thermocycler. The reaction mixtures were initially heated at 94 °C for 4 min, and then subjected to 25 cycles of sequential incubations at 94 °C for 30 s, 56 °C for 30 s, and 65 °C for 4 min, followed by incubation at 65 °C for 7 min. PCR products were analyzed by electrophoresis on a 1.8% agarose gel.

RESULTS

DNase I Footprinting Analyses. Nuclear proteins of HepG2 cells protected five distinct sequence areas within the LS-region and its neighboring region of the normal FIX gene (Figure 1A). Sense strand footprints are designated FP-I (nt +4 to +19), FP-II (nt −16 to −3), FP-III (nt −27 to −19), FP-IV (nt −67 to −49), and FP-V (nt −99 to −77). Footprints on the antisense strand were very similar and showed only minor differences. The overall footprint patterns obtained with NEs of HepG2 cells grown in the presence and absence of testosterone were essentially the same, but slightly clearer footprints for NEs prepared from cells treated with testosterone were observed. Proteins which bind to FP-I and FP-V regions were previously reported as C/EBP and NF-1, respectively (Crossley & Brownlee, 1990). The protein which binds to the FP-III region was recently identified as HNF-4 (LFA-1) (Reijnen et al., 1992; Crossley et al., 1992). Nuclear protein binding at FP-II and FP-III regions which contain FIX–Leyden mutations at nt −20 and −6 was previously reported

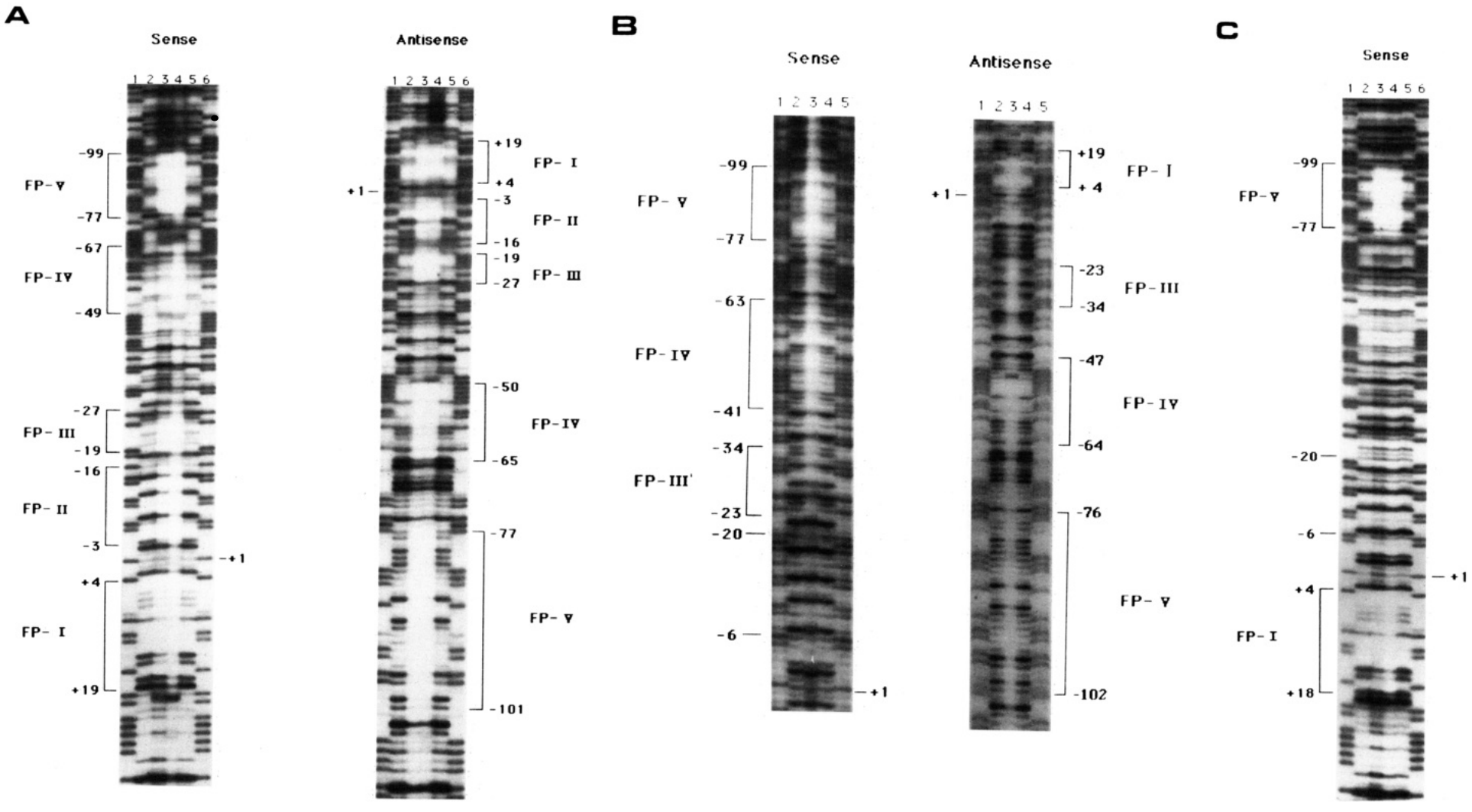

FIGURE 1: DNase I footprinting analyses of the LS-region and its neighboring region of the FIX gene. NE preparations from HepG2 cells as well as BHK cells were used as described in the text. (A) 5′-end region of the normal FIX gene spanning nt −150 to +29. DNA fragments ^{32}P-labeled at the 3′ end of the sense or antisense strand were subjected to DNase I digestion in the absence and presence of 100- or 150-μg NEs of HepG2 cells. For both sense and antisense strand analyses: lanes 1 and 6, purine-specific (A+G) sequence ladders; lanes 2 and 5, without NEs; lanes 3 and 4, with 100- and 150-μg NEs of HepG2 cells, respectively. (B) 5′-end region spanning nt −150 to +29 of a FIX gene with a mutation (T to A) at nt −20 (sense or antisense strand labeled with ^{32}P). For both sense and antisense strands: lanes 1 and 5, purine-specific (A+G) sequence ladders; lanes 2 and 4, without NEs; lane 3, with 150-μg NEs of HepG2 cells grown in the presence of 100 nM testosterone. (C) Same as (A) except 150 and 200 μg of BHK cell NEs were used for lanes 3 and 4, respectively.

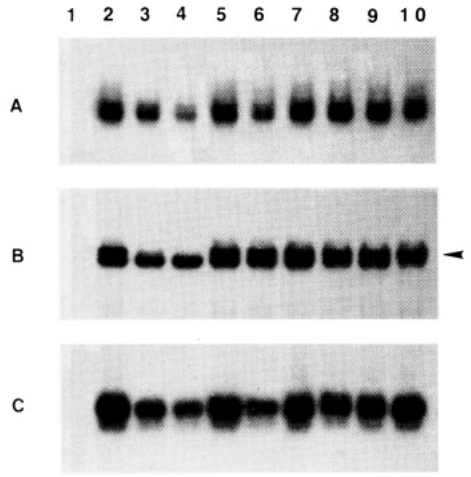

FIGURE 2: EMSA analyses of protein binding at the LS-region. ^{32}P-Labeled ds –20sub (A), –6sub (B), and +13sub (C) were each incubated with 8 μg of HepG2 cell NEs, and DNA–protein complexes formed were analyzed by EMSAs as described in the text. For competition, unlabeled ds oligonucleotide subsequences in molar excess of 50- or 150-fold over that of the labeled oligonucleotides were added to NEs, and the mixtures were incubated at room temperature for 5 min prior to the addition of the labeled probe. In all panels (A–C): lane 1, without NEs; lanes 2–10, with HepG2 cell NEs. (A) EMSAs of ^{32}P-labeled –20sub: lane 2, without competing subsequence; lanes 3 and 4, with 50- and 150-fold unlabeled –20sub (self-competition); lanes 5 and 6, with –20Msub(–20:A); lanes 7 and 8, with –6sub; lanes 9 and 10, with +13sub. (B) EMSAs of ^{32}P-labeled –6sub: lane 2, without competing subsequence; lanes 3 and 4, with 50- and 150-fold unlabeled –6sub (self-competition); lanes 5 and 6, with –6Msub(–6:A); lanes 7 and 8, with –20sub; lanes 9 and 10, with +13sub. The arrow indicates a mobility-shifted band specific to the oligonucleotide sequence. The lower band corresponds to a shifted band due to nonspecific protein binding which was absent when a shorter subsequence spanning nt –13 to +3 was used (data not shown). The shorter subsequence also showed specific protein binding, but with about 5-fold less affinity. (C) EMSAs of ^{32}P-labeled +13sub: lane 2, without competing subsequence; lanes 3 and 4, with 50- and 150-fold unlabeled +13sub (self-competition); lanes 5 and 6, with +13Msub(+13:G); lanes 7 and 8, with –20sub; lanes 9 and 10, with –6sub.

as abstract forms in our studies (Kurachi et al., 1991a,b). FP-IV is a new footprint detected in the present study.

NEs prepared from BHK cells gave only two footprints in the same regions: FP-I, corresponding to C/EBP binding, and FP-V, corresponding to NF-1 binding (Figure 1C). This indicates that footprints FP-II, -III, and -IV are HepG2 cell-specific footprints and are not due to experimental artifacts.

A FIX–Leyden gene sequence with a mutation at nt –20 (T changed to A) showed grossly different footprint patterns on both sense and antisense strands from those of the normal gene in the LS-region (Figure 1B). In the presence of the mutation at nt –20, both FP-II and FP-III were absent, but instead, a new footprint, FP-III′, spanning nt –34 to –23, appeared. The loss of FP-II due to the mutation at nt –20 suggested that the binding of an unidentified protein to FP-II may require the presence of HNF-4 at FP-III for its stable binding.

Protein Binding Analysis of the LS-Region by EMSAs. Synthetic oligonucleotides used in EMSAs (Table 1) are designed to the footprinted areas observed for the normal FIX gene and FIX–Leyden phenotype gene with a T to A mutation at nt –20. With NEs of HepG2 cells, all three ds subsequence oligonucleotides, –20sub, –6sub, and +13sub which harbor sequence areas of FP-III, FP-II, and FP-I, respectively, bound nuclear proteins which were visualized as bands with retarded mobility (Figure 2). These proteins are unrelated to each other (lanes 7–10 of panels A, B, and C).

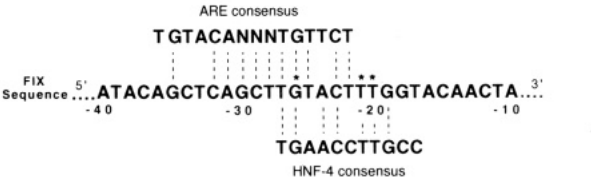

FIGURE 3: LS-region with AR and HNF-4 binding consensus sequences. The 5′-half portion of the LS-region is shown in comparison with consensus sequences of ARE (above) and HNF-4 (below) (Klock et al., 1987; Miao et al., 1992). Matched sequences between the FIX sequence and the consensus sequences are shown by dashed lines. Positions of mutations found in this region of the FIX genes are shown with asterisks. The mutation at nt –20 or –21 results in a Leyden phenotype, while the mutation at nt –26 does not (Reijnen et al., 1992; Crossley et al., 1992).

When these subsequences were tested with NEs of BHK cells, nondetectable or only extremely low levels of protein binding were observed for –20sub or –6sub, while a protein, presumably C/EBP (Crossley & Brownlee, 1990), which was present at an ample level in nuclei of BHK cells, strongly bound to +13sub (data not shown). This was in good agreement with the results of footprinting analysis, as described above. Protein binding affinities of mutant subsequence oligonucleotides, –20Msub(–20:A), –6Msub(–6:A), and +13Msub(+13:G), which contain Leyden phenotype single-base mutations at nt –20, –6, or +13, respectively, were substantially lowered as shown by EMSAs (Figure 2, lanes 5 and 6). Negative effects of the specific mutations on nuclear protein binding agreed well with the grossly reduced *in vivo* and *in vitro* expression of the FIX–Leyden genes containing these mutations (Reitsma et al., 1988, 1989; Crossley et al., 1990, 1992; Hirosawa et al., 1990; Picketts et al., 1993).

The nucleotide sequence of the FP-III region is similar to the HNF-4 binding site (Figure 3) (Sladek et al., 1990) and can bind HNF-4 (Reijnen et al., 1992; Crossley et al., 1992; S. Kurachi, unpublished data). In spite of the similarity in its left half of the FP-III sequence to the glucocorticoid responsive element (GRE), –20sub did not bind rGR, agreeing well with our previous observation that glucocorticoid showed no significant effects on the expression of the FIX gene *in vitro* (Hirosawa et al., 1990). AR also did not bind to –20sub (data not shown).

The region spanning nt –34 to –23, which includes FP-III′, has a substantial similarity to ARE and GRE (Figure 3). Protein binding to –32sub which spans nt –40 to –20 harboring FP-III′ was strongly competed by the unlabeled –32sub (self) (Figure 4A, lanes 3 and 4) as expected, but only very weakly competed by the mutant subsequence –32Msub(–26:C) which contains a G to C mutation at nt –26 in the partially overlapped area of FP-III and FP-III′ (lanes 5 and 6). The protein (HNF-4) binding to –20sub was much less effectively competed against with –32sub (lanes 7 and 8 of Figure 4B) compared to the self-competition by –20sub (lanes 3 and 4). These observations agreed well with those obtained from DNase I footprinting analyses of this region (Figure 1), which showed dominant protein binding at FP-III (presumably HNF-4) of the normal FIX gene over that of protein binding at the partially overlapped region, FP-III′ (Figure 1A), and the loss of its dominance, allowing protein binding to FP-III′ of a Leyden phenotype gene with a mutation at nt –20 (T to A change) (Figure 1B).

Subsequence –32sub could bind purified rAR only very weakly with less than 5% affinity compared to C3 ARE, a *bona fide* ARE present in rat C3 gene which was used as a positive ARE control (Tan et al., 1992) (Figure 5, lanes 1–8; lanes 7 and 8 represent an overexposure of lanes 5 and 6 to

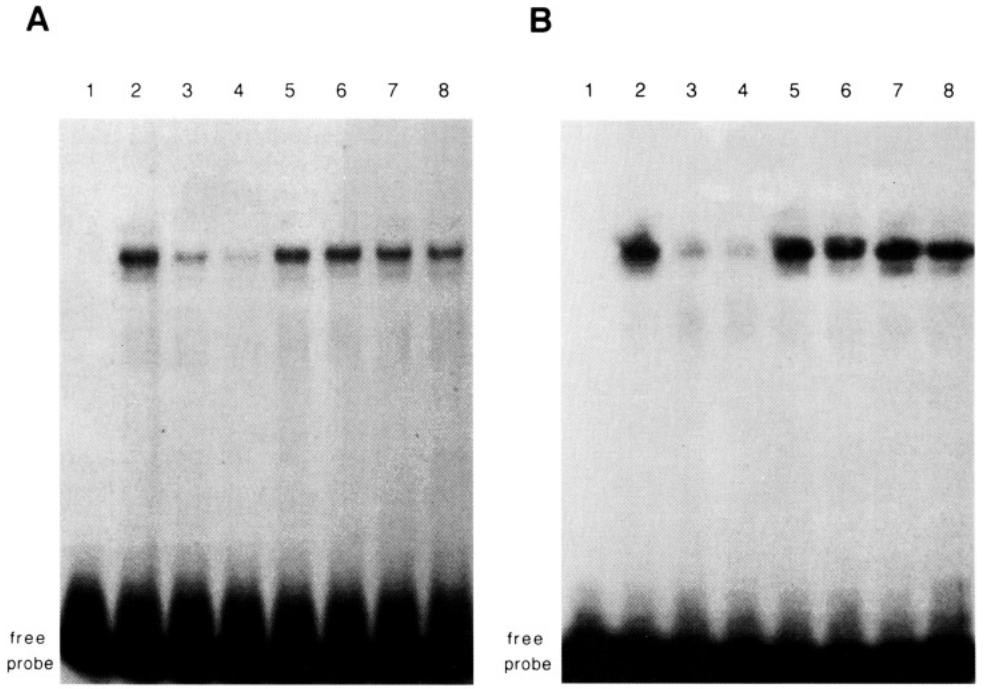

FIGURE 4: Effects of mutation at nt −26 on protein binding to −32sub and −20 sub. EMSAs were performed as described in the text with 8 μg of HepG2 cell NEs without and with 100- or 300-fold molar excess unlabeled subsequences. (A) EMSAs of ^{32}P-labeled −32sub: lane 1, without NEs; lane 2, with NEs alone; lanes 3 and 4, with NEs and unlabeled −32sub (self-competition); lanes 5 and 6, with NEs and unlabeled −32Msub(−26:C); lanes 7 and 8, with NEs and unlabeled −20sub. (B) EMSAs of ^{32}P-labeled −20sub: lane 1, without NEs; lane 2, with NEs alone; lanes 3 and 4, with NEs and unlabeled −20sub (self-competition); lanes 5 and 6, with NEs and unlabeled −20Msub(−26:A); lanes 7 and 8, with NEs and unlabeled −32sub.

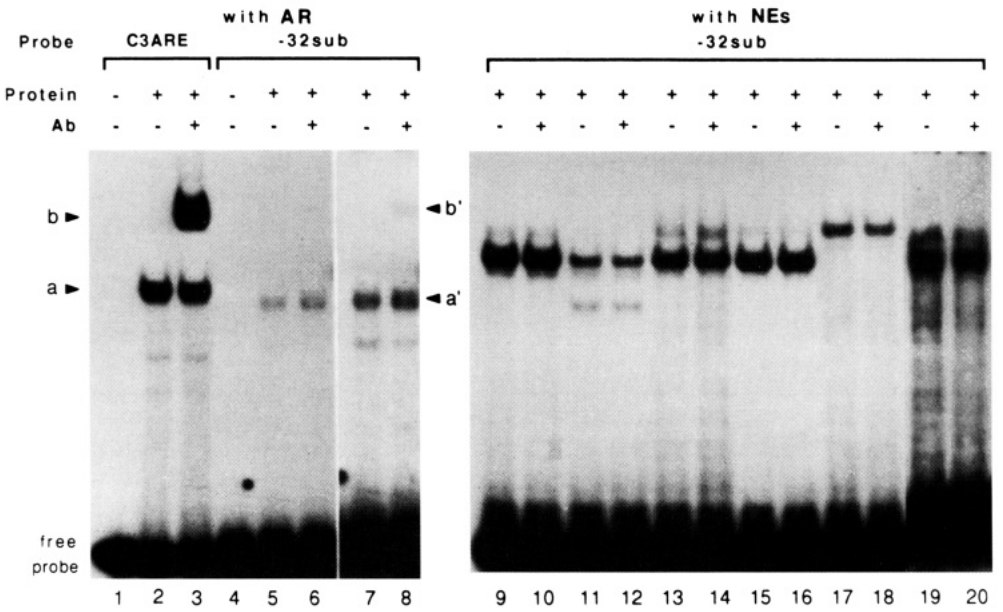

FIGURE 5: DNA binding of AR and nuclear proteins. EMSAs of ^{32}P-labeled C3 ARE (specific activity of 1.2×10^8 cpm/μg) and ^{32}P-labeled −32sub (specific activity of 2.3×10^8 cpm/μg) with rAR were carried out in the presence and absence of anti-AR antibody. All reaction mixtures contained 500 nM dihydrotestosterone. See the text for detailed experimental conditions. Lane 1, C3 ARE probe only; lanes 2 and 3, with 1 μg of rAR, without and with anti-AR antibody, respectively; lane 4, −32sub probe only; lanes 5 and 6, with 1 μg of rAR, without and with anti-AR antibody, respectively; lanes 7 and 8, overexposure of lanes 5 and 6: arrows a and a′ indicate the shifted bands of DNA–protein complexes of C3 ARE and −32sub, respectively; arrows b and b′ indicate supershifted bands of immune complexes of C3 ARE and −32sub (visible upon longer exposure), respectively. Lanes 9–20, EMSAs of ^{32}P-labeled −32sub with 8 μg NEs of AR-positive and -negative cells: lanes 9 and 10, HepG2 cells; lanes 11 and 12, T47D cells; lanes 13 and 14, LNCaP cells; lanes 15 and 16, CV1 cells; lanes 17 and 18, BHK cells; lanes 19 and 20, rat liver NEs. Lanes 10, 12, 14, 16, 18, and 20 contain anti-AR antibodies.

show band b′ in lane 8). The mobility-shifted bands in EMSAs corresponding to complex formation of C3 ARE or −32sub with rAR were further shifted with anti-AR antibodies, as shown by bands b and b′ in Figure 5. Furthermore, subsequence −32sub reproducibly bound nuclear proteins of various cells including AR-positive cells such as HepG2, T47D, and LNCaP, as well as AR-negative cells such as CV1 and BHK cells, strongly indicating that this protein is not AR, but a more abundant nuclear protein (lanes 9–18). NEs prepared from adult male rat liver also contain a protein equivalent to that present in HepG2 cells (lanes 19 and 20). None of these proteins in various cells and liver were recognized by anti-AR

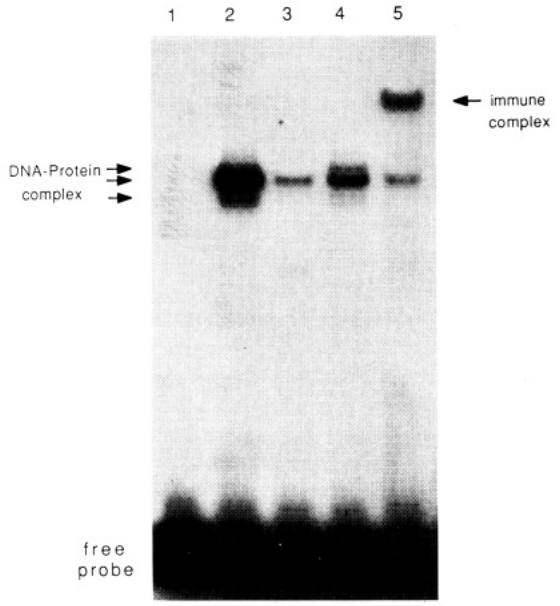

FIGURE 6: Interactions between AR and the protein which binds to –32sub. Experimental conditions were the same as for Figure 5. Recombinant AR (1 μg) and HepG2 cell NEs (10 μg) were incubated for 10 min at room temperature before adding the ^{32}P-labeled –32sub probe. The –32sub probe was then added and incubated for 20 min at room temperature, followed by the addition of 1 μg of anti-AR antibody and 30-min incubation on ice. The reaction mixture was then subjected to electrophoresis. Lane 1, without NEs; lane 2, with NEs alone; lane 3, with NEs premixed with rAR for 10 min; lane 4, with NEs added with anti-AR antibody only; lane 5, with NEs premixed with rAR and added with anti-AR antibody. DNA–protein and supershifted DNA–protein–antibody complexes are indicated by arrows.

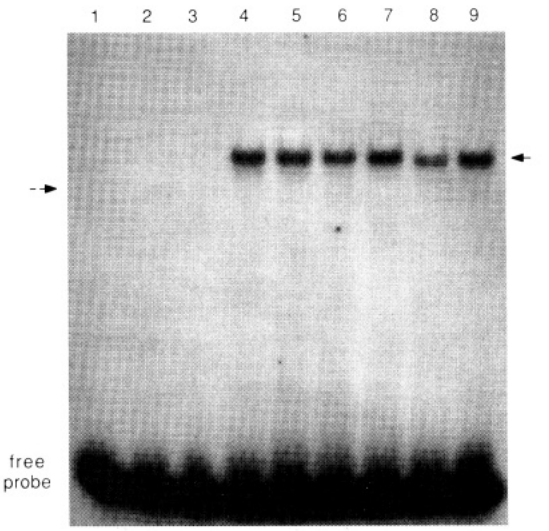

FIGURE 7: Characterization of protein binding to –60sub by EMSAs. Radiolabeled ds –60sub was incubated with AR, GR, and NEs from various cultured cells and electrophoresed on a nondenaturing polyacrylamide gel. Lane 1, without NEs or protein NEs; lane 2, with rAR; lane 3, with rGR; lanes 4 and 5, with HepG2 cell NEs; lanes 6 and 7, with LNCap cell NEs; lanes 8 and 9, with CV1 cell NEs. Lanes 5, 7, and 9 contain anti-AR antibodies. The arrow on the right indicates the shifted bands of DNA–protein complexes formed with nuclear proteins of various cells. The arrow on the left indicates a faint band of DNA–rAR complex in lane 2.

antibody, in further support of the above observation (lanes 10, 12, 14, 16, 18, and 20). The binding of AR to –32sub agrees with the recent observation by Crossley et al. (1992) that AR–protein A fusion protein can bind to the region containing ARE-like sequence. Crossley et al., however, failed to show that the nuclear protein binding to the region is not AR (see Discussion).

When aliquots of purified rAR were premixed with NEs prior to the addition of radiolabeled –32sub probe, the intensity of the shifted band greatly decreased (Figure 6, lane 3), suggesting a possible interaction between rAR and the nuclear protein which binds to –32sub. When –32sub probe and NEs were preincubated prior to the addition of rAR, the intensity of the shifted band (complex of –32sub and a nuclear protein) was changed little from that observed without rAR. No new supershifted band(s) appeared. These results suggest that rAR interacts with the nuclear protein, but once the nuclear protein binds to the DNA, rAR can much less effectively bind to the protein. Upon addition of anti-AR antibody to the reaction mixture of –32sub probe in the presence of rAR, a supershifted band appeared (Figure 6, lane 5). As judged from the band position, this corresponds to the –32sub·rAR·anti-AR complex which was formed only in a large excess of rAR. This shift did not occur when no rAR was added to the reaction mixture (lane 4). These results further support the above observation that the nuclear protein which primarily binds to –32sub is not AR, but a nuclear protein which can interact with AR. Recombinant AR showed no such interactions with HNF-4 or other proteins which bind various other regions of the LS-region (data not shown).

A new footprinted region, FP-IV, has a sequence similarity to ARE in reverse direction of the sequence spanning nt –49 to –65 (5′-GGGAAATAATGTCCATC-3′; the underlined sequences are identical to ARE). The nuclear protein bound to this region was confirmed by EMSAs using –60sub spanning nt –65 to –47 (Figure 7, lanes 4–9). This was neither AR nor GR (lanes 2 and 3). Recombinant AR could bind to –60sub, but extremely weakly only in a purified system with an excess amount of AR (lane 2). Anti-AR antibody did not recognize the nuclear protein bound to –60sub (lanes 5, 7, and 9), further confirming that this protein is not AR. The oligonucleotide –60sub showed no significant competition with –32sub, suggesting that the protein which binds at FP-IV is different from those binding to –32sub (data not shown).

Testosterone Effects on Androgen Receptor-Positive and -Negative Cells. Effects of testosterone on AR-positive cells (HepG2, LNCaP, and T47D cells) and on AR-negative cells (CV1 and BHK cells) were tested with respect to protein binding at the LS-region. In order to carry out these experiments properly, the effects of endogenous testosterone in FBS were tested. Cells cultured in medium supplemented with 10% charcoal-treated FBS, which contains only as low as 0.52 nM testosterone, consistently showed much less response to exogenous testosterone than cells cultured with medium supplemented with 10% regular (non-charcoal-treated) FBS. This diminished response could result from an extensive loss of other serum factors required for normal cell growth and function. The regular FBS at 10% concentration in culture media contains only 1.39 nM testosterone (1.2% of the normal level, 117.4 nM, of adult human males). Accordingly, all the following experiments were carried out with cells cultured in medium supplemented with 10% regular FBS. This experimental condition was found appropriate to observe specific effects of testosterone without any significant background problems. When HepG2 cells were cultured in the presence of testosterone (100 nM), –32sub and –20sub reproducibly increased their binding to nuclear proteins by (3.7 ± 0.3)-fold and (1.75 ± 0.15)-fold, respectively (averages of three experiments) (Figure 8, lanes 8 and 4). Upon cell treatment with testosterone, protein binding to –6sub did not change significantly, while that to +13sub increased 2-fold.

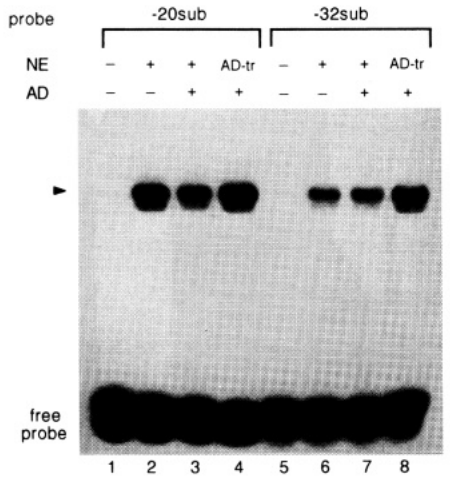

FIGURE 8: Effects of androgen on protein binding to –20sub and –32sub. The ^{32}P-labeled –20sub or –32sub was incubated with NEs of HepG2 cells which were cultured in the presence or absence of 100 nM testosterone (AD-tr). Reaction mixtures were also supplemented with or without 500 nM dihydrotestosterone (AD). DNA–protein complexes were analyzed by gel electrophoresis as described in the text. Lanes 1–4 and lanes 5–8 are for ^{32}P-labeled –20sub probe and ^{32}P-labeled -32sub probe, respectively. Lanes 1 and 5, without NEs; lanes 2 and 6, with NEs prepared from cells cultured in medium without testosterone; lanes 3 and 7, same as in lanes 2 and 6 except AD added to the reaction mixture; lanes 4 and 8, with NEs prepared from cells cultured in the presence of 100 nM testosterone.

Treatment of cells with 1 μM testosterone gave essentially the same results as with 100 nM. Direct addition of dihydrotestosterone (500 nM) to NEs prepared from HepG2 cells which were cultured without testosterone treatment gave only a marginal increase (1.2-fold at most) of the nuclear protein binding to –32sub in EMSAs (Figure 8, lane 7). This indicated that treatment of cells with testosterone is important to observe its effects. NEs prepared from androgen receptor-negative cell lines such as CV1 and BHK did not show any response upon similar testosterone treatment, confirming the specificity of the testosterone effects observed.

Transcriptional Initiation Site of Human FIX Gene in Liver. The preparation of human liver poly(A)$^+$ RNA used in the present study contained a high-quality FIX mRNA with a single discrete band of 3.2 kb with little apparent degradation (Figure 9A). The major transcription initiation site for the human FIX gene determined by primer extension analyses with the RNA preparation was found at nt –176 (Figure 9B), which is much further 5′ upstream to the originally reported site (nt+1) (Anson et al., 1984). The presence of the 5′-upstream major transcription initiation site(s) was further confirmed by reverse transcription-mediated PCR (RT-PCR) analyses of human liver poly(A)$^+$ RNAs (Figure 10). RT-PCR analyses unambiguously showed RT-PCR amplifications with primers designed to various sites located 3′ downstream to nt –176, but not with the primer designed to the nt –300 region, where no obvious initiation signals were detected in the primer extension analysis (Figure 10A, lanes 2–6). Control PCRs with the genomic DNA templates showed good amplifications for all primers used, including that of the –300 region (lanes 7–11). When the RNA preparation was treated with RNase A prior to the reverse transcription in these experiments, no PCR amplifications were observed, eliminating possible genomic DNA contamination in the RNA preparation used (Figure 10B, lane 2). The primary initiation site at 5′ upstream in the neighborhood of nt –176 was also confirmed by RNase A protection assays using an RNA preparation from liver cells of a transgenic mouse carrying a human FIX transgene (data not shown). These results were consistant with the size (3.2 kb) of human liver FIX mRNA (Figure 9A) and the initiation site observed for CAT constructs as well as with the reported initiation sites for dog and mouse FIX genes (Salier et al., 1990; Yao et al., 1991; Evans et al., 1989). Primer extended bands were also found at nt –1 and its neighborhood, in addition to minor bands at nt –214, –87, –45, and –10 regions. These bands may represent true alternative initiation sites, or artifacts in analyses possibly due to unique secondary structures of the RNA sequence in these regions or RNA degradations.

DISCUSSION

The basic mechanism of action of the LS-region, where all Leyden phenotype mutations thus far known are contained, in the regulation of the FIX–Leyden gene was systematically studied in relation to DNA–protein interactions and the transcriptional initiation site of the FIX gene in human liver.

Footprinting analyses of the LS-region and its immediate neighboring region of the normal FIX gene show at least five protein binding sites in the region spanning nt –100 to +29 where the LS-region is contained (Figure 1A). A schematic drawing summarizing these protein bindings is shown in Figure 11. Footprinted areas FP-I, FP-II, and FP-III in the LS-region contain all 11 unique mutations found in 16 FIX–Leyden genes known to date (Thompson, 1991; Giannelli et al., 1992; Kurachi et al., 1992). The importance of the normal sequences of these footprinted areas in the LS-region is now well recognized in relation to FIX gene expression and specific protein binding (Hirosawa et al., 1990; Kurachi et al., 1991a; Picketts et al., 1993; Crossley et al., 1992) (Figure 2). In spite of this amount of information, the precise molecular mechanism underlying the action of the LS-region has been elusive, and the reported data have often contradicted to each others, suggesting a need for highly systematic analyses on this region.

As shown in Figure 1A, FP-III, which presumably corresponds to HNF-4 binding as shown by EMSAs (Reijnen et al., 1992; Crossley et al., 1992), spans nt –27 to –19 on the sense and antisense strands of the normal FIX gene sequence, and does not extend beyond nt –27 on its 5′ side (Figure 1A). The presence of a mutation at nt –20 (T changed to A) in a FIX–Leyden gene, however, caused a loss of FP-III (HNF-4 binding) and a new appearance of FP-III′ (Figure 1B). This indicates that HNF-4 binds to FP-III much more strongly, or is present much more abundantly in the nucleus, than the protein that binds to FP-III′ which partially overlaps with FP-III (Figure 11). This actually agreed very well with the results of EMSAs of –20Msub(–20:A) which also showed a gross reduction in protein binding to the FP-III region (Figure 2A). These results demonstrate for the first time the actual competition in protein binding between FP-III and FP-III′ regions, and the dramatic effect of a single-base mutation at nt –20 on the competition. This mutation-dependent protein binding at FP-III′ may play a key role in the mechanism underlying the amelioration of hemophilia B with the mutation at nt –20 and, perhaps, with some other mutations in the neighborhood such as at nt –6 as discussed below. Unlike other mutations known in the LS-region, a mutation at nt –26, which is located in the partially overlapped region of FP-III and FP-III′, results in a non-Leyden phenotype hemophilia B (Reijnen et al., 1992; Crossley et al., 1992). The reduced protein binding affinity observed, not only to –20Msub(–26:C) but also to –32M(–26:C) (Figure 4), further confirms the importance of protein binding at FP-III′ in

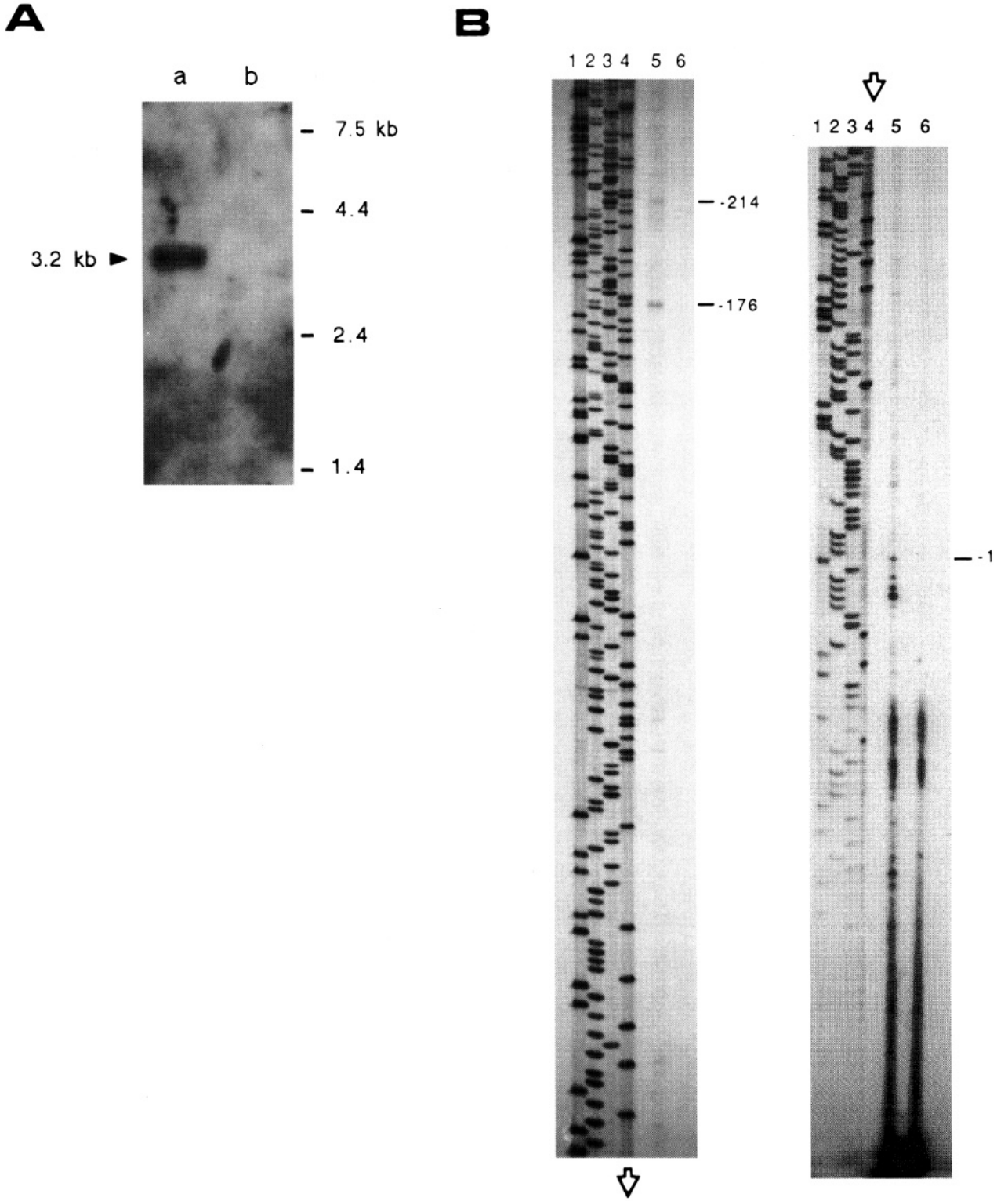

FIGURE 9: Transcriptional start site of the FIX gene in human liver. (A) Northern blot analysis of human liver poly(A)$^+$ RNAs. Human FIX cDNA was used as the ^{32}P-labeled probe. Lane a, poly(A)$^+$ RNA of human liver; lane b, poly(A)$^+$ RNA of BHK cells. An arrow with the size in kb indicates the position of the FIX mRNA band. Positions of RNA size markers (kb) are shown on the right side. (B) Primer extension analyses of human liver poly(A)$^+$ RNA: lanes 1–4, DNA sequencing ladders as size markers; lane 5, primer extension products of poly(A)$^+$ RNA of human liver; lane 6, primer extension products of poly(A)$^+$ RNA of BHK cells. Open arrows indicate the continuation of columns. Positions of prominent extension products are shown by nt number on the right side.

amelioration of the hemophilia B-Leyden phenotype with a mutation at nt −20, a defect which disrupts the required protein binding at FP-III. Crossley et al. (1992) showed by EMSAs using a long oligonucleotide (nt −40 to −9) that mutations at nt −20 and −26 are detrimental to binding of a protein (HNF-4) to the nt −20 region and that recombinant AR–protein A fusion protein can bind to the partially overlapped ARE-like sequence only in the absence of the mutation at nt −26. They concluded that direct binding of AR at the ARE-like sequence is essential for the amelioration of the Leyden phenotype with a mutation at nt −20. In their study, however, no footprinting and/or other analyses using nuclear extracts were carried out to show any direct evidence of mutation-dependent competitive protein binding at the partially overlapped sites.

In the present study, anti-AR antibodies which are capable of recognizing both free and DNA-bound ARs did not

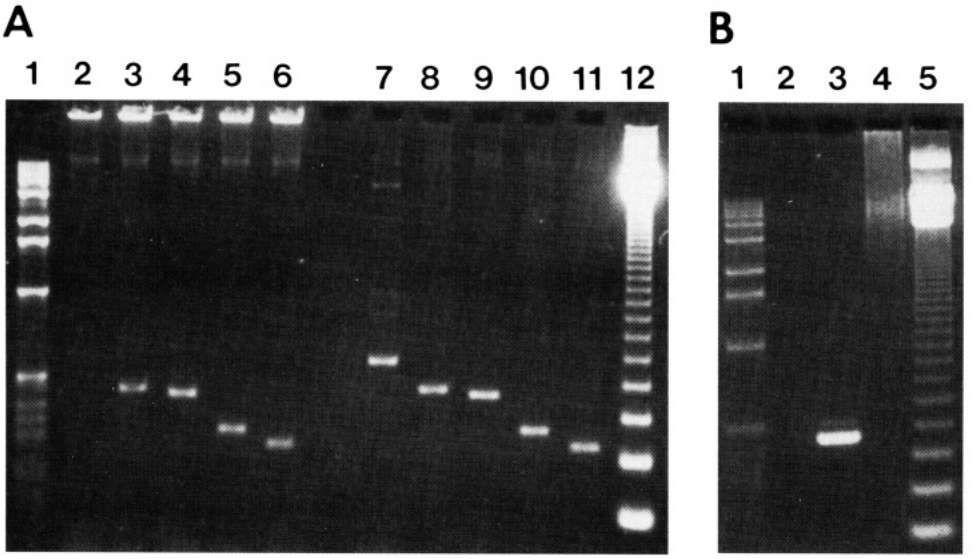

FIGURE 10: RT–PCR analyses of human liver poly(A)$^+$ RNA. (A) Lanes 1 and 12, size markers (*Hin*dIII fragments of λ phage DNA and 123 bp ladders, respectively); lanes 2–6, RT–PCR products generated from human liver poly(A)$^+$ RNAs with primers designed to nt sequences −300 to −270, −176 to −153, −150 to −128, −40 to −20, and +3 to +22, respectively; lanes 7–11, PCR products generated from the human genomic DNA as template with the same set of primers used for lanes 2–6. (B) Controls to (A): lanes 1 and 5, size markers; lane 2, RT–PCR product of poly(A)$^+$ RNA treated with RNase A (primer sequence, nt −150 to −128); lane 3, same as lane 2 except genomic DNA treated with RNase A was used as a template; lane 4, same as lane 4 in (A) except no RTase included.

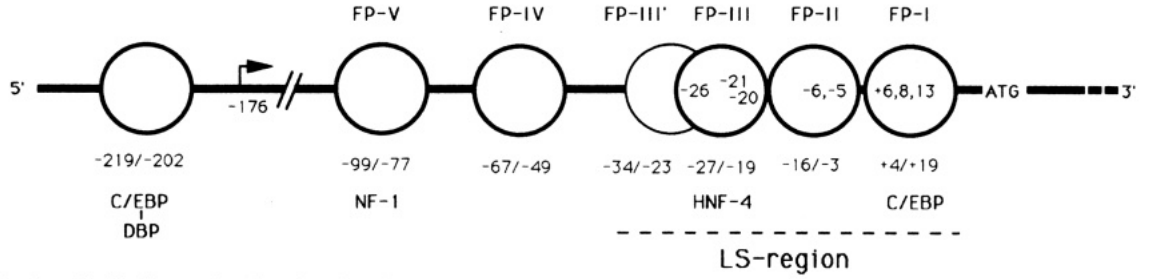

FIGURE 11: Protein binding to the 5′-end region of the human FIX gene. Horizontal solid bars indicate the FIX gene sequence. The arrow indicates the revised transcriptional start site (nt −176). Circles represent nuclear proteins bound at footprinted areas described in the text. Numbers inside circles indicate the locations of natural mutations so far found. Numbers below circles indicate the most 5′ and 3′ nt positions of each footprint. Binding proteins identified to date are also listed. The LS-region arbitrarily defined is shown by a dashed line. Protein binding at the region spanning nt −219 to −202 is based on the data reported by Picketts et al. (1993).

recognize the protein which binds at FP-III′ (Figure 5). Furthermore, this protein was found widely present in various types of cells, including both AR-positive and AR-negative cells as well as rat liver tissues, clearly indicating that the nuclear protein which binds at FP-III′ is not AR. This is consistent with the presence of a few differences in key nucleotide sequences between the FP-III′ region and the typical ARE (Figure 3), which is responsible for the dismal binding of rAR to −32sub (only ≤5% that of C3 ARE, a *bona fide* ARE control) even in the presence of a high amount of rAR (Figure 5, lanes 1–6). Binding of this protein to the area was significantly elevated when NEs of HepG2 cells were cultured in the presence of testosterone, suggesting that testosterone can affect directly or indirectly its nuclear concentration (Figure 8). Interestingly, the protein which binds to FP-III′ apparently interacts with AR without requiring any DNA mediator in the present experimental conditions, suggesting that AR exerts its action through a protein–protein interaction with the protein which binds at FP-III′, but not through a direct DNA–protein interaction in the LS-region (Figure 6). A possibility, however, that AR may confer its activity through its binding at an unidentified site which is quite distant from the LS-region is yet to be tested. These observations are contradictory to the report by Crossley et al. (1992), who claimed the direct AR binding at the ARE-like sequence to be responsible for the amelioration of hemophilia B-Leyden with a mutation at nt −20. They mistakenly reached the conclusion, probably due to multiple reasons, including no appropriate use of their positive ARE control (rat prostatic binding protein ARE) to assess the relative binding affinity of AR to the region with and without NEs as well as of anti-AR antibody to positively identify that the nuclear protein which actually binds to the region is not AR.

After the onset of puberty, hemophilia B-Leyden is ameliorated rather gradually over several years in virtual parallel to the increase of the systemic testosterone level and plasma factor IX which eventually reaches subnormal adult levels (somewhere within about 30–80% of normal in different families). This suggests that the induction of FIX–Leyden gene expression is dependent on the level of available testosterone which may, in turn, be reflected to a gradual increase of nuclear AR or other unidentified proteins. The amelioration of hemophilia B-Leyden with a mutation at nt −20, therefore, may be due to a combination of (i) a mutation-dependent binding of a non-androgen receptor nuclear protein at FP-III′ which is able to substitute for the function of the protein (HNF-4) binding to the intact FP-III site of the normal gene, (ii) a significant increase in the nuclear concentration of this protein upon testosterone surge, resulting in an elevated interaction with nuclear AR, which may be required for the optimal activity of this protein, and possibly (iii) modest, but

collectively significant, increases of HNF-4 and C/EBP binding to FP-II and FP-I, respectively.

A possible interaction between HNF-4 which binds at FP-III and an unidentified protein which binds to FP-II was suggested by the present footprinting results (Figure 1B). This interaction may be required for the optimized binding of HNF-4 to the DNA sequence or for inducing a subtle conformational change in HNF-4 which maximizes enhancer activity. This protein with such a function, however, may not be as crucial as HNF-4 for FIX gene expression, explaining in part why hemophilia B-Leyden patients with mutations at nt −6 or −5 in the FP-II region show mild bleeding disorders with the prepubertal plasma FIX level in a range of 5–10% of normal (Hirosawa et al., 1990; Picketts et al., 1991). The negative effects of these Leyden phenotype mutations on the *in vitro* expression of the genes (Hirosawa et al., 1990) are consistent with the lowered protein binding affinity of the mutant sequences of this region (Figure 2B).

The observation made in the present study agrees with an emerging concept that steroid hormone responsive elements, including ARE, exert their effects on genes via a complex mechanism in the context of multiple short- and long-range protein–protein and DNA–protein interactions (Alder et al., 1993; Yoshinaga et al., 1992).

By using multiple different approaches such as primer extension and RT–PCR, we also have established the major transcriptional start site of the FIX gene in human liver to be at nt −176 (Figures 9 and 10). This location is much further 5′ upstream than the originally assigned site (nt +1), and is consistent with those for dog and mouse FIX genes as well as that found for FIX–CAT constructs (Evans et al., 1989; Yao et al., 1991; Salier et al., 1990). The revised start site is also consistent with our previous observations that any extended deletions of the sequence beyond the nt −207 region in the 5′ to 3′ direction, or any deletions beyond nt −3 in the 3′ to 5′ direction, result in a gross reduction of the expression activity to 10–20% of the nondeleted sequence (Salier et al., 1990; Hirosawa et al., 1990). The nt +1 site may represent one of the alternative initiation sites or an artifact due to a specific secondary structure of mRNA in this region, or degradation products of the RNA preparation used. With the revised transcription initiation site, the entire LS-region is now located within the 5′-untranslated region. This raises an important issue, whether the LS-region mediates its action as dsDNA or as RNA transcripts. Effects of mutations in the LS-region on FIX expression and all the characteristics of proteins (mostly well-characterized enhancers) which bind to this region are consistent with an assumption that the LS-region confers its role as dsDNA. Enhancers such as HNF-4, C/EBP, and NF-1 which bind this region may function through their interactions with dsDNA in this region (Reijnen et al., 1992; Crossley et al., 1992; Crossley & Brownlee, 1990; Picketts et al., 1993). The LS-region apparently lacks any extensive unique structural features such as hairpin loops (Hirosawa et al., 1990) which are often involved in protein binding as shown for various elements in transcripts such as the Tar element of HIV virus (Sharp & Marciniak, 1989) and the iron responsive element of ferritin (Theil, 1990). Single-stranded subsequence oligonucleotides (about 20 bases in size) of the LS-region used in this study do not bind any nuclear proteins, while a much longer ss oligonucleotide spanning nt −40 to −1 binds some nuclear proteins (data not shown). Binding of these proteins is not affected by the presence of Leyden phenotype mutations at nt −6 or nt −20, and is not competed out by the ds subsequences −6sub or −20sub, indicating that these proteins are different from those which bind to ds subsequences. EMSAs, using RNA transcripts of the LS-region, in addition to the determination of stability (half-life) and steady-state concentration levels of FIX–Leyden mRNAs in cells, however, are required to delineate further the basic mechanism responsible for the action of the LS-region.

While this paper was in preparation for publication, Picketts et al. (1993) reported a similar footprint pattern as shown here except for FP-III′ and FP-IV, and the binding of DBP and C/EBP to a region spanning nt −219 to −202 which may be involved in the amelioration of hemophilia B with a nt −5 mutation. The proposed mechanism may provide an alternative explanation for some of the Leyden phenotypes. It, however, does not provide any reasonable explanations for some important phenomena such as the lack of amelioration of the abnormal FIX gene with a mutation at nt −26 (Reijnen et al., 1992) and no significant increase of human FIX gene expression in normal individuals after the onset of puberty. Furthermore, it is important to note that all 16 hemophilia B-Leyden genes thus far identified have specific single mutations in the LS-region, but not in the C/EBP–DBP binding region (nt −202 to −219). This alone, however, does not downplay a possible important role of the C/EBP-DBP binding region, but suggests the complexity of the mechanism-(s) of FIX gene regulation responsible for the hemophilia B-Leyden phenotype and the need for a systematic analysis including site-directed mutagenesis of a much larger area of the 5′-end region.

ACKNOWLEDGMENT

We thank Dr. Diane M. Robins for valuable discussions. We appreciate Nancy Remaklus for her excellent help in preparation of this paper.

REFERENCES

Alder, A. J., Danielsen, M., & Robins, D. M. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 11660–11663.

Anson, D. S., Choo, K. H., Rees, D. J. G., Giannelli, F., Gould, K., Huddleston, J. A., & Brownlee, G. G. (1984) *EMBO J. 3*, 1053–1060.

Briet, E., Bertina, R. M., van Tilburg, N. H., & Veltkamp, J. J. (1982) *N. Engl. J. Med. 306*, 788–790.

Crossley, M., & Brownlee, G. G. (1990) *Nature 345*, 444–446.

Crossley, M., Winship, P. R., Austen, D. E. G., Rizza, C. R., & Brownlee, G. G. (1990) *Nucleic Acids Res. 18*, 4633.

Crossley, M., Ludwig, M., Stowell, K. M., De Vos, P., Olek, K., & Brownlee, G. G. (1992) *Science 257*, 377–379.

Dignam, J. D., Lebovitz, R. M., & Roeder, R. G. (1983) *Nucleic Acids Res. 11*, 1475–1489.

Evans, J. P., Watzke, H. H., Ware, J. L., Stafford, D. W., & High, K. A. (1989) *Blood 74*, 207–212.

Freedenberg, D. L., & Black, B. (1991) *Thromb. Haemostasis 65*, 964.

Freedman, L. P., Luisi, B. F., Korszun, Z. R., Basavappa, R., Sigler, P. B., & Yamamoto, K. R. (1988) *Nature 334*, 543–546.

Giannelli, F., Green, P. M., High, K. A., Sommer, S., Lillicrap, D. P., Ludwig, M., Olek, K., Reitsma, P. H., Goossens, M., Yoshioka, A., & Brownlee, G. G. (1992) *Nucleic Acids Res. 20, Suppl.*, 2027–2063.

Graves, B. J., Johnson, P. F., & McKnight, S. L. (1986) *Cell 44*, 565–576.

Hedner, U., & Davie, E. W. (1989) in *The Metabolic Basis of Inherited Diseases* (Shriver, C. R., Beaudet, A. L., Sly, W. S., & Vallie, D., Eds) 6th ed., pp 2107–2134, McGraw-Hill, New York.

Hirosawa, S., Fahner, J., Salier, J.-P., Wu, C.-T., Lovrien, E., & Kurachi, K. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4421–4425.

Jallat, S., Perraud, F., Dalemans, W., Balland, A., Dieterle, A., Faure, T., Muelien, P., & Pavirani, A. (1990) *EMBO J. 9*, 3295–3301.

Jones, K., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J., & Tjian, R. (1987) *Cell 48*, 79–89.

Klock, G., Strahle, U., Schmid, W., & Schutz, G. (1987) *Horm. Metab. Res., Suppl. Ser. No. 17*, 11–13.

Kurachi, K., & Davie, E. W. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6461–6464.

Kurachi, S., Furukawa, M., & Kurachi, K. (1991a) *Thromb. Haemostasis 65*, 801.

Kurachi, S., Furukawa, M., Yao, S.-N., & Kurachi, K. (1991b) *Am. J. Hum. Genet. 49*, 2429.

Kurachi, K., Yao, S.-N., Furukawa, M., & Kurachi, S. (1992) *Hosp. Practice 27(2)*, 41–51.

Kurachi, K., Kurachi, S., Furukawa, M., & Yao, S.-N. (1993) *Blood Coagulation Fibrinolysis* (in press).

Kusumoto, H., Hirosawa, S., Salier, J.-P., Hagen, F. S., & Kurachi, K. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7307–7311.

Miao, C. H., Leytus, S. P., Chung, D. W., & Davie, E. W. (1992) *J. Biol. Chem. 267*, 7395–7401.

Picketts, D. J., D'Souza, C., Bridge, P. J., & Lillicrap, D. (1992) *Genomics 12*, 161–163.

Picketts, D. J., Lillicrap, D. P., & Mueller, C. R. (1993) *Nature Genet. 3*, 175–179.

Reijnen, M. J., Sladek, F. M., Bertina, R. M., & Reitsma, P. H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6300–6303.

Reijnen, M. J., Peerlinck, K., Maasdam, D., Bertina, R. M., & Reitsma, P. H. (1993) *Blood 82*, 151–158.

Reitsma, P. H., Bertina, R. M., Ploos van Amstel, J. K., Riemens, A., & Briet, E. (1988) *Blood 72*, 1074–1076.

Reitsma, P. H., Mandalaki, T., Kasper, C. K., Bertina, R. M., & Briet, E. (1989) *Blood 73*, 743–746.

Royle, G., Van De Water, N. S., Berry, E., Ockelford, P. A., & Browett, P. J. (1991) *Br J. Haematol. 77*, 191–194.

Salier, J.-P., Diarra-Mehrprour, M., Sesboue, R., Bourguignon, J., Benarous, R., Ohkubo, I., Kurachi, S., Kurachi, K., & Martin, J. P. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8272–8276.

Salier, J.-P., Hirosawa, S., & Kurachi, K. (1990) *J. Biol. Chem. 265*, 7062–7068.

Sharp, P. A., & Marciniak, R. A. (1989) *Cell 59*, 229–230.

Simental, J. A., Sar, M., Lane, M. V., French, F. S., & Wilson, E. M. (1991) *J. Biol. Chem. 266*, 510–518.

Sladek, F. M., Zhong, W., Lai, E., & Darnell, J. E., Jr. (1990) *Genes Dev. 4*, 2353–2365.

Tan, J., Joseph, D. R., Quarmby, V. E., Lubahn, D. B., Sar, M., French, F. S., & Wilson, E. M. (1988) *Mol. Endocrinol. 2*, 1276–1285.

Tan, J., Marschke, K. B., Ho, K.-C., Perry, S. T., Wilson, E. M., & French, F. S. (1992) *J. Biol. Chem. 267*, 4456–4466.

Theil, E. C. (1990) *J. Biol. Chem. 265*, 4771–4774.

Thompson, A. R. (1991) *Prog. Hemostasis Thromb. 10*, 175–214.

Yao, S.-N., DeSilva, A. H., Kurachi, S., Samuelson, L. C., & Kurachi, K. (1991) *Thromb. Haemostasis 65*, 52–58.

Yoshinaga, S. K., Peterson, C. L., Herskowitz, I., & Yamamoto, K. R. (1992) *Science 258*, 1598–1604.

Yoshitake, S., Schach, B. G., Foster, D. C., Davie, E. W., & Kurachi, K. (1985) *Biochemistry 24*, 3736–3750.