MOLECULAR AND CELLULAR BIOLOGY, Jan. 1990, p. 95–102
0270-7306/90/010095-08$02.00/0

Vol. 10, No. 1

# A Naturally Occurring Gamma Globin Gene Mutation Enhances SP1 Binding Activity

KATHRYN SYKES[1] AND RUSSEL KAUFMAN[1,2]*

*Departments of Biochemistry[1] and Medicine,[2] Duke University, Durham, North Carolina 27710*

Received 19 July 1989/Accepted 27 September 1989

**Transcription of the human fetal globin genes in erythroid cells is tightly regulated during different stages of development and differentiation. Two naturally occurring mutations 202 base pairs upstream of the duplicated gamma globin genes are associated with incorrectly regulated gamma globin gene expression; elevated levels of fetal globin are synthesized during adult life. A C-to-G base substitution upstream of the $^{G}\gamma$-globin gene is highly correlated with a dramatic increase in gene expression. It increases the similarity of the region to the consensus Sp1 recognition site. We determined that the mutated DNA had a 5- to 10-fold-higher affinity for Sp1 than did the normal gamma globin gene sequence. We also observed a reduction in normal factor-binding activity. A different substitution at −202, C to T, upstream of the $^{A}\gamma$-globin gene was associated with a more moderate increase in fetal globin expression. This mutation decreased the similarity of the sequence to an Sp1 recognition site. We determined that it did not result in enhanced Sp1 binding but did alter normal factor binding. We suggest that these changes in nuclear protein-binding properties detected in vitro are responsible for the enhanced gamma globin gene expression found in −202 $^{G}\gamma\beta^{+}$ patients with hereditary persistence of fetal hemoglobin.**

The embryonic, fetal, and adult genes of the human beta-like globin gene family are expressed in mature erythrocytes in a precise developmental order. Production of embryonic (epsilon) globin terminates near week 8 of gestation and is supplanted by the synthesis of fetal (gamma) globin. Gamma globin remains the predominant beta-like globin chain present until birth, when it is replaced by adult (beta) globin synthesis. Although earlier investigations have clearly established that these changes in globin gene expression are regulated at the level of transcription (4, 5, 9, 42), the mechanisms responsible for these regulatory processes remain undefined.

A number of studies have provided insights into the role of *cis*-acting DNA elements in transcriptional control of the fetal, gamma globin genes. For example, sequence comparisons have revealed the presence of short DNA segments that are homologous to other known regulatory elements (2, 10). In vitro mutagenesis and in vivo expression studies have shown that upstream sequences are sufficient to stimulate transcription from a gamma globin or neutral promoter in both tissue culture cells and transgenic mice (1, 26, 32, 40, 41). In addition, analyses of naturally occurring upstream mutations in the gamma globin genes suggest a role for multiple regions in the regulation of transcription (6–8, 14–16, 18–21, 30, 35, 37). Some of the most recent work has focused on the *trans*-acting factors that interact with these *cis* DNA elements. In vitro binding assays have shown that a 300-base-pair (bp) DNA segment located 5′ to the human gamma globin genes specifically interacts with factors present in nuclear extracts of both erythroid and nonerythroid cell lines (22, 27, 28, 34). Mutations within some of these protein-binding sites that disrupt binding activity have been shown to correlate with altered promoter activity in erythroid and nonerythroid cells (29, 31, 36). Finally, the gene encoding one of these *cis* element binding proteins, GF-1, has been isolated (38).

Several classes of naturally occurring gamma globin gene mutations are associated with elevated gamma and reduced beta globin gene expression in adults. This is a clinical condition known as hereditary persistence of fetal hemoglobin (HPFH). Patients found with genetically defective beta globin expression in concert with one of these gamma globin gene mutations are relatively asymptomatic. One of the most dramatic increases in promoter activity is strongly correlated with a single C-to-G substitution 202 bp upstream of the cap site. A 30- to 40-fold increase in $^{G}\gamma$-globin gene expression during adult life is observed. In all known cases, this mutation affects only the $^{G}\gamma$-globin gene and not the normal, *cis*-linked $^{A}\gamma$-globin gene. It is the only nonpolymorphic change from the normal sequence over a 1,500-bp range (6, 8). Another base substitution (C to T) 202 bp upstream of the $^{A}\gamma$-globin gene is associated with a three- to sixfold increase in gene activity (21). The mechanism by which these point mutations may lead to the adult-stage overproduction of gamma globin in adults is unknown.

We report our comparative study of the normal and HPFH-associated upstream regions of the human gamma globin genes. We demonstrate that a C-to-G mutation at position −202 present in $^{G}\gamma\beta^{+}$ HPFH patients produces avid binding by the general transcriptional activator Sp1 concomitant with reduced binding by another factor(s). We propose that the molecular basis of this strong HPFH phenotype may relate to both the emergence of stage-nonspecific Sp1 binding and reduction of binding by another potential regulatory factor. We also show that a different base change 202 bp upstream of the $^{A}\gamma$-globin gene, found in a group of HPFH patients with a relatively moderate increase in fetal globin levels, does not bind Sp1 but may alter the normal factor interactions.

## MATERIALS AND METHODS

**Oligonucleotides.** Double-stranded oligonucleotides were synthesized to correspond to either the normal or mutated sequences spanning the gamma globin gene −200 region and the simian virus (SV40) Sp1 binding-site region. Each of the

* Corresponding author.

TABLE 1. Comparison of SV40 and gamma globin sequences to the consensus SP1 recognition and binding site

| DNA | Recognition site[a] | Relative Sp1 binding activity |
|---|---|---|
| SP1 consensus | G G GC<br>G GGC GG<br>T A AT | |
| | −81 −92 | |
| SV-V | à att ctg caG ATG GGC GGA GTaggga | +++ |
| $\gamma 27^{MUT}$ | aat tcT CTT GGG GGC GCC TTC CCC ACA | ++ |
| | −213 −191 | |
| $\gamma 27^{WT}$ | aat tCT CTT GGG GGC CCC TTC CCC ACA | + |
| $\gamma 27^{CT}$ | aat tCT CTT GGG GGC TCC TTC CCC ACA | + |

[a] Top strand of double-stranded oligonucleotide is given. Orientations is 5′ (left) to 3′ (right). Globin and SV40 nucleotides are capitalized; restriction site linker sequences are in lowercase. Numbers above gamma globin sequences designate nucleotide positions relative to the start site; numbers above SV40 sequences designate nucleotide positions according to Tooze (39). Underlines define positions of identity with the optimal consensus Sp1 binding site; the dot identifies a position of identity with a weak Sp1 binding site, HSV-TK(I) (24). Asterisks identify sites of −202 mutations.

synthetic DNA fragments has a restriction enzyme site at both the 5′ and 3′ ends to facilitate manipulations (Table 1).

**DNA probes.** End-labeled oligonucleotide probes for the mobility shift assays were generated by 5′ phosphorylating the double-stranded synthetic DNAs with [$\gamma$-$^{32}$P]ATP and T4 polynucleotide kinase. Unincorporated radioactive nucleotides were removed by passage over a Sephadex G-25 column. The probes were then gel fractionated, and the double-stranded oligonucleotides were excised and eluted by standard techniques.

**Cells.** All cells were grown in RPMI 1640 supplemented with 5% fetal calf serum. All culture densities were maintained at $2 \times 10^5$ to $5 \times 10^5$ cells per ml, or subconfluency.

**Nuclear extracts.** Nuclear protein extracts were prepared from cultured cells by the nucleus separation method of Dignam et al. (11). The nuclear extract was dialyzed against BC-100 (20 mM Tris [pH 7.9], 15% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol, 0.5 mM phenylmethylsulfonyl fluoride) for 3.5 h. Final protein concentrations were determined by the dye-binding method (Bio-Rad Laboratories). Standard curves were generated by using known amounts of bovine serum albumin, and several concentrations of nuclear extract within the linear range of this curve were assayed. Extracts were stored in BC-100 buffer in liquid nitrogen. Typically, 6 mg of protein was obtained from approximately $10^8$ cells.

**Protein fractionation.** Crude nuclear extracts prepared from K562 cells were fractionated on a 100-ml Superose 6 sizing column (Pharmacia, Inc.). Typically, 10 to 20 mg of nuclear proteins was loaded and run at a rate of 1 ml/min. Fractions were collected in 1.5-ml volumes. Protein concentrations were determined by the dye-binding method, and binding activity was assayed by mobility shift.

**Purified protein and antibody.** The rabbit polyclonal anti-Sp1 and preimmune antisera and the purified Sp1 protein were the kind gifts of S. Jackson and R. Tjian.

**Electrophoretic mobility shift assays.** For analysis of crude nuclear extracts, end-labeled DNA fragments (0.2 to 0.5 ng) were incubated with nuclear extracts (8 to 15 μg) from K562 and HeLa cells as described by Singh et al. (33), with minor modification. The reactions were carried out in binding buffer A (10 mM Tris, 50 mM NaCl, 1 mM dithiothreitol, 1 mM EDTA, 5% glycerol) in the presence of 0.5 μg of poly(dI-dC) · poly(dI-dC) and in a total volume of 25 μl. Binding experiments with fractionated proteins (10 to 15 μl of the specified fractions) included 0.2 mg of bovine serum albumin per ml. Binding was allowed to proceed for 30 min at 4°C. Sp1 mobility shifts were carried out by incubating the end-labeled DNA fragments with purified Sp1 (approximately 1 ng) in binding buffer Z[e] (20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid [HEPES], 100 mM KCl, 1 mM EDTA, 1 mM dithiothreitol) and without poly(dI-dC) · poly(dI-dC). Sp1 binding was conducted at 4°C for 15 min. Dye was added to the reactions containing crude or fractionated extracts or purified protein, and the samples were electrophoresed through a 6 or 7% nondenaturing acrylamide gel in 0.5× TBE buffer (45 mM Tris, 45 mM boric acid, 0.5 mM EDTA [pH 8.3]) at 150 V for 1 to 2.5 h. Gels were dried and autoradiographed.

Competition experiments included a 20- to 200-fold molar excess of a specific unlabeled DNA, as described in the figure legends. For each reaction, the buffer, poly(dI-dC) · poly(dI-dC), unlabeled specific competitor, and labeled DNA were incubated for 5 min before addition of the protein component.

The antibody shift experiment was conducted as described for the mobility shift assays, with minor modification. All components of the binding reaction except the labeled DNA were incubated with 0.75 μl of rabbit polyclonal antiserum or preimmune serum for 10 min on ice. The labeled DNA then was added, and the reaction was incubated for an additional 20 min before the gel was loaded.

## RESULTS

**Identification of nuclear factors that interact with the gamma globin gene upstream sequences.** Our studies were directed by a group of naturally occurring gamma globin gene mutations that suggest a regulatory function for sequences 200 bp upstream of the transcriptional start site (6, 8, 14, 21, 37). This region is identical in the evolutionarily duplicated $^{G}\gamma$- and $^{A}\gamma$-globin genes. A 27-bp oligonucleotide, $\gamma 27^{WT}$, was prepared that represents 23 bp of gamma globin sequence extending from nucleotides −213 to −191 (relative to the cap site) (Table 1). To determine whether specific binding sites for proteins are present within the −200 region, we prepared nuclear extracts from an erythroid and a nonerythroid source. K562 is a human erythroleukemia cell line that can synthesize the embryonic (epsilon) and fetal (gamma) globin chains; HeLa is a human epithelium-derived cell line.

Incubation of the labeled gamma globin oligonucleotide $\gamma 27^{WT}$ with the nuclear proteins resulted in several bound complexes when analyzed by mobility shift (Fig. 1, lanes 1, 3, and 5). A 20- to 200-fold excess of unlabeled $\gamma 27^{WT}$ effectively competed for these DNA-bound factors, whereas a 1,000- to 2,000-fold excess of nonspecific DNA competitor did not (see Fig. 5B, lanes 1 to 4). Factors found in both

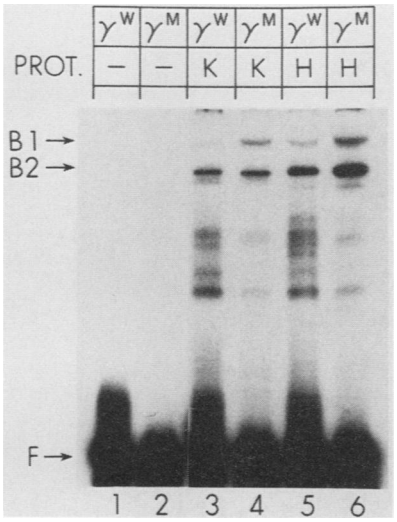


FIG. 1. Specific complexes formed between the gamma globin upstream region and nuclear proteins from erythroid and nonerythroid cells. The HPFH-related C-to-G mutation alters these complexes. The mobility shift assays were performed as described in Materials and Methods. $\gamma^W$ ($\gamma 27^{WT}$) and $\gamma^M$ ($\gamma 27^{MUT}$) were labeled and used in the binding reactions. PROT., Nuclear proteins added to the reaction as indicated; K, K562 cell nuclear extracts; H, HeLa cell nuclear extracts; B1 and B2, protein-bound labeled DNA; F, protein-free labeled DNA.

K562 and HeLa nuclear protein sources specifically interacted with $\gamma 27^{WT}$ to produce a major complex, B2, and a minor complex, B1. The remaining faster-mobility complexes also appeared with both extracts; however, they accumulated with prolonged storage and incubation times. Although we cannot exclude the possibility that they contained distinct factors, we suggest that these fast-mobility complexes may have contained degradation products of factors in B1 or B2. We have focused our attention on B1 and B2.

**The HPFH-related C-to-G mutation leads to an altered protein-binding pattern.** To examine the molecular basis of the gamma globin mutant phenotype HPFH, we synthesized an oligonucleotide containing the mutated gamma globin sequence. $\gamma 27^{MUT}$ differs from $\gamma 27^{WT}$ at only one nucleotide position, containing the same C-to-G substitution at position −202 that is associated with a 30- to 40-fold increase in adult-stage expression of the $^G\gamma$ (fetal) globin gene. Since this mutation is strongly correlated with dramatically altered gamma globin gene expression, we performed mobility shift assays to determine whether the mutation causes a change in any nuclear protein interaction.

End-labeled $\gamma 27^{MUT}$ was incubated with extracts from human K562 and HeLa cells in the presence of unlabeled nonspecific competitor DNA. Comparison of the mobility shift pattern produced by $\gamma 27^{WT}$ with that produced by $\gamma 27^{MUT}$ demonstrated that the two oligonucleotides interacted differently with the nuclear proteins. The single-base change enhanced the intensity of complex B1 and in some experiments modestly enhanced that of complex B2 (Fig. 1, lanes 2, 4, and 6). This pattern, one major complex associated with $\gamma 27^{WT}$ (B2) and two major complexes associated with $\gamma 27^{MUT}$ (B1 and B2), was also observed in extracts from hemin-induced K562 cells and human bone marrow (data not shown).

**Comparison of the gamma globin and SV40 binding sites.** The upstream gamma globin sequence centered about −200 relative to the cap site is similar to the sequence recognized by the general transcription factor Sp1. The C-to-G base substitution at position −202, correlated with HPFH, increased the similarity to the region of the Sp1 recognition and binding site such that it matched the decanucleotide consensus, optimal site in 7 of the 10 positions (Table 1). Although one base within the core GC box differed from the consensus, transcriptionally active Sp1 binding sites have been previously characterized that deviate from the recognition site within the core (24). Consequently, it has been suggested that the −202 mutation might produce a novel Sp1 binding site and stimulate transcription (4, 6). We designed experiments to determine whether the binding pattern differences between the wild-type and mutated gamma globin sequences were attributable to differential interaction with Sp1.

To study the gamma globin DNA-protein interactions in a less complex protein environment, the K562 cell nuclear extracts were size fractionated over a Superose 6 column. We assayed the DNA-binding activities of the fractions with both $\gamma 27^{WT}$ and the high-affinity Sp1 binding site V of SV40 (12, 17), SV-V. Figure 2 illustrates the relative binding activities of each protein fraction to the globin and SV40 DNAs. Two distinct but overlapping peaks of activity were produced. We believe that the fractions producing the $\gamma 27^{WT}$ binding activity peak contained a factor that specifically interacted with the normal globin sequence and that the fractions producing the SV-V activity peak contained Sp1.

Fractions 21 to 23 were pooled, and the interactions of the three labeled binding sites, $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V, were compared. $\gamma 27^{WT}$ and $\gamma 27^{MUT}$ interacted with the fractions to produce a band pattern similar to that in Fig. 1 (Fig. 3, lanes 1 and 4). Incubation of the fractions with labeled SV-V produced two specific complexes. They were similar in mobility to B1 and B2; however, the intensity of complex B1 produced with the SV40 DNA (B1-SV-V) was greater than that of the globin sequences (B1-$\gamma 27^{WT}$ and B1-$\gamma 27^{MUT}$) (Fig. 3, lane 7). Although the three binding sites exhibited similar binding patterns, the complexes differed quantitatively. $\gamma 27^{WT}$ produced B2, whereas B1 was not detectable. In contrast, SV-V primarily produced B1. $\gamma 27^{MUT}$ interacted with the fractionated proteins to form a binding pattern intermediate between that of the normal globin sequence and the SV40 Sp1 binding site.

**Sp1 binds the gamma globin upstream region in vitro.** We conducted an Sp1 antibody mobility shift experiment to determine whether complex B1, predominantly observed with SV-V and $\gamma 27^{MUT}$, contains Sp1. $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V were incubated with the K562 fractions 21 to 23 as previously described. To these DNA-protein reactions we added either the anti-Sp1 antibody or the normal preimmune serum, and the mixtures were fractionated on a nondenaturing gel (Fig. 3). If B1 contained the Sp1 protein, then we predicted that addition of the Sp1 reactive antibody would either associate with the [$^{32}$P]DNA-labeled complex and alter its mobility or disrupt complex formation. Consistent with the latter, addition of antibody to labeled SV-V and protein caused a reduction in complex B1-SV-V (lanes 7 to 9). A similar shift of complex B1-$\gamma 27^{MUT}$ was observed (lanes 4 to 6). In contrast, the $\gamma 27^{WT}$ complexes were not altered by addition of Sp1 antibody (lanes 1 to 3). We conclude that B1, a protein complex efficiently formed with the SV40- and HPFH-associated sequences but not the normal DNA, contained Sp1.

To demonstrate that the globin sequences interact with purified Sp1, we performed mobility shift binding studies

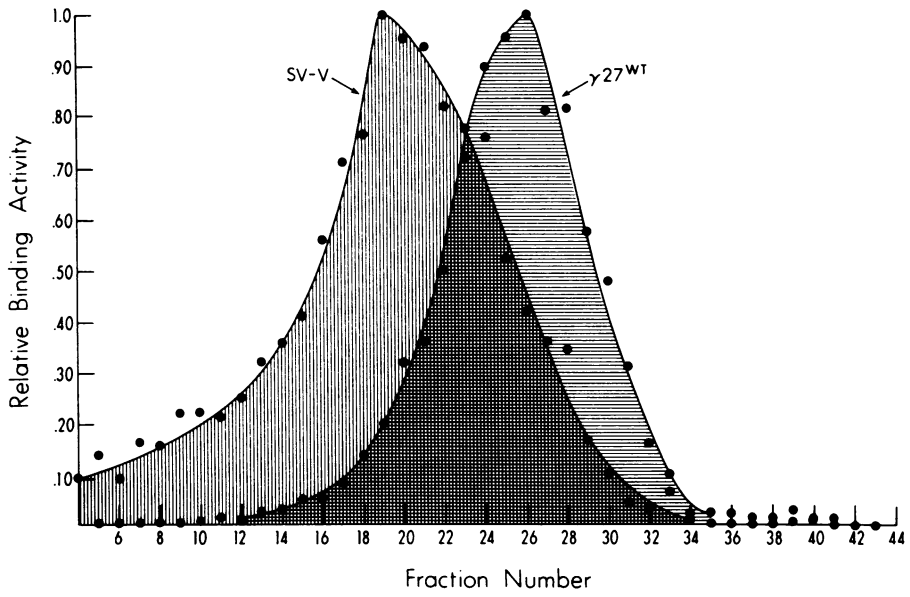


FIG. 2. Fractionation of K562 cell nuclear proteins. The *x* axis represents the fractions eluted from the Superose 6 (Pharmacia) column. The *y* axis indicates binding activity assayed by mobility shift and quantified by densitometric scans. For each binding site, the fraction with the highest activity was designated 1.0, and relative activities from each fraction were calculated. The relative activities of the fractions for one site are not calculated relative to activities for the other.

with a preparation of 60 to 90% pure Sp1 (3) (a gift of S. Jackson and R. Tjian). Radiolabeled gamma globin upstream sequences ($\gamma 27^{WT}$ and $\gamma 27^{MUT}$) and a high-affinity SV40 binding site (SV-V) were each incubated with the purified protein, and binding was analyzed on a nondenaturing gel. Purified Sp1 formed a specific complex with the SV40 and the mutated gamma globin sequences (Fig. 4, lanes 4 and 6). The normal globin sequence also formed a complex with Sp1 but showed a dramatically reduced affinity relative to that of the mutated DNA (Fig. 4, lanes 2 and 4). Densitometric scanning of these data demonstrated that $\gamma 27^{MUT}$ bound 8- to 9-fold more Sp1 than did $\gamma 27^{WT}$, whereas SV-V bound 7- to 10-fold more than did $\gamma 27^{MUT}$. The relative binding activities of the SV40 and globin sequences are summarized in Table 1.

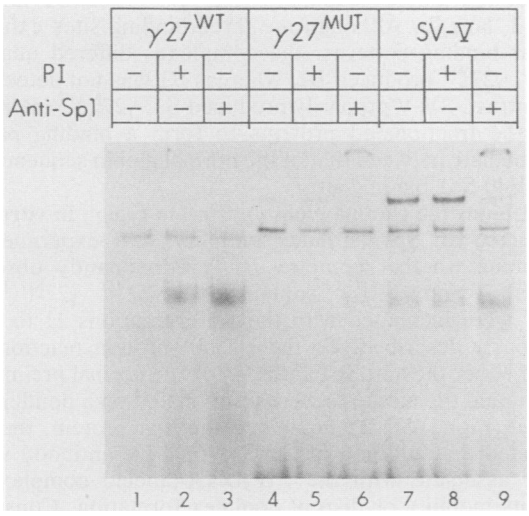


FIG. 3. Specific interaction of the anti-Sp1 antibody with complex B1. The experiment is described in Materials and Methods. Addition of either anti-Sp1 or preimmune serum to the binding reaction led to an increase in the fast-mobility complexes described earlier. The sera may have contained proteases that produced DNA-binding degradation products of the B1 and B2 factors and caused these complexes to appear. $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V are labeled DNAs. Each reaction contained 10 μl of pooled K562 cell fractions 21 to 23 (Fig. 2). PI, Preimmune serum present (+) or absent (−) from the reaction; Anti-Sp1, anti-Sp1 antibody serum present in (+) or absent from (−) the reaction.

**The globin and SV40 sequences share two binding activities.** A mobility shift titration experiment was conducted to ascertain the relative binding affinities of the SV40 and gamma globin sequences for Sp1 in erythroid extracts.

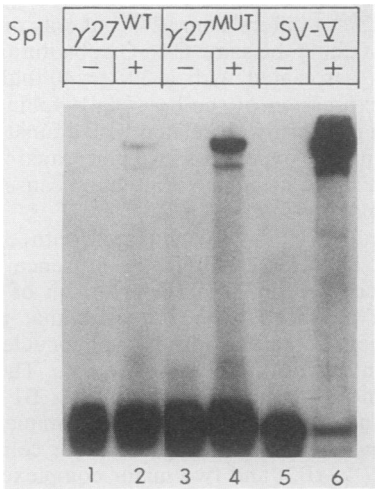


FIG. 4. Binding of purified Sp1 protein to the gamma globin upstream sequences. $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V were labeled and incubated with (+) or without (−) 60 to 90% pure Sp1 protein. Binding was assayed by mobility shift as described in Materials and Methods.

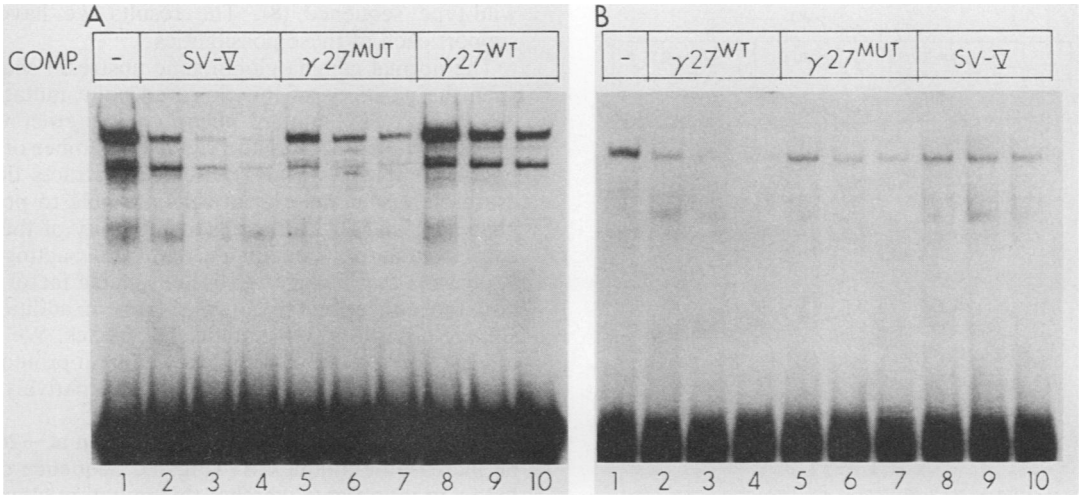


FIG. 5. Demonstration that the gamma globin and SV40 sequences share two different binding activities. (A) Titration of proteins bound to labeled SV-V with unlabeled SV40 and gamma globin DNAs, analyzed by mobility shift. Lanes: 1, no unlabeled competitor; 2 to 4, unlabeled SV-V added in 20-, 100-, and 200-fold excess over the probe; 5 to 7, unlabeled $\gamma 27^{MUT}$ added in 20-, 100-, and 200-fold excess over the probe; 8 to 10, unlabeled $\gamma 27^{WT}$ added in 20-, 100-, and 200-fold excess over the probe. (B) Titration of proteins bound to labeled $\gamma 27^{WT}$ with unlabeled gamma globin and SV40 DNAs, analyzed by mobility shift. Lanes: 1, no unlabeled competitor; 2 to 4, unlabeled $\gamma 27^{WT}$ added to the binding reaction in 20-, 100-, and 200-fold excess over the probe; 5 to 7, unlabeled $\gamma 27^{MUT}$ added in 20-, 100-, and 200-fold excess over the probe; 8 to 10, SV-V added in 20-, 100-, and 200-fold excess over the probe. Each reaction contained 15 μl of pooled K562 cell fractions 21 to 23. COMP., Competitor DNA.

Labeled SV-V incubated with the K562 fractions 21 to 23 was titrated with 20-, 100-, and 200-fold excesses of unlabeled SV-V, $\gamma 27^{MUT}$, and $\gamma 27^{WT}$ (Fig. 5A). The competition analysis demonstrated that $\gamma 27^{MUT}$ successfully competed for the SV-V-bound factors. Although less efficient than SV-V, $\gamma 27^{MUT}$ competed approximately 5- to 10-fold more efficiently than did $\gamma 27^{WT}$ (Fig. 5A). These estimated relative titration efficiencies are consistent with the relative binding affinities of these sites for purified Sp1.

We were interested in determining the relative binding affinities of the factor(s) in B2 for $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V. In a similar mobility shift titration experiment, labeled $\gamma 27^{WT}$ was incubated with K562 fractions 21 to 23 in the presence of 20-, 100-, and 200-fold excesses of unlabeled globin and SV40 binding-site DNAs (Fig. 5B). The factor in complex B2 bound to $\gamma 27^{WT}$ was competed for by SV-V 10-fold less efficiently than by $\gamma 27^{WT}$. $\gamma 27^{MUT}$ exhibited an intermediate B2 factor titration efficiency.

As with the SV-V-bound factor, the $\gamma 27^{WT}$-bound factors were shared by all three sequences ($\gamma 27^{WT}$, $\gamma 27^{MUT}$, and SV-V) but differed in their relative binding affinities for these sequences. The factor exhibiting the most efficient SV-V binding, Sp1, exhibited the least efficient $\gamma 27^{WT}$ binding. Conversely, the factor exhibiting the most efficient $\gamma 27^{WT}$ binding showed the least efficient SV-V binding.

**An HPFH-correlated C-to-T mutation does not acquire Sp1 binding activity.** Another naturally occurring mutation 202 bp upstream of the $^{A}\gamma$ globin gene cap site is also found associated with elevated gamma globin expression. In contrast to the 30- to 40-fold increase correlated with a C-to-G transversion, a C-to-T transition is correlated with only a three- to sixfold elevation in fetal globin levels (20). We have shown that the C-to-G substitution enhances Sp1 binding activity. Since the C-to-T substitution decreases its similarity to an Sp1 site, we postulated that this HPFH-related sequence should not exhibit enhanced Sp1 binding activity.

We synthesized an oligonucleotide representing the upstream region identical to $\gamma 27^{WT}$ but containing the C-to-T change at position −202 ($\gamma 27^{CT}$). Fractionated K562 proteins were incubated with $\gamma 27^{CT}$, $\gamma 27^{MUT}$, and $\gamma 27^{WT}$ and then compared by a mobility shift analysis. Unlike results for the C-to-G mutation, complex B1 was not enhanced by the C-to-T base change and remained a minor band (Fig. 6, lanes 2 and 6). Complex B2 appeared relatively unaltered by the mutation by direct binding analysis (Fig. 6, lanes 4 and 6). However, we conducted a competitor titration analysis to discern any differences in the relative binding efficiencies of the normal and mutated globin sequences for the B2 factor(s). The binding of labeled $\gamma 27^{WT}$ to K562 fractions 24 to 28 was titrated with increasing amounts of unlabeled $\gamma 27^{WT}$,

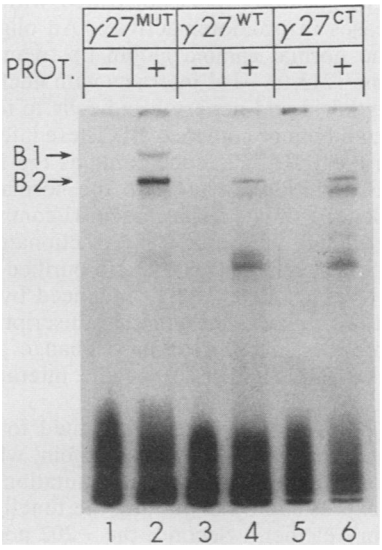


FIG. 6. Demonstration that the HPFH-related C-to-T mutation does not enhance complex B1. $\gamma 27^{WT}$, $\gamma 27^{MUT}$, and $\gamma 27^{CT}$ were labeled DNAs in the binding reaction. PROT., Presence (+) or absence (−) of K562 fractions 20 to 24 (Fig. 2).

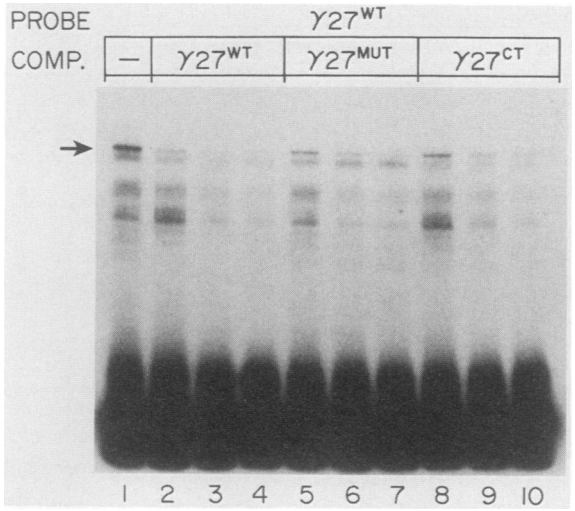


FIG. 7. Reduced normal factor-binding efficiencies of both HPFH mutations. End-labeled $\gamma 27^{WT}$ was incubated with pooled K562 fractions 24 to 28 (Fig. 2) and analyzed by competitor titration. Lanes: 1, no unlabeled competitor; 2 to 4, unlabeled $\gamma 27^{WT}$ added in 20-, 100-, and 200-fold excess over the probe; 5 to 7, unlabeled $\gamma 27^{MUT}$ added in 20-, 100-, and 200-fold excess over the probe; 8 to 10, unlabeled $\gamma 27^{CT}$ added in 20-, 100-, and 200-fold excess over the probe. COMP., Competitor DNA.

$\gamma 27^{MUT}$, or $\gamma 27^{CT}$ (Fig. 7). The competition analysis demonstrated that $\gamma 27^{WT}$ competed more efficiently for the B2 factor(s) than did either mutated sequence. We conclude that only $\gamma 27^{MUT}$ exhibited enhanced Sp1 binding activity but that both −202 mutations reduced the binding efficiencies of the normally bound protein(s).

## DISCUSSION

We have investigated the DNA region upstream of the gamma globin gene transcription unit to which a group of naturally occurring HPFH-associated mutations has been localized. Our analyses have been focused on a C-to-G transversion at position −202, correlated with one of the highest increases in promoter activity. An oligonucleotide containing the normal gamma globin upstream sequences from positions −213 to −191 interacts with nuclear proteins from both erythroid and nonerythroid cells to form a major complex, B2, and minor complex, Bl. These interactions are altered when the DNA sequence contains the HPFH-associated C-to-G mutation. Relative to the normal DNA, the mutated sequence exhibits an increased B1 complex. In vitro binding experiments with crude and fractionated erythroid proteins, an Sp1-specific antibody, and purified Sp1 protein have led us to conclude that B1, enhanced by the HPFH C-to-G mutation, contains the general transcriptional activator Sp1. Moreover, a different base change at the same position does not enhance the specific interaction of the sequence with Sp1.

Several hypotheses have been presented to explain the adult-stage overproduction of gamma globin, which is associated with a −202 mutation. First, the mutation may lead to constitutive expression by disrupting the function of a negative regulatory element. Second, the −202 point mutation may augment transcription by increasing the efficiency of a positive regulatory element normally spanning the region. Last, the mutated sequences may stimulate transcription by creating a positive regulatory element that is absent from the wild-type sequence (8). The results we have presented support each of these possibilities.

The normal gamma globin gene upstream sequences that span the positions of the clustered point mutations associated with HPFH mutant phenotypes interact with unfractionated nuclear factors to produce a number of specifically bound complexes (Fig. 1). Nuclear extracts derived from erythroid and nonerythroid cells are able to produce complexes B1 and B2. Although the ubiquity of the complexes suggests that they contain a general *trans*-acting factor(s), it is possible that a stage- or tissue-specific factor is bound in both the induced and uninduced states. In addition to B1 and B2, we observe faster-mobility complexes. We believe that at least some of these complexes contain proteolyzed products of the factors in B1 or B2; alternatively, they may contain other DNA-binding factors.

The HPFH-correlated C-to-G mutation at −202 results in an increase in complex B1 (Fig. 1). Sequence comparisons led us to investigate whether this mutation-altered complex contains Sp1. Using a polyclonal antibody produced against Sp1 and fractionated nuclear proteins in a mobility shift experiment, we establish that complex B1, enhanced by the HPFH-related mutation, contains Sp1. Analysis with purified protein allowed us to determine that the mutated upstream sequence binds purified Sp1 with an eight- to ninefold higher affinity than does the normal sequence (Fig. 3 and 4).

These results provide evidence in support of the second and third hypotheses proposed to explain the −202 constitutive phenotype. The mutated sequence stimulates transcription by augmenting or creating a positive regulatory element, namely, an Sp1 binding site. Other groups have previously demonstrated that a single Sp1 binding site is sufficient for transcriptional activation (24, 25). Although we detect low-level Sp1 binding by the normal sequence in vitro, this interaction appears to be minor compared with that produced by the normal binding factor, B2.

Since Sp1 binding sites were originally characterized in SV40, we compared the DNA-protein interactions produced by the gamma globin or SV40 DNA incubated with fractionated K562 cell nuclear proteins. A competitor titration experiment allowed us to distinguish two different binding activities associated with the gamma globin and SV40 sequences. The factor(s) in complex B2 that binds to $\gamma 27^{WT}$ also binds, with decreasing affinities, to $\gamma 27^{MUT}$ and SV-V. Reciprocally, the factor in B1, Sp1, maintains the highest affinity for SV-V but also binds, with decreasing affinities, to $\gamma 27^{MUT}$ and $\gamma 27^{WT}$ (Fig. 5).

The binding properties of $\gamma 27^{MUT}$ suggest that the combined effect of the loss of binding by one factor, possibly a repressor, and the enhanced binding by a transcriptional activator, Sp1, could produce the dramatic elevation in $^{G}\gamma$-globin expression associated with the C-to-G mutation. The factor in complex B2-$\gamma 27^{WT}$ is competed for 2- to 10-fold less efficiently by $\gamma 27^{MUT}$ than by the wild-type DNA (Fig. 5). One would anticipate that this would result in a reduction in the intensity of complex B2-$\gamma 27^{MUT}$ relative to B2-$\gamma 27^{WT}$. The intensity of complex B2 is not diminished and may appear modestly increased by the mutation (Fig. 1). These data are most consistent with the intensity of B2-$\gamma 27^{MUT}$ being a summation of the intensities of $\gamma 27^{MUT}$ bound to the factors that complex with both the wild-type globin and SV40 DNAs (Fig. 3, lanes 1, 4, and 7). Two different but overlapping complexes would not only create a seeming increase in the globin-bound factor(s) but would obscure any loss in globin factor-binding activity by the mutated relative to the normal globin sequence. Alternatively, the binding

differences may not be detectable under the excess binding-site conditions of Fig. 1, although they are measurable when the DNA sites are in competition with one another (Fig. 5).

The natural occurrence of another mutation at position −202 associated with elevated fetal globin expression in adults suggested that this mutation may also alter normal factor binding. However, since this C-to-T base change decreases the similarity of the region to an Sp1 site, we hypothesized that it would not resemble the binding pattern of the C-to-G mutation. We compared the protein-binding properties of the normal and mutated DNAs and demonstrated that the C-to-T change does not enhance Sp1 binding (Fig. 6). A competitor titration analysis demonstrates that both mutations moderately diminish B2 factor-binding efficiency (Fig. 7). We believe that the modest three- to sixfold increase in $^{A}\gamma$-globin expression associated with this C-to-T base substitution may be caused by an altered or reduced interaction with the normally bound factor(s). The disruption of a negative regulatory element would support one of the proposed hypotheses to explain the HPFH phenotype.

Complex B2 may contain a repressor factor that is active during nonfetal developmental stages. The core sequence of the wild-type −200 region is TGGGGGCCCCT. Similar sequences are present upstream of other globin genes, including the human epsilon (embryonic) globin gene, and their functions have yet to be determined. However, Emerson et al. (13) have recently identified an identical GC element upstream of the chicken beta globin gene and demonstrated that it is involved in activation and repression of gene expression in their in vitro transcription studies. We are currently developing similar in vitro systems to determine whether the altered interactions of the mutated sequences with Sp1 and the normally bound factor(s) are responsible for the associated changes in gene expression levels.

Finally, the cellular phenotype of the C-to-G −202 HPFH mutation is consistent with the proposed molecular mechanism by which this condition may occur. The entire population of erythroid cells in $-202\ ^{G}\gamma\beta^{+}$ HPFH patients contains elevated levels of fetal hemoglobin. This pancellular distribution of fetal cells is distinct from the distribution of another class of HPFH conditions in which only a subpopulation of erythroid cells have increased fetal hemoglobin levels. Therefore, the mutation apperas to allow the general factor Sp1 to influence transcription and lead to a general effect on all erythroid cells. Although these data do not determine whether the normal gamma globin region is involved in a general or specific mechanism of gene control, the protein-binding analyses of the two naturally occurring mutations, with their associated phenotypes, strongly suggest the importance of this region in gamma globin gene regulation.

## LITERATURE CITED


1. **Anagnou, N., S. Karlsson, A. Moulton, G. Keller, and A. Nienhuis.** 1986. Promoter sequences required for function of the human gamma globin gene in erythroid cell. EMBO J. **5:** 121–126.
2. **Anagnou, N., A. Moulton, G. Keller, S. Karlsson, T. Papapyannopoulou, G. Stamatoyannopoulos, and A. Nienhuis.** 1985. Cis-acting sequences that affect the expression of the human fetal gamma globin genes. Prog. Clin. Biol. Res. **191:**163–182.
3. **Briggs, M., J. Kadonaga, S. Bell, and R. Tjian.** 1986. Purification and biochemical characterization of the promoter-specific transcription factor, Spl. Science **234:**47–52.
4. **Chao, M., P. Mellon, P. Charnay, T. Maniatis, and R. Axel.** 1983. The regulated expression of beta globin genes introduced into mouse erythroleukemia cells. Cell **32:**483–493.
5. **Charnay, P., and T. Maniatis.** 1983. Transcriptional regulation of globin gene expression in the human erythroid cell line K562. Science **220:**1281–1283.
6. **Collins, F., C. Boehm, P. Waber, C. Stoeckert, S. Weissman, B. Forget, and H. Kazazian.** 1984. Concordance of a point mutation 5′ to the $^{G}$gamma globin gene with $^{G}$gamma beta$^{+}$ hereditary persistence of fetal hemoglobin in the black population. Blood **64:**1292–1296.
7. **Collins, F., J. Metherall, M. Yamakawa, J. Pan, S. Weissman, and B. Forget.** 1985. A point mutation in the $^{A}$gamma globin gene promoter in the Greek type of hereditary persistence of fetal hemoglobin. Nature (London) **313:**325–326.
8. **Collins, F., C. Stoeckert, G. Sergeant, B. Forget, and S. Weissman.** 1984. $^{G}$gamma beta$^{+}$ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5′ to the $^{G}$gamma gene. Proc. Natl. Acad. Sci. USA **81:** 4894–4898.
9. **Dean, A., T. Ley, R. Humphries, M. Ferdis, and A. Schechter.** 1983. Inducible transcription of five globin genes in K562 human leukemia cells. Proc. Natl. Acad. Sci. USA **80:**5515–5519.
10. **Dierks, P., A. van Ooyen, M. Cochran, C. Dobken, J. Reiser, and C. Weissman.** 1983. Three regions upstream from the cap site are required for accurate transcription of rabbit beta-globin gene in mouse 3T6 cells. Cell **32:**695–706.
11. **Dignam, J., R. Lebovitz, and R. Roeder.** 1983. Accurate transcription initiation by RNA pol II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **2:**1475–1489.
12. **Dynan, W., and R. Tjian.** 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. Cell **35:**79–87.
13. **Emerson, B., J. Nickol, and T. Fong.** 1989. Erythroid-specific activation and derepression of the chick beta-globin promoter in vitro. Cell **57:**1189–1200.
14. **Gelinas, R., M. Bender, C. Letshaw, P. Walser, H. Kazazian, and G. Stamatoyannopoulos.** 1986. Chinese $^{A}$gamma HPFH C to T substitution at position −196 of the $^{A}$gamma gene promoter. Blood **67:**1777–1779.
15. **Gelinas, R., B. Endlich, C. Pfeiffer, M. Yagi, and G. Stamatoyannopoulos.** 1985. G to A substitution in the distal CCAAT box of the $^{A}$gamma globin gene in Greek hereditary persistence of fetal hemoglobin. Nature (London) **313:**323–325.
16. **Gelinas, R., M. Yagi, B. Endlich, C. Letshaw, H. Kazazian, and G. Stamatoyannopoulos.** 1985. Sequences of $^{G}$gamma, $^{A}$gamma and beta genes of the Greek ($^{A}$gamma) HPFH mutant: evidence for a distal CCAAT box mutation in the $^{A}$gamma gene, p. 123–139. *In* G. Stamatoyannopoulos and A. Nienhuis (ed.), Experimental approaches for the study of hemoglobin switching. Alan R. Liss, Inc., New York.
17. **Gidoni, D., J. Kadonaga, H. Barrera-Soldana, K. Takahashi, P. Chambon, and R. Tjian.** 1985. Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. Science **230:** 511–517.
18. **Giglioni, B., C. Casini, R. Mantovani, S. Merli, P. Comi, S. Ottolenghi, G. Serglio, C. Camaschella, and V. Mazzo.** 1984. A molecular study of a family with Greek hereditary persistence of fetal hemoglobin and beta$^{+}$ thalassemia. EMBO J. **3:**2641–2645.
19. **Gilman, J., and T. Huisman.** 1985. DNA sequence variation associated with elevated fetal $^{G}$gamma globin production. Blood **66:**783–787.
20. **Gilman, J., F. Kutlar, M. Johnson, and T. Huisman.** 1987. A G nucleotide substitution 161 base pairs 5′ of the $^{G}$gamma globin gene site in a high $^{G}$gamma non-anemic person, p. 383–390. *In* G. Stamatoyannopoulos and A. Nienhuis (ed.), Developmental control of globin gene expression. Alan R. Liss, Inc., New York.
21. **Gilman, J., N. Mishima, X. Wen, F. Kutlar, and T. Huisman.** 1988. Upstream mutations associated with a modest elevation of fetal hemoglobin expression in human adults. Blood **72:**78–81.
22. **Gumucio, D., K. Rood, T. Gray, M. Riordan, C. Sartor, and F. Collins.** 1988. Nuclear proteins that bind the human gamma globin gene promoter: alterations in binding produced by point mutations associated with hereditary persistence of fetal hemo-

globin. Mol. Cell. Biol. 8:5310–5322.
23. **Jones, K., J. Kadonaga, P. Luciu, and R. Tjian.** 1986. Activation of the AIDS retrovirus promoter by the cellular transcription factor, Spl. Science 232:755–759.
24. **Jones, K., Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter in vitro. Cell 42:559–572.
25. **Kadonaga, J., K. Jones, and R. Tjian.** 1986. Promoter-specific activation of RNA polymerase II transcription by Spl. Trends Biol. Hematol. Sci. 11:20–23.
26. **Lin, H., N. Anagnou, T. Rutherford, T. Shimada, and A. Nienhuis.** 1987. Activation of the human beta-globin promoter in K562 cells by DNA sequences 5′ to the fetal gamma or embryonic delta globin genes. J. Clin. Invest. 80:374–380.
27. **Lingrel, J., J. Weimer, and A. Menon.** 1987. Binding of nuclear factors to an upstream region of the human $^{A}$globin gene, p. 201–210. *In* G. Stamatoyannopoulos and A. Nienhuis (ed.), Developmental control of globin gene expression. Alan R. Liss, Inc., New York.
28. **Mantovani, R., N. Malgaretti, B. Giglioni, P. Comi, N. Cappellini, S. Nicolis, and S. Ottolenghi.** 1987. A protein factor binding to an octamer motif with gamma globin promoter disappears upon induction of differentiation and hemoglobin synthesis in K562 cells. Nucleic Acids Res. 15:9349–9364.
29. **Martin, D., S. Tsai, and S. Orkin.** 1989. Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid specific DNA-binding factor. Nature (London) 338:435–438.
30. **Ottolenghi, S., S. Nicolis, R. Taramelli, N. Malgaretti, R. Mantovani, P. Comi, B. Giglioni, M. Longinotti, F. Dore, L. Oggiano, D. Pistidda, A. Serra, C. Camaschella, and G. Saglio.** 1988. Sardinian $^{G}$gamma HPFH: A T→C substitution in a conserved "octamer" sequence in the $^{G}$gamma globin promoter. Blood 71:815–817.
31. **Rixon, M., and R. Gelinas.** 1988. A fetal globin gene mutation in $^{A}\gamma$ nondeletion hereditary persistence of fetal hemoglobin increases promoter strength in a nonerythroid cell. Mol. Cell. Biol. 8:713–721.
32. **Rutherford, T., and A. Nienhuis.** 1987. Human gene promoter sequences are sufficient for specific expression of a hybrid gene transfected into tissue culture cells. Mol. Cell. Biol. 7:398–402.
33. **Singh, H., R. Sen, D. Baltimore, and P. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) 319:154–158.
34. **Superti-Furga, G., A. Barberis, G. Schaffner, and M. Busslinger.** 1988. The −117 mutation in Greek HPFH affects the binding of three nuclear factors to the CCAAT region of the gamma-globin gene. EMBO J. 7:3099–3107.
35. **Surrey, S., K. Delgrosso, P. Malladi, and E. Schwartz.** 1988. A single base change at position −175 in the 5′ flanking region of the $^{G}$gamma globin gene from a Black with $^{G}$gamma beta+ HPFH. Blood 71:807–810.
36. **Stoeckert, C., J. Metherall, M. Yamakawa, S. Weissman, and B. Forget.** 1987. Expression of the affected $^{A}\gamma$ gene associated with Greek nondeletion hereditary persistence of fetal hemoglobin. Mol. Cell. Biol. 7:2999–3003.
37. **Tate, V., W. Wood, and D. Weatherall.** 1986. The British form of hereditary persistence of fetal hemoglobin. Results from a single mutation adjacent to an S1 hypersensitivity site in the $^{A}$gamma globin gene. Blood 68:1389–1393.
38. **Tsai, S., D. Martin, L. Zon, A. D'Andrea, G. Wong, and S. Orkin.** 1989. Cloning of cDNA for the major DNA-binding protein of the erythroid lineage through expression in mammalian cells. Nature (London) 339:446–451.
39. **Tooze, J.** 1981. Molecular biology of tumor viruses, 2nd ed., part 2, revised. DNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
40. **Trudel, M., J. Magreim, L. Bruckner, and F. Constantini.** 1987. Upstream $^{G}\gamma$-globin and downstream β-globin sequences required for stage-specific expression in transgenic mice. Mol. Cell. Biol. 7:4024–4029.
41. **Waber, P., M. Bender, R. Gelinas, C. Kattamis, A. Karaklis, K. Safroniadou, G. Stamatoyannopoulos, F. Collins, B. Forget, and H. Kazazian.** 1986. Concordance of a point mutation 5′ to the A globin gene with $^{A}$gamma beta$^{+}$ hereditary persistence of fetal hemoglobin in Greeks. Blood 67:551–554.
42. **Wright, S., A. Rosenthal, R. Flavell, and F. Grosveld.** 1984. DNA sequences required for regulated expression of beta-globin genes in murine erythroleukemia cells. Cell 38:265–273.