The T→C substitution at −198 of the Aγ-globin gene associated with the British form of HPFH generates overlapping recognition sites for two DNA-binding proteins

Klaus-Dieter Fischer and Joachim Nowock*
Heinrich-Pette-Institut für Experimentelle Virologie und Immunologie an der Universität Hamburg, Martinistrasse 52, 2000 Hamburg 20, FRG

Received July 11, 1990; Revised and Accepted September 3, 1990

ABSTRACT

Defects in the developmental changes of human hemoglobin production characterized by the continued expression of fetal globin during adult life are classified as hereditary persistence of fetal hemoglobin (HPFH). Among the various molecular lesions associated with this phenotype, the non-deletion forms with point mutations in the promoter region are thought to provide mechanistic clues for γ-globin gene regulation. The natural occurrence of four different base substitutions mapping within six nucleotides of a homopurine.homopyrimidine motif in the upstream promoter region demarcate a potential control element. To assess its importance for transcriptional activity, we compared the −202 (C→G), −198 (T→C) and −196 (C→T) HPFH mutations with the normal sequence in binding studies with nuclear proteins from erythroid and non-erythroid cells. Wildtype DNA and HPFH mutations at −202 or −196 showed only a weak protein interaction of unclear functional significance. In contrast, −198 (T→C) generated overlapping, high-affinity binding sites for two ubiquitous nuclear proteins. One cognate protein was identified as the transcription factor Sp1. The second one was termed NF-G.C as it interacted strongly with the homopolymer poly(dG).poly(dC). The generation of additional recognition sites for trans-acting factors by the −198 HPFH mutation correlated with a modest increase in promoter activity in vitro specifically with nuclear extracts from erythroid cells. The activation appears to be mediated by binding of Sp1, but it requires interaction with an erythroid-specific factor, most likely GF-1. Templates containing the −196 HPFH mutation showed a transcriptional activity identical to wildtype. This suggests that despite the topological proximity of the mutations, the HPFH phenotype may be established by different mechanisms.

INTRODUCTION

In man, γ-globins constitute the dominant β-type chains of hemoglobin during the fetal phase. During the perinatal period they are replaced by adult chains (δ, β), and only a residual synthesis of fetal hemoglobin persists. This phenomenon is called the fetal to adult hemoglobin switch.

Hereditary disorders which show a persistent expression of fetal globin (HPFH) have attracted much interest owing to the expectation that they might shed light on the mechanism regulating the switch. Among the varied genotypes that cause this phenotype (reviewed in 1), the group of non-deletion HPFH is particularly thought to disclose regulatory DNA sequences that are involved in the switching of fetal globin. The genetic alterations which correlate with HPFH are single point mutations in the promoter region of the Gγ- or Aγ-globin genes. For the −117 and −175 HPFH mutations (2, 3) it was shown that the base alterations affected the binding of transcription factors in vitro (4, 5) and led to a moderate increase of transcriptional activity after transfection into erythroid cells (6, 7).

Four single base substitutions associated with HPFH have been found in an upstream promoter region (8−11). The mutations mapped at three closely spaced sites at position −202, −198 and −196. Further, the −196 C→T transition was detected in three different populations (9, 12, 13) indicating independent mutational events (Fig. 1).

The three affected sites reside in a short segment of oligopurine.oligopyrimidine (oligo(R.Y)) asymmetry with high GC content which may adopt a non-B conformation (reviewed in 14). Indeed, when cloned into plasmids and probed with S1, a site immediately 5′ of the HPFH mutations was hypersensitive (10). This suggests that an unorthodox structure may be a control element in the developmental regulation of γ-globin gene expression. Alternatively, based on the proposal that the −202 C→G substitution could convert this sequence to a recognition site for the transcription factor Sp1 (8), altered protein-DNA interactions might be the basis of the phenotypic expression of the HPFH mutations.

* To whom correspondence should be addressed

To test whether the generation of additional protein-binding sites would affect promoter activity, we compared the −202 C→G, −198 T→C and −196 C→T HPFH mutations with the normal sequence by in vitro protein-DNA binding experiments and by in vitro transcription studies.

MATERIALS AND METHODS

Plasmid constructs

Oligonucleotides encompassing the GC-rich region of the γ-globin genes (15) from position −208 to −193 were synthesized on a Pharmacia Gene Assembler Plus and represented the wildtype configuration or the HPFH mutations −202, −198 and −196:

−208 G C T −193

5′-GATCCGGGGGCCCCTTCCCCA-3′.

They were cloned as BamHI-BglII fragments in pUC19, which was modified by insertion of the oligonucleotide 5′-CAGATCTG-3′ (BglII site) into the polylinker HincII site. A BstNI-BalI fragment containing the CACCC box region from −155 to −130 was ligated blunt-end into the SmaI site of pUC19. The inserts were excised with EcoRI and HindIII for use in the protein-DNA binding reactions. Templates for in vitro transcription were derived from pMMTV-CAT (16) by replacing the MMTV-LTR by Gγ-globin sequences from −383 or −203 to +52. The resulting constructs were p-383γ-CAT and p-203γ-CAT. p-198 HPFH-CAT and p-196 HPFH-CAT were generated by inserting synthetic ApaI-NcoI fragments containing the respective HPFH mutations into p-383γ-CAT. In pF-MuLV-CAT sequences containing the Friend murine leukemia virus 3′ LTR from a ClaI site in the env region to a KpnI site in the R region (17) replaced the LTR of the parental construct. Construct integrity was verified by sequencing of the relevant regions.

Cell lines

K562 and HEL are human erythroleukemic cell lines that synthesize embryonal and fetal globin chains (18, 19). Treatment of HEL cells with TPA induces a shift to a macrophage-like phenotype (20). B8/3 is a subline of the Friend cell line B8 (21). HEL and K562 cells were grown in IMDM supplemented with 2 mM glutamine and 10% fetal calf serum. The embryonic stem cell line ES-CCE (22) was maintained in BRL conditioned medium (60/40) containing 20% fetal calf serum. All other cells were kept in MEM, 2 mM glutamine, and 10% fetal calf serum.

Preparation of nuclear extracts

Nuclear extracts from HEL, B8/3 and HeLa cells were prepared essentially as described by Dignam et al. (23). All other cells or tissues were processed following the procedure of Gorski et al. (24). To minimize proteolytic breakdown, a set of protease inhibitors as recommended by Thornell et al. (25) was added to all buffers. HEL and HeLa nuclear extracts were chromatographed on wheat germ agglutinin columns to enrich for Sp1 as described by Jackson and Tjian (26). Further purification was achieved by two cycles of DNA-affinity chromatography (27) with a matrix containing a retroviral homology to the Sp1 consensus binding sequence. GF-1 was purified by two cycles of affinity chromatography on a matrix containing Gγ-promoter sequences from position −193 to −165 containing the −175 HPFH mutation.

Band shift assay

Reaction mixtures contained in 15 μl 0.5 to 2.0 fmoles DNA fragment, 2.5 to 10 μg protein, 2 mM $MgCl_2$, 1 mM DTE, 600 ng poly(dA-dT), 300 ng poly(dI-dC), 0.025% NP-40 and 4% Ficoll. The end-labeled probe was added last after a 15 min preincubation on ice, and the reaction was continued for additional 15 min. Samples were loaded on a 5% polyacrylamide (29:1) gel in 0.5×TBE and electrophoresed at 10 V/cm for 2.5 hrs. Gels were dried and exposed to Kodak XAR-5 film.

Methylation interference analysis

End-labeled fragments were partially methylated with dimethyl sulfate (28). The binding reaction was scaled up 15-fold as compared to the band shift assay described above. Free and protein-bound probes were eluted from the gel and processed as described by Sturm et al. (29), followed by the G > A reaction. The samples were analyzed on a 10% polyacrylamide (19:1)/7 M urea gel.

Missing contact probing

Premodification of the end-labeled probes was performed as described by Brunelle and Schleif (30). DNA fragments were partially depurinated or depyrimidated by the G+A and C+T reactions of Maxam and Gilbert (31). Premodified DNA was reannealed in 0.2 M NaCl at 65°C for 2 hrs. Binding reactions, gel separation and recovery of the probes were as described for the methylation interference assay. Strand cleavage occurred in 100 μl freshly diluted 1 M piperidine at 90°C for 30 min.

DNase protection

In the standard binding reaction (30 μl) 5 to 10 fmoles DNA fragment and 50 to 200 μg protein were incubated in a buffer containing 20 mM Hepes-KOH, pH 7.9, 60 mM KCl, 0.5 mM EDTA, 6.25 mM $MgCl_2$, 2 mM DTE, 10% glycerol, 0.025% NP-40, 1 μg poly(dA-dT), 1 μg poly(dI-dC) for 30 min on ice. Then 5 μl of 10 to 100 μg/ml DNaseI in 35 mM $MgCl_2$ were added and incubated at 20°C for 90 sec. Reactions were terminated with 170 μl of a mixture containing 10 mM EDTA, 0.5% sarkosyl, 500 mM NaCl and 3 μg glycogen as carrier. After phenol/chloroform extraction and ethanol precipitation, the samples were loaded onto an 8% sequencing gel.

In vitro transcription

In the standard transcription reactions 90 fmoles of supercoiled test template, 5 fmoles of pCMV-CAT (32) as reference template and pUC18 up to 600 ng total DNA were incubated with nuclear extracts at protein concentrations of 4 to 20 mg/ml in 15 μl of a buffer containing 25 mM Hepes-KOH, pH 8.0, 60 mM KCl, 6 mM $MgCl_2$, 0.6 mM rNTPs, 1 mM DTE, 10% glycerol and

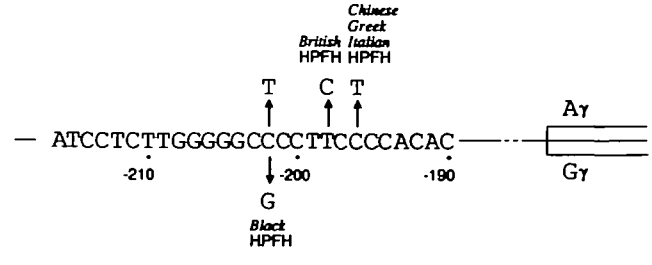

Figure 1. Occurrence of four different HPFH mutations in a GC-rich sequence of the γ-globin gene promoter.

0.025% NP-40 at 30°C for 45 min. Reaction products were purified by adding 180 µl 0.5 M LiCl, 1 M urea, 0.25% SDS, 20 mM sodium citrate and 2.5 mM CDTA (final pH 6.8) and 5 µl of 1 mg/ml proteinase K. Incubation at 50°C for 30 min was followed by phenol/chloroform extraction and LiCl/ethanol precipitation (33). Transcripts were hybridized to an oligonucleotide complementary to CAT RNA (5'-CCATTTT-AGCTTCCTTAGCTCC-3') and analyzed by primer extension as described by Südhoff et al. (34). Competition assays contained in 20 µl 25 fmoles test template, 25 fmoles pF-MuLV-CAT as internal reference and 5 pmoles of restriction fragments as competitor. Total DNA was brought to 400 ng by pUC18.

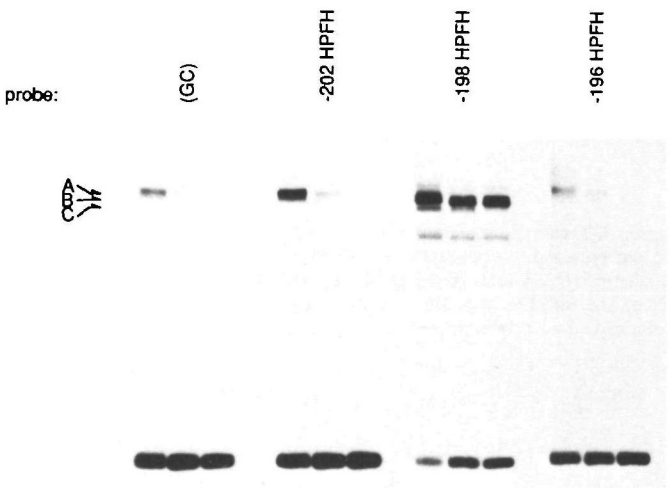

Figure 2. Band shift analysis of nuclear proteins binding to the −208 to −193 region of γ-globin. Fragments containing the wildtype sequence (GC) or the indicated HPFH mutations were incubated with 10, 5 and 2.5 µg of nuclear proteins of HEL cells.

Nuclear extracts were incubated with competitor DNA for 30 min on ice before the addition of templates and ribonucleotides. The incubation was continued for 45 min at 30°C. The synthesis of specific RNAs was sensitive to low levels of α-amanitin indicative of RNA-polymerase II products.

RESULTS

−198 T→C HPFH generates overlapping binding sites for two nuclear proteins

The cluster of three base substitutions in the GC-rich region of the γ-globin genes that are associated with non-deletional HPFH could affect a single regulatory element and thus act through a common mechanism. Since specific protein-DNA interactions might be involved in this mechanism, we analyzed the binding potential of wildtype and mutant DNA fragments spanning the region from −208 to −193 with nuclear proteins from erythroid cell lines expressing fetal or adult globin genes. Extracts from HEL cells synthesizing fetal globins showed only a weak complex with the wildtype fragment in a band shift assay (Fig. 2, band A). A similarly weak band with identical mobility appeared with the mutant fragments −202 HPFH and −196 HPFH. In contrast, the −198 mutation formed two strong complexes of faster mobility (band B and C), each of which could be resolved as double bands. Faster migrating complexes are probably due to proteolytic degradation. Under conditions of protein excess, complex B dominates (not shown). Nuclear extracts from K562 cells and the murine Friend cell line B8/3, which synthesize fetal and adult globins, respectively, generated the same binding pattern as HEL (not shown). When larger restriction fragments with additional flanking sequences were used as binding probes, no other complexes were observed in the GC-rich region, excluding the possibility that the shorter oligonucleotides provided insufficient sequence information. Thus, from the three HPFH mutants only the T→C transition at −198 causes alterations in protein-DNA interactions detectable in this assay. Interestingly,

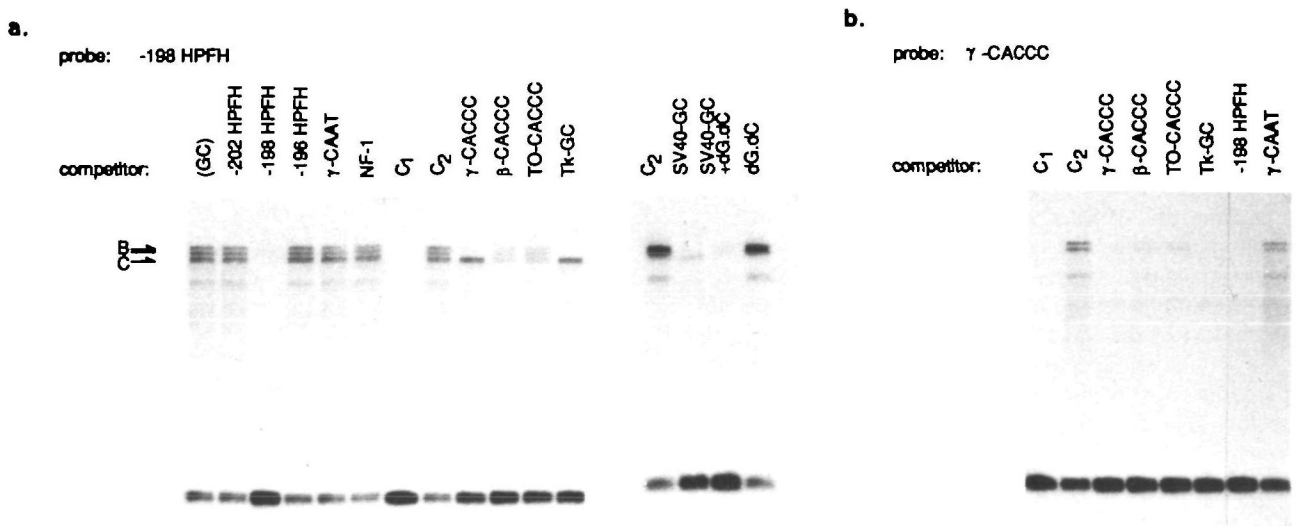

Figure 3. Competition band shift assays. **a:** Binding of HEL nuclear proteins to 1 fmole of the −198 HPFH oligonucleotide was competed with 15 ng of various DNA fragments representing a 250- to 750-fold molar excess over probe. The following competitor sequences were used: (GC), HPFH: normal γ-globin region from position −208 to −193 or HPFH mutants; γ-CAAT: γ-globin CAAT boxes (−130 to −52); NF-1: consensus binding sequence of nuclear factor 1 (16); γ-CACCC: γ-globin CACCC motif (−154 to −131); β-CACCC: β-globin CACCC motif (−127 to −77); TO-CACCC: CACCC motif of rat TO gene (−455 to −439); Tk-GC: GC-box of Tk gene (−109 to −81); SV40-GC: GC-boxes (21 bp repeats) of SV40 (−128 to −40); dG.dC: poly(dG).poly(dC). C_1 was a control devoid of protein. In C_2 pUC18 was used as unspecific competitor. **b:** The γ-globin CACCC region was used as probe. All other conditions as in **a**.

this mutation generates new binding sites for trans-acting factors rather than decreasing the affinity of a site in the normal gene.

On inspection of wildtype and mutant sequences it became apparent that the −198 HPFH mutation resulted in an increased match to the consensus sequence of the transcription factor Sp1 (35−37). We compared the γ-globin sequences with the prototypical Sp1 binding sites (GC boxes) of the herpes simplex thymidine kinase (Tk) gene promoter (38) by in vitro competition binding. An additional GC box homology exists in the CACCC motif of the γ-globin gene (when read in the opposite orientation) around position −140. It has been indicated that the related sequence in the β-globin gene can bind to Sp1 and to proteins with similar DNA recognition characteristics (39). Therefore, the CACCC region from the human β- and γ-globin promoter and from the rat tryptophan oxygenase (TO) gene (40) were included in the competition binding analysis. Fig. 3A shows that neither the normal GC-rich region of γ nor HPFH mutations −202 and −196 competed for the −198 HPFH sequence, consistent with the direct binding data. However, the Sp1 binding sites of the Tk promoter and the γ-globin CACCC box efficiently suppressed the two signals of complex B, but left complex C unaffected. Sequences containing CACCC motifs of β-globin and the TO gene competed complex B to a lesser extent. Complex C was selectively competed with the synthetic polymer poly(dG).poly(dC) (Fig. 3a). When using the γ-globin CACCC box as probe, only a doublet occurred with the same mobility as complex B for the −198 HPFH region (Fig. 3b). The competition pattern was consistent with the experiment using −198 HPFH sequences as tracer.

We conclude that the T→C transition at −198 allows for two types of specific protein interactions which can be discriminated by their binding behavior and not only by protein size and/or charge. The recognition sequence of complex B shares strong homology with the γ-globin CACCC motif, and most likely they represent a binding site for Sp1. The protein(s) in complex C bind to the uniform dG.dC polymer reminiscent of the G-string binding protein (BGP1) detected in chicken (41). We will refer to this activity as NF-G.C. As reported for Sp1 and BGP1, proteins in complex B and C required Zn^{++} for binding (data not shown).

Tissue distribution of proteins binding to the −198 HPFH sequence

Since the mutation at −198 accommodated the alternate binding of two different proteins, the question of their relation to the HPFH phenotype was addressed. The mechanism by which the mutant acts could by itself be cell-specific or it might operate within the context of the stage- and cell-specific expression of γ-globin, as would be expected if the ubiquitous transcription factor Sp1 contributed to the effect.

We therefore analyzed the tissue distribution of proteins that generated complex B or C. A DNA fragment containing the −198 mutation was incubated with extracts from erythroid cells (HEL, K562, B8/3) and from other hematopoietic cells representing the myeloid (WEHI-3B; TPA-induced HEL) or the lymphoid lineage (thymus) as well as non-hematopoietic cells (HeLa, embryonic cells CCE). HEL and HeLa cell extracts purified by sequential affinity chromatography on a wheat germ agglutinin (26) and a sequence-specific DNA column (27) containing the GC box homology of a retroviral promoter also gave rise to protein-DNA interactions identical to the complex B formed with crude extracts

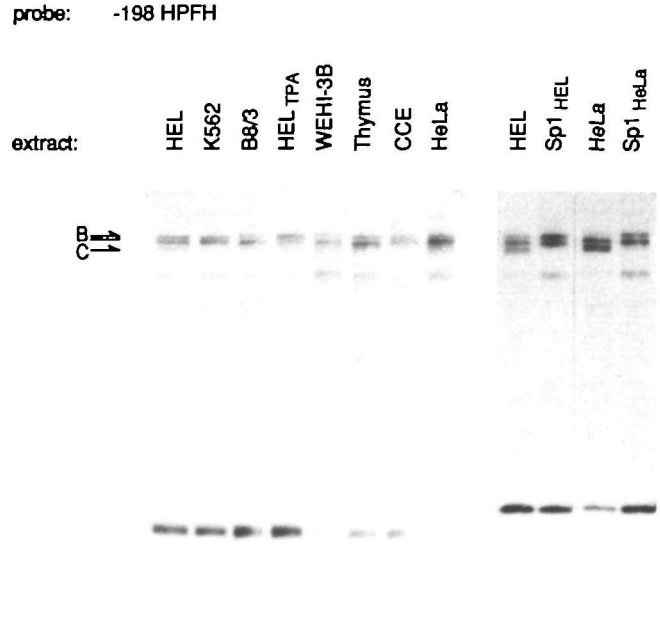

Figure 4. Tissue distribution of proteins binding to the −198 HPFH mutation. An oligonucleotide containing the −198 HPFH mutation was incubated with 5 μg of nuclear proteins from erythroid and non-erythroid cells. $Sp1_{HEL}$ and $Sp1_{HeLa}$ denote Sp1 fractions from HEL and HeLa cells highly purified by wheat germ lectin and specific DNA-affinity chromatography.

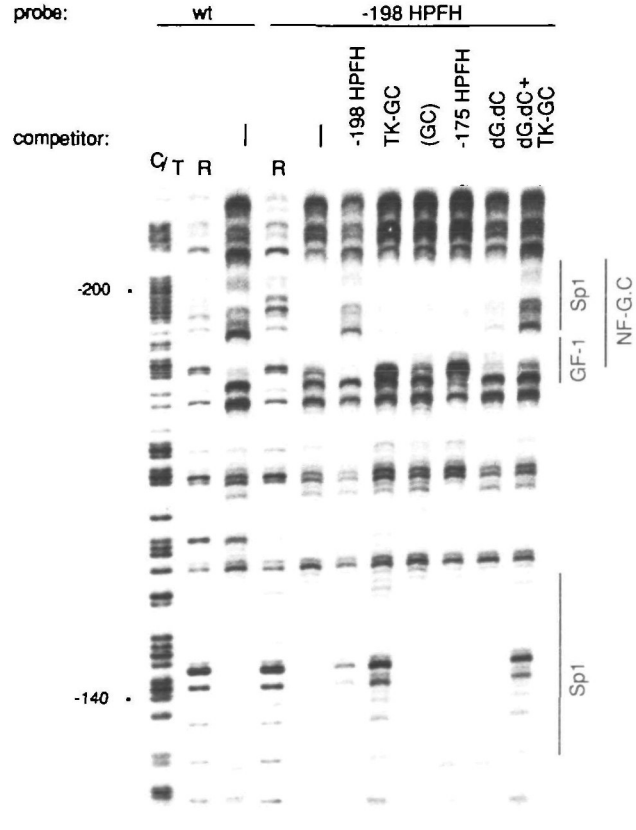

Figure 5. In vitro footprinting. DNA fragments extending from −383 to −54 of wildtype or −198 HPFH sequences were labeled at the promoter-proximal 3′ end and incubated with nuclear extract from HEL cells. Binding to the −198 HPFH sequence was challenged with GC-rich oligonucleotides as described in Fig. 3. A sequence from position −193 to −165 containing a −175 HPFH mutation (7) was used as competitor for GF-1 binding. R is the reference reaction containing BSA instead of nuclear extract. A C+T chemical sequencing reaction (C/T) was used as position marker.

(Fig. 4). (The band of lower mobility is probably due to a breakdown product.) This provided further evidence that complex B is formed by binding of Sp1.

Although the relative amount of complex C varied in the extracts from the different sources (Fig. 4), its presence became evident on longer exposures of the autoradiograms. Thus, both proteins that show high affinity binding to the -198 HPFH mutant have a ubiquitous distribution.

Characterization of protein interactions with the −198 HPFH and CACCC motifs

DNAse I footprinting

To further characterize the protein interactions with the −198 HPFH and CACCC box motifs, we used footprinting and base modification techniques. Incubation of a promoter fragment containing the −198 T→C mutation with nuclear extract from

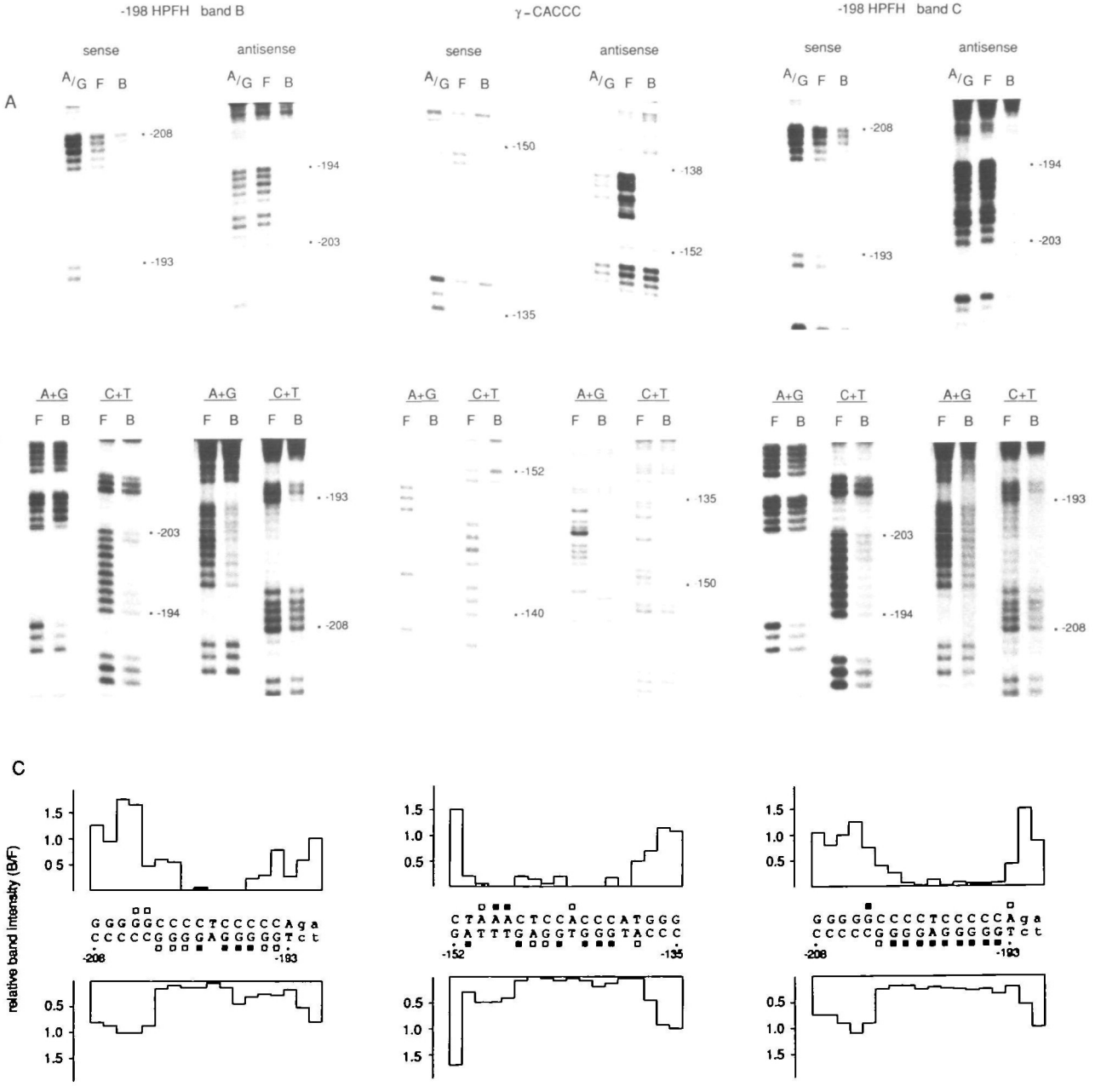

Figure 6. Methylation interference and base contact probing. **A:** Fragments containing the −198 HPFH mutation or the γ-CACCC motif were selectively labeled at either 3′ end by the Klenow fill-in reaction and partially methylated with DMS. Following preparative band-shifting, samples of the protein-bound (B) and free (F) DNA were treated by the G>A chemical cleavage reaction (A/G). Equal amounts of radioactivity were separated on a 10% sequencing gel. Map positions are indicated at the margins and can be compared with the sequences in C. Sense and antisense denote the upper and lower strand in C. **B:** The same fragments as in A were partially depurinated (A+G) or depyrimidated (C+T). After preparative band-shifting, protein-bound (B) and free (F) DNA was recovered, treated with piperidine and analyzed on a 10% polyacrylamide/urea gel. **C:** Summary of the premodification analyses. Autoradiograms from experiments in A and B were quantitated by densitometric scanning. Squares indicate methylated purines that reduce band intensity of the bound fraction to <50% (□) or <10% (■) compared to free DNA. The effect of missing bases on binding strength is given relative to the unbound fraction in the block diagram. Base pairs assigned by small characters in the −198 HPFH sequences denote vector DNA.

HEL cells revealed a footprint on the CACCC region from position −153 to −134 on the upper strand, in agreement with published data (5). Further upstream, a larger protected region extended approximately from position −208 to −184 (Fig. 5). From competition binding experiments, it was deduced that this footprint was generated by the closely spaced binding of the erythroid-specific factor GF-1 (7) and Sp1 or, alternatively, by binding of NF-G.C alone. Binding of the latter factor appeared to displace GF-1. No protected region was observed 5′ of the GF-1 footprint with the −196 HPFH mutation (data not shown).

Methylation interference assays

Since the bands of each doublet that represented complex B and C, respectively, reacted identically in competition band shift assays (Fig. 3), they were regarded as variants of a single polypeptide. Therefore these complexes were combined for the premodification analysis. Interference probing of the −198 HPFH and the γ-globin CACCC box fragments with wheat germ lectin-purified HEL extract, in which NF-G.C binding activity was removed but Sp1 activity retained (26), revealed a similar grouping of purines which when methylated decreased binding (Fig. 6A,C). The affected guanines on the bottom strands coincided with the consensus binding sequence of Sp1 (35):

GCC C

5′-ATTCCGCCCA-3′

(conventionally written in the reverse orientation). Furthermore, the interference pattern matched the data reported for Sp1 sites by others (36, 37). When the CACCC motif was probed for binding to a wheat germ lectin fraction of HeLa instead of HEL cells, an identical pattern was observed (data not shown). Analysis of complex C revealed a different type of interaction. Methylation in the block of purines from position −203 to −194 uniformly affected binding.

Missing contact probing

A method that extends the premodification interference approach to each individual base is missing contact probing (30), and, unlike the steric hindrance of an added methyl group, it detects protein contacts directly. The DNA probe is subjected to limited depurination or depyrimidation; then binding reaction and analysis of bases involved in protein contacts are performed as for the methylation interference method. The determination of contacts relevant for Sp1 binding to the −198 HPFH and the CACCC motif showed (Fig. 6B,C) that removal of a base at either strand in the region from −203 to −195 of CACCC strongly reduced binding. The marginal positions appeared to make one-sided contacts. Remarkably, missing contacts 5′ of the consensus sequence produced opposite effects on Sp1 binding in the −198 HPFH or CACCC motif. However, it should be cautioned that unambiguous contact assignment by base missing probing relies on the absence of conformational alterations due to a missing base. Sp1 binding could require or impose a conformational change in the DNA.

Base missing probing of complex C showed a pattern different from Sp1 binding. The base requirements for efficient binding coincided with the methylation interference data. Again, contact effects may be superseded by conformational features, especially in a sequence with the potential to adopt different conformations.

Effect of Sp1 binding to the CACCC motif on transcriptional activity in vitro

As an estimate of the biological significance of the binding data, we performed in vitro transcription analyses. First we wanted to know whether Sp1 binding to the CACCC motif was essential for promoter activity. Templates containing Gγ-globin sequences from −203 to +52 linked to a CAT reporter gene were incubated with nuclear extract from HEL cells and with a 200-fold molar excess of the DNA fragments used in the in vitro competition binding reactions. Fig. 7 shows that sequences which competed for Sp1 binding to the CACCC box also reduced transcriptional activity in vitro. Loss of Sp1 binding decreased promoter activity to a level comparable to a deletion construct which only contained sequences downstream of the CACCC box (J. Nowock, unpublished).

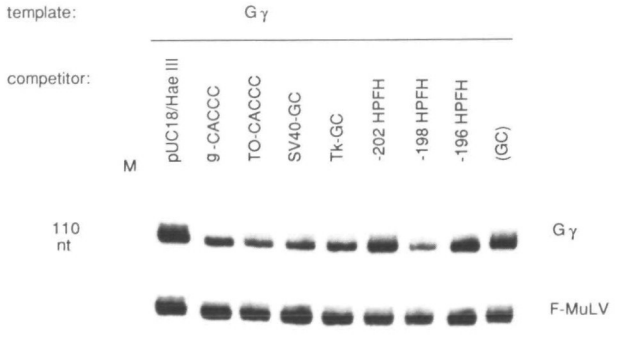

Figure 7. Importance of Sp1 binding to the CACCC motif for γ-globin gene promoter activity. In vitro transcription directed by plasmid p-203γ-CAT as template was performed in HEL extract in the presence of unspecific competitor DNA (HaeIII-digested pUC18) or of fragments containing GC-rich sequence motifs also used for the competition binding experiments in Fig. 3. A template containing the enhancer/promoter of F-MuLV, which is not responsive to Sp1, was included as internal standard. Transcriptional activity was evaluated by quantitative primer extension assay yielding expected reverse transcripts of 113 nt for γ-globin and 100 nt for F-MuLV.

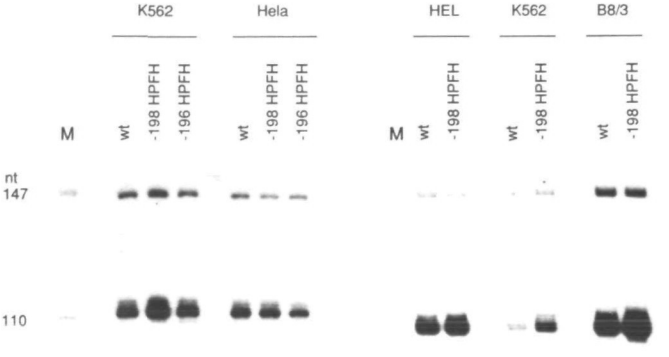

Figure 8. Effect of −198 and −196 HPFH mutations on γ-globin gene transcription in vitro. Plasmids p-383γ-CAT, p-198 HPFH-CAT and p-196 HPFH-CAT were transcribed in nuclear extracts from erythroid (K562, HEL, B8/3) and from HeLa cells. A CMV-CAT construct served as internal standard. Primer extension analysis of the synthesized RNA produced reverse transcripts of 113 nt for γ-globin and of 150 nt for the CMV templates. M denotes size markers of indicated lengths.

HPFH mutation at −198 increases transcriptional activity in vitro in erythroid extracts

Next we wanted to know whether the introduction of an additional cis-element by the −198 HPFH mutation increased transcriptional activity and whether this effect would lack cell specificity consistent with the ubiquitous nature of the cognate proteins. Wildtype, −198 HPFH and −196 HPFH templates containing Gγ-globin sequences from −383 to +52 were transcribed in vitro with extracts from erythroid (HEL, K562, B8/3) and from HeLa cells. The T→C substitution caused a 2.0- and 3.9-fold increase in promoter activity when tested with two different preparations of K562 extracts. The stimulation was 2.0- and 1.6-fold with extracts from HEL and B8/3 cells, respectively. The −196 mutation exhibited wildtype activity (Fig. 8). In nuclear extracts from non-erythroid cells transcription was not promoted by the −198 HPFH mutation (Fig. 8, 9). The increase in transcriptional activity by the −198 HPFH mutation is, though consistent, rather small when assayed in vitro. However, only a likewise moderate stimulation (4- to 5-fold) has been observed with transient transfection assays in a recent study by Ronchi et al. (42). The binding experiments had shown that two distinct proteins could interact with the −198 HPFH motif. In vitro transcription under conditions which compete out NF-G.C by poly(dG).poly(dC) did not significantly reduce promoter activity. This indicates that binding of Sp1 is sufficient to enhance transcription. Conditions which competed out Sp1 but allowed binding of NF-G.C or which removed both proteins resulted in the same decrease of transcriptional activity. Since in these cases the CCACCC box was also depleted of Sp1, an effect of protein binding to the −198 HPFH motif may not have been transmitted downstream (data not shown). That the activating potential of Sp1 on binding to the −198 HPFH site is only exerted by interaction with some erythroid-specific factor was suggested by the wildtype transcriptional activity with extracts from non-erythroid cells which contained Sp1 levels comparable to erythroid extracts (Fig. 8; see also Fig. 4).

Supplementation of nuclear extract from the myelomonocytic cell line WEHI-3B with affinity-purified Sp1 from HeLa or HEL cells did not result in a higher activity of the −198 HPFH promoter (Fig. 9). Thus there is no evidence for an erythroid-specific variant of Sp1. In contrast, addition of extract from K562 cells increased transcription directed by a wildtype template (p-383γ-CAT), and an additional slight promotion was observed with the −198 HPFH mutation. Supplementation with an affinity-purified fraction from HEL cells enriched in GF-1 yielded similar results, which suggests that transcriptional activation required binding of GF-1 to its upstream recognition site.

DISCUSSION

The independent occurrence in different populations of mutations associated with HPFH, mapping within 6 bp in the upstream region of the γ-globin genes, strongly suggests a causal connection between the base substitutions and the phenotype. The mechanism by which the downregulation of the fetal genes is prevented remains unclear. It was proposed for the −202 mutant that the base substitution increased the affinity for Sp1 (8). Although an increased binding compared to wildtype has been observed (43), the affinity is still 10-fold less than for the prototypical Sp1 site of SV40. Our protein-DNA binding data revealed only strong interactions with the −198 HPFH mutation. This T→C transition generated high affinity recognition sequences for two different proteins. One factor was indeed identified as Sp1 based on the following evidence: 1. In vitro binding could be competed with prototypical Sp1 recognition sequences; 2. the methylation interference pattern was in agreement with data published for Sp1 binding (36); 3. DNA binding required Zn^{++}; 4. purification by sequential wheat germ lectin and sequence-specific DNA affinity chromatography yielded the same binding complexes as crude nuclear extracts. This is in agreement with recent data of Ronchi et al. (42) who similarly showed Sp1 binding to the −198 HPFH mutation. The

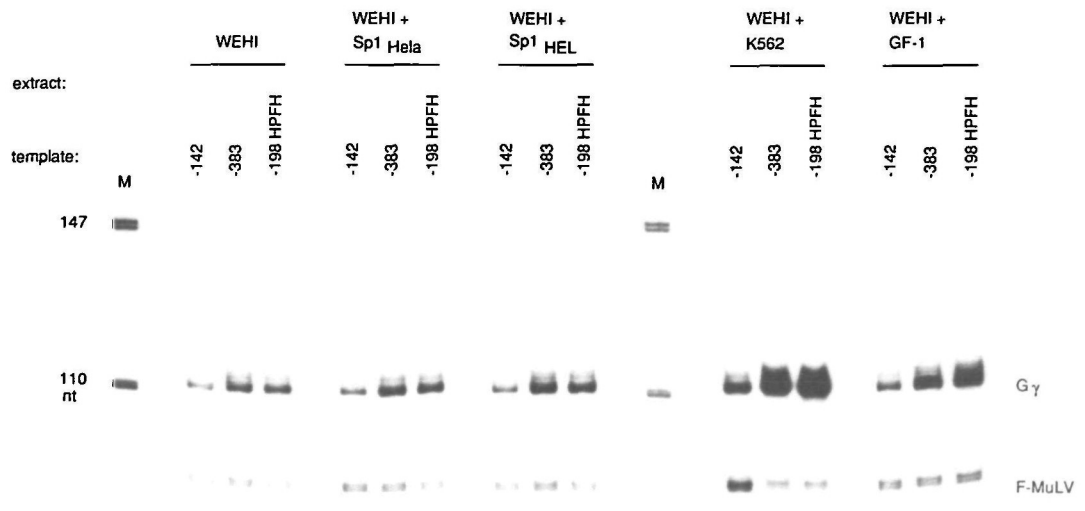

Figure 9. Supplementation of nuclear extract from WEHI-3B cells with various protein fractions. Plasmids p-142γ-CAT, containing a truncated promoter with ATAAA and CCAAT motifs, p-383γ-CAT and p-198 HPFH-CAT were transcribed with nuclear extract from the myelomonocytic cell line WEHI-3B. This extract was further supplemented with affinity-purified Sp1 from HeLa or HEL cells, with nuclear extract from K562 or with affinity-purified GF-1 from HEL cells. pF-MuLV-CAT served as internal standard. M denotes size markers of indicated length.

second protein that recognized the −198 HPFH sequence was termed NF-G.C owing to its high affinity to poly(dG).poly(dC). A protein with similar binding characteristics has been described (BGP1) which binds to a G-string in the chicken β-globin gene promoter (41). In contrast to the ubiquitous nature of the protein binding to −198 HPFH, BGP1 was erythroid-specific. However, the existence of a protein with homologous binding specificity and more general distribution has already been indicated (44). In vitro transcription data indicate that binding of Sp1 is sufficient for the −198 (T→C)-dependent transcriptional activation. Two observations argue against an involvement of NF-G.C: 1. Competitive depletion of NF-G.C does not abolish the −198 HPFH effect; 2. footprinting analyses (Fig. 6) showed that binding of NF-G.C displaces the erythroid-specific factor GF-1. However, transcriptional enhancement by the −198 HPFH mutation is not observed in non-erythroid extracts. In vitro transcription assays rather demonstrate the requirement for an erythroid factor, most likely GF-1, for the −198 HPFH effect to become apparent. An interaction between Sp1 and GF-1 is in agreement with the in vitro footprinting results.

Does the acquisition of an additional cis-element that is recognized by Sp1 or NF-G.C adequately explain the phenotypic expression of the −198 mutation? In vivo (42) and in vitro transcription experiments (this paper) yielded only a moderate increase in promoter activity, which falls short of the 20- to 30-fold increased level of fetal hemoglobin observed in adults with −198 HPFH. The British form of HPFH is thoroughly documented by hematological data. Hemoglobin analyses showed that fetal expression of HbF is normal. Further, adults with the −198 HPFH allele have a balanced synthesis of α- and β-type globin chains (45), i.e. the postnatal persistence of elevated γ-chain production is compensated by a reduced synthesis of β-chains. Thus the phenotype has to be explained in the context of the developmental regulation of the entire β-domain (46). These data question the simple interpretation that an increased transcriptional activity mediated by the acquisition of an additional Sp1 interaction generates the HPFH phenotype. Further, an explanation solely based on transcriptional enhancement does not provide a coherent model that integrates the −202 and −196 HPFH mutations which presumably affect the same cis-element.

A noteworthy feature of the region encompassing the HPFH mutations is the sequence arrangement itself. There is a 22 bp homopurine.homopyrimidine stretch (−215 to −194) containing a $(dG)_5.(dC)_5$ interspersion of reverse polarity (see Fig. 1). Homopurine.homopyrimidine sequences undergo structural transitions to non-B conformations under superhelical stress and/or low pH (47).

An unorthodox structure in the GC-rich region of the γ-globin gene was detected by S1 nuclease cleavage reactions. S1 recognizes, in addition to unpaired bases, conformational perturbations of the phosphodiester backbone. An S1-hypersensitive site was found centering at position −220 to −208 (10). In the −198 HPFH mutation, the cutting probability was slightly increased as compared to the wildtype sequence. In contrast, −202 C→G and −196 C→T substitutions reduced S1 sensitivity (K.-D. Fischer and J. Nowock, unpublished).

The biological role of this purine-pyrimidine structure in the perinatal switch of fetal globin is not clear, but we propose that a conformational alteration is associated with the downregulation of gene activity. When the structural transition is prevented, the promoter will remain active leading to a HPFH phenotype. This could be accomplished in two ways: 1. by mutations generating within this specific sequence a recognition site for DNA-binding proteins (−198 HPFH); the protein-DNA complexes would then stabilize the 'active' conformation; 2. by mutations which directly shift the conformation equilibrium to the 'active' state (−202, −196 HPFH). However, since part of the HPFH-associated mutations are conservative pyrimidine substitutions (see Fig. 1), other features in addition to a homopurine.homopyrimidine configuration have to be considered.

The β-globin locus is regulated by a region 5′ to the ε-gene, which was termed dominant control region (DCR) (48) or locus activation region (LAR) (49). DCR/LAR sequences act as cell-specific enhancers on transfection into erythroid cells (50, 51) or in transgenic mice (52). When linked to either the fetal or adult human globin gene, expression is not related to the stage of development. However, when DCR/LAR are linked to both genes in their normal genomic arrangement, developmental regulation is restored (46). Therefore it has been hypothesized that globin switching is controlled through a competitive interaction between DCR/LAR and regulatory elements of the γ- or β-globin gene, resulting in an exclusive expression at the respective developmental stages. Non-deletion HPFH mutations would then shift the competition equilibrium in an adult environment in favor of the fetal gene.

ACKNOWLEDGEMENTS

We thank Roland Schüle (Munich) for providing oligonucleotides with the TO-CACCC motif. We are grateful to Gökhan Arman-Kalcek for preparing the figures. This work is part of the doctoral thesis of K.-D.F. and was supported by the Deutsche Forschungsgemeinschaft Grant No. 89/5. The Heinrich-Pette-Institut is financially supported by Freie und Hansestadt Hamburg and by Bundesministerium für Jugend, Familie, Frauen und Gesundheit.

REFERENCES

1. Stamatoyannopoulos, G. and Nienhuis, A.W. (1987) In Stamatoyannopoulos, G., Nienhuis, A.W., Leder, P. and Majerus, P.W. (eds), The Molecular Basis of Blood Diseases. W.B. Saunders, Philadelphia.
2. Waber, P.G., Bender, M.A., Gelinas, R.E., Kattamis, C., Karaklis, A., Sofroniadou, K., Stamatoyannopoulos, G., Collins, F.S., Forget, B.G. and Kazazian, H.H., Jr. (1986) Blood, **67**, 551−554.
3. Surrey, P., Delgrosso, K., Malladi, P. and Schwartz, E. (1988) Blood, **71**, 807−810.
4. Superti-Furga, G., Barberis, A., Schaffner, G. and Busslinger, M. (1988) EMBO J., **7**, 3099−3107.
5. Gumucio, D.L., Rood, K.L., Gray, T.A., Riordan, M.F., Sartor, C.I. and Collins, F.S. (1988) Mol. Cell. Biol., **8**, 5310−5322.
6. Stoeckert, C.J., Jr., Metherall, J.E., Yamakawa, M., Weissman, S.M. and Forget, B.G. (1987) Mol. Cell. Biol., **7**, 2999−3003.
7. Martin, D.I., Tsai, S.-F. and Orkin, S.H. (1989) Nature, **338**, 435−438.
8. Collins, F.S., Stoeckert, C.J., Jr., Serjeant, G.R., Forget, B.G. and Weissman, S.M. (1984) Proc. Natl. Acad. Sci. USA, **81**, 4894−4898.
9. Giglioni, B., Casani, C., Mantovani, R., Merli, S., Comi, P., Ottolenghi, S., Saglio, G., Camaschella, C. and Mazza, U. (1984) EMBO J., **3**, 2641−2645.
10. Tate, V.E., Wood, W.G. and Weatherall, D.J. (1986) Blood, **68**, 1389−1393.
11. Gilman, J.G., Mishima, N., Wen, X.J., Kutlar, F. and Huisman, T.H.J. (1988) Blood, **72**, 78−81.
12. Gelinas, R., Bender, M., Lotshaw, C., Waber, P., Kazazian, H., Jr. and Stamatoyannopoulos, G. (1986) Blood, **67**, 1777−1781.
13. Ottolenghi, S., Giglioni, B., Pulazzini, A., Comi, P., Camaschella, C., Serra, A., Guerrasio, A. and Saglio, G. (1987) Blood, **69**, 1058−1061.
14. Wells, R.D. and Harvey, S.C. (1988) Unusual DNA Structures. Springer Verlag, New York.

15. Shen, S., Slightom, J.L. and Smithies, O. (1981) Cell, **26,** 191–203.
16. Miksicek, R., Borgmeyer, U. and Nowock, J. (1987) EMBO J., **6,** 1355–1360.
17. Koch, W., Zimmermann, W., Oliff, A. and Friedrich, R. (1984) J. Virol., **49,** 829–840.
18. Lozzio, C. and Lozzio, B. (1975) Blood, **45,** 321–334.
19. Martin, P. and Papayannopoulou, T. (1982) Science, **216,** 1233–1235.
20. Papayannopoulou, T., Nakamoto, B., Yokochi, T., Chait, A. and Kannagi, R. (1983) Blood, **62,** 832–845.
21. Ostertag, W., Roesler, G., Krieg, C.J., Kind, J., Cole, T., Crozier, T., Gaedicke, G., Steinheider, G., Kluge, N. and Dube, S. (1974) Proc. Natl. Acad. Sci. USA, **71,** 4980–4985.
22. Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986) Nature, **323,** 445–448.
23. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res., **11,** 1475–1489.
24. Gorski, K., Carneiro, M. and Schibler, U. (1986) Cell, **47,** 767–776.
25. Thornell, A., Hallberg, B. and Grundström, T. (1988) Mol. Cell. Biol., **8,** 1625–1637.
26. Jackson, S.P. and Tjian, R. (1989) Proc. Natl. Acad. Sci. USA, **86,** 1781–1785.
27. Kadonaga, J.T. and Tjian, R. (1986) Proc. Natl. Acad. Sci. USA, **83,** 5889–5893.
28. Siebenlist, U. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA, **77,** 122–126.
29. Sturm, R., Baumruker, T., Franza, B.R., Jr. and Herr, W. (1987) Genes Dev., **1,** 1147–1160.
30. Brunelle, A. and Schleif, R.F. (1987) Proc. Natl. Acad. Sci. USA, **84,** 6673–6676.
31. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzymol., **65,** 499–560.
32. Foecking, M.K. and Hofstetter, H. (1986) Gene, **45,** 101–105.
33. Birnboim, H.D. (1988) Nucleic Acids Res., **16,** 1487–1497.
34. Südhoff, T.C., van der Westhuyzen, D.R., Goldstein, J.L., Brown, M.S. and Russell, D.W. (1987) J. Biol. Chem., **262,** 10773–10779.
35. Kadonaga, J.T., Jones, K.A. and Tjian, R. (1986) TIBS, **11,** 20–23.
36. Evans, T., DeChiara, T. and Efstratiadis, E. (1988) J. Mol. Biol., **199,** 61–81.
37. Gustafson, T.A. and Kedes, L. (1989) Mol. Cell. Biol., **9,** 3269–3283.
38. McKnight, S.L. (1982) Cell, **31,** 355–365.
39. Xiao, J.-H., Davidson, I., Macchi, M., Rosales, R., Vigneron, M., Staub, A. and Chambon, P. (1987) Genes Dev., **1,** 794–807.
40. Schüle, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988) Nature, **332,** 87–90.
41. Lewis, C.D., Clark, S.P., Felsenfeld, G. and Gould, H. (1988) Genes Dev., **2,** 863–873.
42. Ronchi, A., Nicolis, S., Santora, C. and Ottolenghi, S. (1989) Nucleic Acids Res., **17,** 10231–10241.
43. Sykes, K. and Kaufman, R. (1990) Mol. Cell. Biol., **10,** 95–102.
44. Emerson, B.M., Nickol, J.M. and Fong, T.C. (1989) Cell, **57,** 1189–1200.
45. Wood, W.G., MacRae, I.A., Darbre, P.D., Clegg, J.B. and Weatherall, D.J. (1982) Br. J. Haematol., **50,** 401–414.
46. Enver, T., Raich, N., Ebens, A.J., Papayannopoulou, T., Costantini, F. and Stamatoyannopoulos, G. (1990) Nature, **344,** 309–313.
47. Pulleyblank, D.A., Haniford, D.B. and Morgan, A.R. (1985) Cell, **42,** 271–280.
48. Grosveld, F., van Assendelft, G.B., Greaves, D. and Kollias, G. (1987) Cell, **51,** 975–985.
49. Forrester, W.C., Takegawa, S., Papayannopoulou, T. and Stamatoyannopoulos, G. (1987) Nucleic Acids Res., **15,** 10159–10177.
50. Tuan, D.Y.H., Solomon, W.B., London, I.M. and Lee, D.P. (1989) Proc. Natl. Acad. Sci. USA, **86,** 2554–2558.
51. Collis, P., Antoniou, M. and Grosveld, F. (1990) EMBO J., **9,** 233–240.
52. van Assendelft, G.B., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989) Cell, **56,** 969–977.