Expression of the \(\gamma\)-globin gene is silenced in adult humans. However, certain point mutations in the \(\gamma\)-globin gene promoter are capable of maintaining expression of this gene during adult erythropoiesis, a condition called non-deletion hereditary persistence of fetal hemoglobin (HPFH). Among these, the British form of HPFH carrying a \(T\rightarrow C\) point mutation at position \(-198\) of the \(\gamma\)-globin gene promoter results in \(4-10\%\) fetal hemoglobin in heterozygotes. In this study, we used nuclear extracts from murine erythroleukemia cells to purify a protein complex that binds the HPFH \(-198\) \(\gamma\)-globin gene promoter. Members of this protein complex were identified by mass spectrometry and include DNMT1, the transcriptional coactivator p52, the protein SNEV, and RAP74 (the largest subunit of the general transcription factor IIF). Sp1, which was previously considered responsible for HPFH \(-198\) \(\gamma\)-globin gene activation, was not identified. The potential role of these proteins in the reactivation and/or maintenance of \(\gamma\)-globin gene expression in the adult transcriptional environment is discussed.

The human \(\beta\)-globin gene cluster consists of five functional globin genes (\(\epsilon, \xi, \upsilon, \delta,\) and \(\beta\)) arranged in the locus according to the order of their expression during development. Genetic and biochemical evidence indicates that the expression of these genes during development depends on interactions between the individual globin gene promoters and the locus control region located \(6-22\) kb upstream of the \(\epsilon\)-globin gene (1). Two developmental expression switches occur in the \(\beta\)-globin locus: one from the embryonic \(\epsilon\) gene to the fetal \(\gamma\) genes and later from the fetal \(\gamma\) genes to adult \(\delta\) and \(\beta\) genes. Adult individuals express very low levels of the \(\gamma\)-globin genes (usually \(-0.5\%\) of the total hemoglobin). However, single point mutations occurring in either the \(\xi\), \(\upsilon\), \(\delta\), or \(\beta\)-globin gene promoter result in continued expression of the \(\gamma\) gene in the adult, a condition termed non-deletion hereditary persistence of fetal hemoglobin (HPFH) (reviewed in Refs. 1 and 2). Structural studies have shown that non-deletion HPFH point mutations are clustered in three regions of the \(\gamma\) gene promoters centered around positions \(-200, -175,\) and \(-115\) relative to the transcriptional start site (2). The \(-200\) region is a highly GC-rich region known to be the target for five different but closely spaced point mutations affecting the \(\xi\)-\(\gamma\) promoter at position \(-202\) (\(C\rightarrow G\)) and the \(\upsilon\)-\(\gamma\) promoter at position \(-202\) (\(C\rightarrow T\)), \(-198\) (\(T\rightarrow C\)), \(-196\) (\(C\rightarrow T\)), or \(-195\) (\(C\rightarrow G\)), respectively (1, 2). Two hypotheses have been proposed to explain the increased \(\gamma\)-globin gene expression in adults carrying non-deletion HPFH mutations. The first proposes that these point mutations decrease the binding of a transcriptional repressor or a complex that is involved in the silencing of \(\gamma\)-globin expression in the adult. Alternatively, these mutations may create binding sites that enhance the binding for a transcriptional activator or complex, thus increasing \(\gamma\) gene expression in the adult. Early in vitro studies focused on characterizing the effects of these mutations on the binding of different DNA-binding proteins. The sequence similarity between the HPFH \(-198\) mutation and the cognate Sp1 response element prompted in vitro studies, which suggested that this mutation increases the binding affinity for the transcriptional activator Sp1 (3–5). At least two other unidentified factors capable of binding to this region were also identified.

Transgenic mice carrying the \(-117, -175,\) and \(-198\) mutations display the phenotype of HPFH, providing direct evidence for the mechanistic relationship between mutation and phenotype (6–8). We have shown previously that the HPFH \(-198\) mutation is able to retain \(\gamma\) gene expression in adult transgenic mice when the CACCC box is disrupted (9). Because the CACCC box is indispensable for \(\gamma\) gene expression in the adult, these results suggested that the HPFH \(-198\) mutation creates a new element that is able to substitute for the function of the CACCC box in adults. This study focused on the biochemical purification and characterization of proteins that bind to the \(\upsilon\)-\(\gamma\)-globin gene carrying the HPFH \(-198\) (\(T\rightarrow C\)) point mutation using non-biased methods. Chromatography and mass spectrometry identified a group of proteins that specifically eluted from an HPFH \(-198\) oligonucleotide affinity column. Using commercially available antibodies, we were able to confirm the identity and HPFH \(-198\) binding specificity of a subset...
of these proteins, including DNMT1 (DNA methyltransferase 1), RAP74 (the largest subunit of the general transcription factor (TF) IIIP), and the coactivator p52. Sp1 was not found in this complex. The implication and potential roles of these proteins are discussed in the context of regulating the expression of the \( \gamma \)-globin gene in adults.

**EXPERIMENTAL PROCEDURES**

**Gel Shift Assays**—The DNA oligonucleotide sequences used in this work are shown in Fig. 1A. Position −198 and the CACCC box present in the human \( \Lambda \ \gamma \) promoter are underlined, and their mutant counterparts are depicted in italic type. Radiolabeled double-stranded oligonucleotide probes (1.5 \( \times \) 10^4 cpm, \( \sim \)4 fmol) were incubated with 3–6 \( \mu \)g of either murine erythroleukemia (MEL) cell nuclear extract or protein from column fractions for 20 min at room temperature in binding buffer (20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM dithiothreitol, 12.5% glycerol, 5 mM MgCl\(_2\), 0.05% Nonidet P-40, and 0.3 \( \mu \)g of oligo(dI-dC)\(_{20}\)). When necessary, a 100-fold excess of unlabeled double-stranded competitor oligonucleotides or the amount of antibodies indicated in the figures was added and preincubated with proteins under similar conditions prior to adding the probe. Samples were subjected to electrophoresis on a 5% polyacrylamide gel containing 5 mM MgCl\(_2\) in 1\( \times\) Tris borate/EDTA buffer at 4 °C, followed by autoradiography.

**Nuclear Extract Preparation, Protein Fractionation, and Western Blotting**—Nuclear extracts were prepared from MEL cells as described previously (10) with modifications. Briefly, \( \sim1.8 \times 10^{10} \) logarithmic phase MEL cells were harvested and washed twice with cold phosphate-buffered saline, followed by washing with a 5-fold packed cell volume of hypotonic buffer (10 mM HEPES-KOH (pH 7.9), 1.5 mM MgCl\(_2\), 10 mM KCl, 0.5 mM dithiothreitol, and 0.2 mM EDTA) supplemented with protease inhibitor mixture (phenylmethylsulfonyl fluoride, pepstatin, leupeptin, bestatin, and aprotinin). MEL cells were suspended in a 3-fold packed cell volume of hypotonic buffer plus protease inhibitor mixture, incubated on ice for 10 min, and homogenized by 10 strokes of a Dounce type A pestle. The suspension was centrifuged at 3300 \( \times \) g for 15 min at 4 °C, suspended in a 3.5-fold packed cell volume of ice-cold nuclear suspension buffer (10 mM HEPES-KOH (pH 7.9), 3 mM MgCl\(_2\), 100 mM KCl, 0.1 mM EDTA, and 0.5 mM dithiothreitol) plus protease inhibitor mixture, and lysed by 15 strokes of a Dounce type A pestle. The suspension was complemented with the dropwise addition and mixing of 0.1 volume of ice-cold 3 M ammonium sulfate (pH 7.5) and incubated at 4 °C for 30 min with rocking. Chromatin was precipitated by centrifugation at 65,000 rpm for 1 h at 4 °C using a Beckman Ti-70 rotor. Proteins present in the supernatant were precipitated with a 65% cut using solid ammonium sulfate. Pelleted proteins were suspended in an \( \sim1\)-fold packed cell volume of ice-cold buffer C (20 mM HEPES-KOH (pH 7.8), 0.2 mM EDTA, 0.25 mM phenylmethylsulfonyl fluoride, and 15% glycerol) containing 100 mM KCl (BC100) and supplemented with protease inhibitor mixture, dialyzed overnight at 4 °C against the same buffer, and stored at −70 °C for further use.

Fractionation of MEL cell nuclear extract was done at 4 °C unless indicated otherwise. Approximately 500 mg of protein was loaded onto a packed column containing phosphocellulose resin (120 ml) previously equilibrated with BC100. The flow-through fraction was collected; the resin was washed extensively with BC100; and bound proteins were eluted in three batches with buffer C containing 300 (BC300), 500 (BC500), and 1000 (BC1000) mM KCl, respectively. For each elution, fractions (2.6 ml each) were collected, and their protein concentrations were estimated by Bradford assay (Bio-Rad). Fractions containing >0.1 \( \mu \)g/ml protein were pooled and dialyzed against BC100 prior to analysis of their HPFH −198 binding activity by gel shift assay.

We packed Source Q resin (GE Healthcare) into a Tricorn column (GE Healthcare) following the manufacturer’s instructions to create our own ÄKTA-compatible Source Q column (9.5 ml of resin). The pooled 0.5 M phosphocellulose fraction (\( \sim100\) ml, 54 mg of total protein) was loaded onto this column, which was previously conditioned with BC100. The flow-through fraction was collected, and the column was washed with the same buffer. Bound proteins were eluted with BC1000 using a 21-column volume three-step programmed gradient consisting of 15 column volumes of 0–40% buffer C, 3 column volumes of 40–100% buffer C, and 3 column volumes of 100% buffer C. The collected fractions (2 ml) were dialyzed and analyzed for HPFH −198 binding activity in gel shift assays before they were pooled.

Protein markers from a size exclusion chromatography kit (GE Healthcare) were used to calibrate a Superdex S-200 HR 10/30 gel filtration column (23.56 ml; GE Healthcare) under stringent buffer conditions (BC500 containing 0.1% Nonidet P-40) using an ÄKTA chromatography system (GE Healthcare). Nuclear extract from MEL cells (1 mg, 7.6 \( \mu \)g/ml) was fractionated in this column under the same buffer conditions. Fractions (0.5 ml each) were collected, dialyzed against BC100, and stored at −70 °C for further analysis.

Aliquots derived from fractions from either the gel filtration (20 \( \mu \)l) or affinity chromatography (200–500 \( \mu \)l precipitated, see below) column were analyzed by Western blotting. Briefly, samples were separated on an 11% mini SDS-polyacrylamide gel, transferred to nitrocellulose, and blocked for 1 h at room temperature with 5% nonfat milk in 0.3 M Tris/Tween-buffered saline (20 mM Tris-HCl (pH 7.5), 0.1% Tween 20, 300 mM NaCl). Blots were washed with the same buffer and incubated overnight at 4 °C with antibody against DNMT1 (1:5000 dilution; Abcam Inc., Cambridge, MA), coactivator p52 (1:1000 dilution; ProteinOne, Bethesda, MD), or RAP74 (1:700 dilution; ProteinOne) with rocking.

Membranes were washed with 0.3 M Tris/Tween-buffered saline and incubated with goat anti-rabbit antibodies conjugated with horseradish peroxidase (1:2500 dilution; Santa Cruz Biotechnology, Inc., Santa Cruz, CA) for 1.5 h. Finally, membranes were washed with 0.3 M Tris/Tween-buffered saline, and specific proteins were detected on x-ray films using an ECL chemiluminescence kit (Amersham Biosciences).

**Preparation of Affinity Resins, Tandem Oligonucleotide Affinity Purification, and Mass Spectrometry Identification of Proteins**—The following forward (F) and reverse (R) phosphorylated single-stranded oligonucleotides comprising the human...
Proteins That Bind to the γ-Globin HPFH – 198 Mutation

wild-type (WT) or HPFH – 198 (T → C) point mutation γ-globin sequence were used for the preparation of WT and HPFH – 198 oligonucleotide affinity resins: 5′-phos-GATCTTTTAGGGGCCCCCTCCCACATATATTGGGAAAGGAGGGAGGGAGGGAGGGCCCT-3′ (WT-R), 5′-phos-GATTCTTTTAGGGGCCCCCTCCCACATATATTGGGAAAGGAGGGAGGGAGGGCCCT-3′ (WT-R), 5′-phos-GATCTTTTAGGGGCCCCCTCCCACATATATTGGGAAAGGAGGGAGGGAGGGCCCT-3′ (WT-R). The only difference in these sequences is the underline point mutation at position – 198. These oligonucleotides are identical to those used for the gel shift assays shown in Fig. 1A. An additional 9-nucleotide overhang sequence was added at each end to create sticky ends once they were annealed. Online data base searches indicated that this 9-bp extension did not harbor any potential DNA-binding sequence.

Single-stranded oligonucleotides were purified by denaturing gel electrophoresis and annealed to form complementary WT and HPFH – 198 double-stranded oligonucleotides. Five-hundred μg of each was then ligated with T4 DNA ligase (New England Biolabs) to form concatamers (10 copies on average), which were then coupled to 10 ml of CNBr-activated Sepharose 4B beads following published protocols (49) to create the WT and HPFH – 198 oligonucleotide affinity columns.

Before chromatography, the pooled Source Q protein fraction was supplemented with 5 mM MgCl2, 0.05% Nonidet P-40, and 3 μg/ml oligo(dI-dC), thus bringing the buffer conditions similar to those used in gel mobility shift assays. This mixture was incubated on ice for 15 min and centrifuged at 14,000 rpm at 4 °C for 10 min, and the supernatant was used as the input material for the tandem oligonucleotide affinity chromatography step. Usually 2 ml of input material was gently rocked for 30 min at room temperature in a column containing an equal volume of packed WT oligonucleotide resin previously equilibrated with affinity buffer (20 mM HEPES-KOH (pH 7.8), 1 mM dithiothreitol, 5 mM MgCl2, 0.05% Nonidet P-40, and 10% glycerol) containing 100 mM KCl (AF100). The column was placed on a stand, and the flow-through fraction was collected on ice by gravity and reapplied to the column nine more times at room temperature. The last flow-through fraction was collected; the WT – 198 oligonucleotide column was washed with 10 column volumes of AF100; and the bound proteins were eluted with 1.5 column volumes of affinity buffer containing 1000 mM KCl (AF1000) and 2000 mM KCl (AF2000), respectively. The last flow-through fraction from the WT oligonucleotide column was combined with its first column volume wash and used as the input material for the HPFH – 198 oligonucleotide column. After a similar incubation and reapplication process, the flow-through fraction from the HPFH – 198 oligonucleotide column was collected; the resin was washed with 10 column volumes of AF100; and the bound proteins were eluted with 1.5 column volumes of AF500, AF1000, and AF2000, respectively. Collected fractions were dialyzed against BC100 and frozen at −70 °C.

To visualize the eluted proteins from the tandem oligonucleotide affinity chromatography step, 150–250-μl aliquots of each of the column fractions were precipitated overnight at −20 °C with 4 volumes of cold acetone in the presence of 10 μg of human recombinant insulin (Sigma) as a carrier. This mixture was centrifuged at 14,000 rpm for 15 min at 4 °C, and the pelletted proteins were washed twice with 1 ml of cold acetone, dried at 37 °C, and suspended in 20 μl of suspension buffer (50 mM Tris-HCl (pH 8.0), 1% SDS, and 5% glycerol). Samples were complemented with 6× SDS loading buffer and run on an 11% mini SDS-polyacrylamide gel, and proteins were either transferred to nitrocellulose for Western blot analysis (see above) or silver-stained following the kit instructions (Bio-Rad).

To identify proteins in the HPFH – 198 oligonucleotide column fractions, bands were excised from the silver-stained gel, processed, and in gel-digested with trypsin (Roche Applied Science) following described protocols (11, 12). The peptides obtained were identified with a capillary liquid chromatography-atmospheric pressure ionization quadrupole orthogonal accelerator time-of-flight hybrid tandem mass spectrometer (Micromass Ltd., Manchester, UK). Online data mining was done with the Matrix Science Mascot tandem mass spectrometry (MS/MS) ion search algorithm (www.matrixscience.com) using both the mouse and NCBI non-redundant protein data bases.

RESULTS

HPFH – 198 Binding Activity Is Related to the CACCC Box but Differs from Sp1—We performed gel shift assays to analyze the binding specificity of the HPFH – 198 probe using MEL cell nuclear extract. There were three major retarded bands (Fig. 1B, lane 1, arrows a–c) that were competed away by the unlabeled HPFH – 198 oligonucleotide (lane 4), but not by the WT – 198 oligonucleotide (lane 5). These bands were partially competed away with a 100-fold excess of unlabeled WT CACC oligonucleotide (lane 2), but not by the mutant CACC oligonucleotide (lane 3), suggesting that the proteins forming these retarded bands are associated with the CACCC box, as reported previously (3–5, 9). Below the three bands, there were several fast migrating bands (bracket d). The occurrence and intensity of these bands varied in different MEL cell extracts. In addition, these bands were not competed away by unlabeled oligonucleotides, including HPFH – 198 (lanes 2–6), suggesting that they are nonspecific binding products.

A number of reports have characterized the transcriptional activator Sp1 as a potential factor binding to the HPFH – 198 mutation (3–5). We were able to reproduce these results by demonstrating that unlabeled Sp1 oligonucleotide competed away three major complexes as shown in Fig. 1B (lane 6, arrows a–c). However, we found that anti-Sp1 antibody had no effect on the retarded bands generated on the HPFH – 198 probe (lane 8). A similar outcome was observed when anti-IgG control antibody was used (lane 9). Used as a positive control, the same anti-Sp1 antibody clearly supershifted the retarded band formed when human recombinant Sp1 protein bound to the Sp1 probe (compare lanes 10 and 11). Titration of either the nuclear extract or the antibodies under these conditions resulted in similar results (data not shown). The presence of Sp1 in MEL cell extracts was confirmed by Western hybridization (Fig. 1C). MEL cell extracts generated strong signals in the immunoblot assay using anti-Sp1 antibody (Fig. 1C, lanes 1–4). Recombinant Sp1 served as a positive control (lane 5). Moreover, the Sp1 abundance decreased during the purification process. For instance, only a small amount of Sp1 was detected in
Proteins That Bind to the γ-Globin HPFH −198 Mutation

Source Q column fractions (lanes 6 and 7). Thus, these experiments strongly suggest that the proteins that binds specifically to the HPFH −198 probe (arrows a–c) are different from Sp1. Moreover, these proteins are able to bind the Sp1 consensus motif, whereas Sp1 protein is unable to bind the HPFH −198 oligonucleotide.

Purification Profile and MEL Cell Nuclear Extract Fractionation—Fig. 2A outlines the scheme of purification of the HPFH −198 mutation-binding proteins. Approximately 0.5 g of MEL cell nuclear extract was fractionated by column chromatography using a phosphocellulose resin. The flow-through fraction was collected; the resin was washed extensively with BC100; and bound proteins were eluted into three batches with BC300, BC500, and BC1000, respectively. Fig. 2B shows the HPFH −198 gel shift activity of these batch-eluted fractions after dialysis against BC100. As shown at the top of Fig. 2B, the activity of each batch was then competed by a 100-fold excess of either unlabeled HPFH or WT −198 oligonucleotide as described previously for Fig. 1A. Most of the HPFH −198 binding activity resided in the fractions that eluted at 0.5 and 1.0 M. After four independent chromatography fractionation experiments, the percentage of the total nuclear protein found in the flow-through and 0.3, 0.5, and 1.0 M fractions was 39, 47.1, 11.6, and 2.3%, respectively.

The 0.5 M phosphocellulose fraction was then loaded into an ÄKTA Source Q column and fractionated using a triphasic isocratic gradient. The elution profile is shown in Fig. 2C. Column fractions were dialyzed against BC100, and their HPFH −198 binding activity was analyzed by gel shift assay (Fig. 2D). We used a shallow gradient and found that the activity consistently eluted between 150 and 370 mM KCl, thus explaining the wide spread among the collected fractions. After quantitative protein analysis of each fraction and considering the elution and activity profiles (Fig. 2, C and D), we pooled only fractions 27–40, thus sacrificing some HPFH −198 binding activity for protein purity. The pooled activity thus represents 8% of the total protein loaded onto the Source Q column and ~1% of the starting MEL cell nuclear extract.

To take advantage of the single point mutation difference between the WT and HPFH −198 oligonucleotides, we designed a tandem oligonucleotide affinity purification step. Following incubation of the Source Q pool with the concatenated WT −198 oligonucleotide resin, the unbound fraction was combined with the first column volume wash and used as the input material for HPFH −198 oligonucleotide affinity chromatography. The WT resin was extensively washed, and bound proteins were eluted in two steps with 1.5 volumes of AF1000 and AF2000, respectively. Similarly, after incubation with the HPFH −198 resin, the flow-through fraction was collected; the resin was washed extensively; and bound proteins were eluted in 1.5 volumes of AF500, AF1000, and AF2000, respectively. The bulk of the HPFH −198 mutation-bound proteins were present in the first two fractions (see Fig. 4, upper panel), which were combined and dialyzed against BC100. Proteins present in aliquots (200–500 µl) of this combined fraction were precipitated with acetone in the presence of human recombinant insulin (10 µg) and separated on an 11% SDS-polyacrylamide gel. Fig. 3 is a representative silver-stained gel showing the proteins that eluted from the HPFH −198 oligonucleotide column, those in the input material, and molecular mass markers. Bands present in the HPFH −198 lane were excised; proteins were digested with trypsin; and peptides were

| Peptide          | NCBI accession no. | Molecular mass | No. of peptides identified | Protein coverage |
|------------------|--------------------|----------------|---------------------------|-----------------|
| DNMT1            | AAH53.047          | 183,174        | 60                        | ~40             |
| CDC5-like        | NP690023           | 92,133         | 6                         | ~10             |
| RAP74            | NP598562           | 57,206         | 6                         | ~11             |
| SNEV             | NP598890           | 55,204         | 4                         | ~7              |
| p52              | CAC34945           | 37,403         | 4                         | ~13             |
| Unnamed BAB28490 |                    | 37,548         | 5                         | ~18             |

TABLE 1

Peptides identified in the putative HPFH −198 complex purified using an affinity column
identified using the liquid chromatography-atmospheric pressure ionization quadrupole orthogonal accelerator time-of-flight hybrid tandem mass spectrometer. Data mining was done online with the Matrix Science Mascot MS/MS ion search algorithm using both the mouse and NCBI non-redundant protein data bases. Bands identities are shown on the right of the gel and are summarized in Table 1.

Peptides Identified by Mass Spectrometry—The highest molecular mass protein (183 kDa) that bound to the HPFH\(^{198}\) oligonucleotide was identified as DNMT1 in seven independent experiments. Mass spectrometry detected a total of 60 unique peptides identified to be portions of DNMT1, representing \(\sim 40\%\) of its amino acid sequence. DNMT1 has been implicated in maintaining methylation patterns established during development and newly synthesized DNA at replication foci in eukaryotes (13, 14). DNMT1 does not bind DNA directly, but is most likely recruited via protein-protein interactions.

The protein band migrating at \(\sim 100\) kDa was identified in four independent experiments as mouse CDC5-like protein (Fig. 3). We based our identification on six different peptides covering \(\sim 10\%\) of its amino acid sequence. This protein is an ortholog of the G2/M cell cycle regulator protein Cdc5 from *Schizosaccharomyces pombe*.

The third protein identified in two independent experiments with a total of six peptides covering \(\sim 10\%\) of its amino acid sequence was RAP74, the largest subunit of TFIIF. Human RAP74 associates with the smaller subunit RAP30 to form a tetramer and, as such, associates with RNA polymerase II (15, 16).

The mouse nuclear matrix protein SNEV, which migrated as a ghost band (i.e. negatively stained with silver) in the gel at \(\sim 55\) kDa, was identified as the forth protein binding to the HPFH.
Proteins That Bind to the γ-Globin HPFH −198 Mutation

FIGURE 4. Western blot analysis of DNMT1, RAP74, and p52 in eluates from an affinity chromatography column. Aliquots of eluted fractions from both the WT −198 (lanes 2 and 3) and HPFH −198 (lanes 4–7) oligonucleotide affinity chromatography steps (shown at the top) were acetone-precipitated; separated on an 11% SDS-polyacrylamide gel, transferred to nitrocellulose; and analyzed by Western blotting using commercially available antibodies for DNMT1, RAP74, and p52. Arrows indicate the specific reactive bands. Migration of molecular mass markers (M) is shown in lane 8, and the input material is shown in lane 1. Both DNMT1 and p52 bound specifically to the HPFH −198 oligonucleotide column, whereas RAP74 had less specificity. FT, flow-through fraction.

FIGURE 5. Anti-DNMT1 and anti-RAP74 antibodies specifically inhibit HPFH −198 binding activity. Prior to addition of the HPFH −198 probe, MEL cell nuclear extract was preincubated with 1, 2, or 4 µg of anti-DNMT1 (lanes 2–4), anti-IgG (lanes 5–7), anti-RAP74 (lanes 8–10), or anti-p52 (lanes 11–13) antibody as indicated. Lane 1 shows the control in the absence of antibody. Arrows a–c point to the three retarded bands specifically formed with the HPFH −198 probe, and arrow (d) indicates a new band that appeared after addition of antibody (lanes 3 and 4).

−198 mutation. Four different polypeptides covering −7% of this protein were identified in five independent experiments. SNEV is 99% identical to the human nuclear matrix protein NMP200 (17), which is in turn related to the human splicing factor PRP19, thus potentially implicating SNEV in splicing.

The fifth protein identified in two independent experiments with four different peptides covering ∼13% of the protein was the coactivator p52. This protein is derived from an alternatively spliced 15-exon gene encoding both p52 and a larger protein called p75/lens epithelium-derived growth factor (18, 19). The C terminus of p52 is highly charged and shows some similarity to human HMG1, a non-histone multifunctional protein involved in different aspects of gene regulation (20).

The last protein identified was an unnamed mouse protein of unknown function (NCBI accession number BAB28490) that migrated at ∼44 kDa on a denaturing SDS gel (Fig. 3). Five different polypeptides were identified by MS/MS in three independent affinity chromatography purifications covering ∼18% of this protein. Attempts to find more information using online BLAST searches against the NCBI non-redundant protein database and motif searches using different browser algorithms were unsuccessful.

Western Blot Analysis of Affinity-purified WT and HPFH −198 Oligonucleotide Column Fractions—To confirm both the identification of the proteins described above and their binding specificity for the HPFH −198 mutation, we acetone-precipitated the eluted fractions from the WT and HPFH −198 oligonucleotide affinity chromatography and analyzed the presence of some of these proteins in these fractions by Western blotting. Fig. 4 shows the results using commercially available antibodies for DNMT1, RAP74, and the coactivator p52. Both DNMT1 and p52 were specifically eluted under high salt conditions from the HPFH −198 mutant oligonucleotide column (lanes 5 and 6), but not from the WT −198 oligonucleotide column (lanes 2 and 3), confirming their MS/MS identification and binding specificity for the HPFH −198 mutation. On the other hand, RAP74 seemed to bind equally well to both the WT and HPFH −198 oligonucleotide columns because it was found in the eluates of both columns (lanes 2 and 3 and lanes 5 and 6, respectively), suggesting a potential lack of binding specificity. Independent of this observation, these results confirm the MS/MS identification and binding specificity for the DNMT1, RAP74, and p52 proteins for the HPFH −198 mutation, with RAP74 having the lowest specificity of all three.

Binding Specificity Demonstrated by Inhibition of Gel Shift Activity Using Antibodies—To validate the presence of DNMT1, RAP74, and p52 in the retarded bands, we used specific antibodies against these proteins to see if migration of these bands could be disrupted or supershifted in gel shift assays. Fig. 5 shows the gel autoradiography of this experiment. Addition of 1 µg of rabbit anti-DNMT1 polyclonal antibody (lane 2) enhanced the binding of the bands indicated by arrows a–c and promoted the appearance of a new diffuse band (arrow d) compared with the control (lane 1). Addition of increasing amounts of the same antibodies (lanes 3 and 4) resulted in almost complete disruption of the specific bands. This effect was not seen when similar amounts of anti-IgG control antibodies were used (lanes 5–7), suggesting that the anti-DNMT1 effect is specific and most probably because of the presence of DNMT1 in the bands formed with the HPFH −198 probe.

A similar titration of anti-RAP74 antibody showed a more dramatic effect, disrupting most of the specific bands with the lowest amount tested (Fig. 5, lane 8) and completely inhibiting the formation of the specific bands with higher amounts (lanes 9 and 10). Thus, similar to DNMT1, RAP74 seemed to be present in these bands. These results contrast with those obtained upon addition of anti-p52 antibody (lanes 11–13). Although the
Proteins That Bind to the γ-Globin HPFH − 198 Mutation

specific bands were enhanced in the presence of this antibody, they were neither disrupted nor supershifted. Increasing or reducing the amount of antibody in this assay led to similar results (data not shown). We concluded that, at least upon addition of this particular anti-p52 antibody, we could not accurately confirm the presence of p52 in the retarded bands specifically formed with the HPFH − 198 probe. In summary, these experiments strongly correlate with the Western blot findings of Fig. 4 and confirm that at least DNMT1 and RAP74 were present in the retarded bands specifically formed with the HPFH − 198 probe.

Correlation of HPFH − 198 Binding Activity and the Identified Peptides—To determine whether these proteins are individually recruited to the HPFH − 198 γ-globin gene promoter or whether they constitute a large protein complex, we fractionated MEL cell nuclear extracts in a Superdex gel filtration column under stringent buffer conditions (0.5 M KCl and 0.1% Nonidet P-40). Fig. 6A shows the protein fractionation profile of this column. Analysis by gel shift assay showed that the binding activities were separated into two groups. The first one peaked around fraction 10 and the second in fraction 17 (Fig. 6B). The activity in the first peak corresponded to the retarded bands specific for the HPFH − 198 probe in the gel shift assay (Fig. 1, arrows a–c); the second peak contained activity that bound equally well to both the WT and HPFH − 198 probes (data not shown). Extrapolation of the elution volume from the first activity peak in the calibration graph predicted a molecular mass of ~420 kDa for this activity (Fig. 6A, inset), suggesting that the HPFH − 198 binding activity is composed of a large protein complex.

The elution profiles of the DNMT1, RAP74, and p52 proteins in this column were analyzed by Western blotting and are shown in Fig. 6C. The elution of DNMT1 overlapped with the distribution of the specific HPFH − 198 binding activity (fractions 4–14). Compared with DNMT1, the elution distribution of RAP74 was better defined (fractions 8–11); its peak correlated well with both the HPFH − 198 binding activity and DNMT1 elution profile, suggesting that both proteins are components of a potential HPFH − 198 mutation-binding complex. On the other hand, the p52 protein did not seem to be part of this potential complex because its elution pattern (fractions 11 and 12) did not correlate with the other two. Because we found that the coactivator p52 was able to bind specifically to the HPFH mutation (Fig. 4), it is likely that p52 has a low affinity for the HPFH − 198 complex, resulting in later

FIGURE 6. DNMT1 and RAP74 co-migrate with HPFH − 198 binding activity. A, a fractionation profile of MEL cell nuclear extract on a Superdex S-200 HR 10/30 gel filtration column. Protein was detected by UV absorbance at 280 nm (dotted curve). The thick bar denotes the spread of the specific binding activity measured by gel shift assay shown in B. Inset, column calibration plot using different protein markers to determine the molecular mass of the HPFH − 198 binding activity. BSA, bovine serum albumin; Ovo, ovalbumin; Chymotryp, chymotrypsin; mAU, milli-absorbance units. B, gel shift assay of HPFH − 198 binding activity. The indicated fractions from gel filtration chromatography were dialyzed, and equal amounts (10 μl) were analyzed for their activity. Activity in the input material (In) is shown as a control in the first lane. C, Western blot analysis. Aliquots (20 μl) from Sephadex S-200 column fractions were run on 11% SDS-polyacrylamide gels; transferred to nitrocellulose; and analyzed with commercially available antibodies against DNMT1, RAP74, and p52. Specific reactive bands are indicated by arrows. MEL cell nuclear extract (NE) loaded onto the last lane was used as a control.
elution during gel filtration compared with the other components in the complex.

**DISCUSSION**

In this study, we have identified a set of proteins (DNMT1, CDC5-like protein, RAP74, SNEV, the coactivator p52, and a protein of unknown function) that are capable of binding to the HPFH –198 (T → C) mutation present in the ^γ-globin gene promoter. Using commercially available antibodies, we were able to confirm that at least two of these proteins, DNMT1 and p52, associate specifically with the HPFH –198 mutation (Fig. 4). In addition, we confirmed that DNMT1 and RAP74 are present in the retarded bands specifically formed with the HPFH –198 probe (Fig. 5) and that they co-elute as a potential complex using gel filtration chromatography (Fig. 6). Taken together, these results identify a set of proteins potentially involved in the regulation of expression of the γ-globin gene carrying the HPFH –198 mutation in adult erythropoiesis. In this study, we used nuclear extracts derived from MEL cells, which express the adult β-major and β-minor globin genes. We reasoned that MEL cells represent a transcriptional environment similar to adult erythroid cells, in which the HPFH –198 mutation is able to activate human ^γ-globin gene expression.

Early reports suggested that the most likely protein that binds to the HPFH –198 mutation is the transactivator Sp1 (3–5). However, Sp1 was not found among the proteins we identified. Differences in the transcriptional environments present in the different nuclear extracts used previously may explain this discrepancy. We think that our approach is more broadly based and unbiased compared with the early reports, as we first biochemically fractionated the extract and then purified proteins based on their binding affinity for the HPFH –198 mutation. Approaches similar to ours had been used to identify and purify the erythroid-specific transcription factor GATA-1 (21), NF-AT (22), and RelA (23), among others (24).

Of all the proteins that bound to the HPFH –198 mutation reported in this study, DNMT1 had the greatest number of unique peptides matching its sequence, covering ~40% of the protein. This identification was confirmed biochemically in three independent approaches using Western blotting after affinity purification (Fig. 4), by its co-elution with the HPFH –198 binding activity (Fig. 6), and by disruption of the HPFH –198 binding activity using a specific anti-DNMT1 antibody (Fig. 5). DNMT1 is regarded as a maintenance methyltransferase because it is responsible for copying DNA methylation patterns after DNA replication (13), thus maintaining the genomic epigenetic information in the cell. Because of this and its inability to specifically bind DNA, we were originally surprised to find DNMT1 among those proteins that bound specifically to the HPFH –198 mutation. However, recent reports indicate that DNMT1 has alternative functions independent of its CpG methylation activity that depend on the ability of its non-catalytic N-terminal portion to associate with an array of different proteins involved in transcriptional repression, chromatin regulation, and histone modifications (13, 25–29). Thus, DNMT1 has been shown to repress E2F-dependent transcription independently of its methyltransferase activity through a direct association with the tumor suppressor protein Rb and related family proteins (28, 30). DNMT1 was also shown to interact with HDAC1 and HDAC2 to partially repress transcription independently of histone deacetylation (25, 28). This is believed to be mediated through DNMT1 interaction with the corepressor DMAP1 (25) and its direct interaction with the repressor protein encoded by tsg101 (tumor susceptibility gene 101) (31). On the other hand, methylation by DNMT1 is an established epigenetic mechanism silencing gene expression, in particular, methylation of cognate CG-rich DNA sites such as those used by Sp1 and Krüppel-like family proteins to their cognate CG-rich sites (32). A direct link between methylation of CG-rich regions affecting Sp1 transactivation has been demonstrated in different reports (33–37). These results are extended by the recent finding suggesting that DNMT1 knockdown is responsible for the general activation of Sp1-dependent transcription, a process that is independent of both its methyltransferase- and histone deacetylase-recruiting activities (38). Thus, a general picture is emerging in which, upon association with one set of partners, DNMT1 could act independently of its methyltransferase activity to repress gene expression. Alternatively, association with another set of proteins involved in chromatin regulation and modification such as HP1 and SUV39H1 (27) or MBD2 and MBD3 (39) promotes gene silencing by increasing DNA methylation and chromatin condensation. The mechanism by which DNMT1 is involved in activation of the HPFH –198 γ-globin gene promoter in adult erythropoiesis is intriguing.

Upon induction of differentiation, MEL cells are capable of inducing expression of adult globin genes by 20–100-fold. A recent report showing the dynamic changes in transcription factors during erythroid maturation in MEL cells demonstrated that the nuclear matrix protein SNEV is associated with p18/MafK, the small subunit of the erythroid-specific transcription factor NF-E2, before but not after differentiation (40). The p18/MafK protein plays a dual role during erythroid maturation, shifting from a repressive to an activating function depending on its association with different protein partners (40). Thus, the finding that SNEV binds to the HPFH –198 mutant ^γ-globin promoter (this report) is in agreement with the notion that SNEV may be involved in the regulation of γ-globin gene expression during adult erythropoiesis.

We identified the coactivator p52 as another member of the protein complex that binds specifically to the HPFH –198 mutant ^γ-globin promoter. Interestingly, p52 is derived from an alternatively spliced product of a larger transcript encoding p75/LEDGF protein, which, among other things, associates with p18/MafK after MEL cell maturation when p18/MafK is functioning as an activator of erythroid gene expression (40). Because splicing occurs such that p75 contains all but the last 8 C-terminal amino acids of p52 (18), it is likely that a region common to p52/p75 is responsible for the interaction with p18/MafK. An in vitro reconstituted transcription system functionally identified p52 and p75 proteins as general transcriptional coactivators (18, 41). Notably, p52 is a more efficient coactivator compared with p75 in promoting the activation of Sp1-dependent transcription (18, 41). Unpublished in vitro studies mentioned in Ref. 18 stated that p52 is capable of interacting with RAP74 and two subunits of RNA polymerase II. Interest-
ingly, RAP74 is another protein that we found associated with the HPFH −198 mutant 3 gyr-globin promoter. Thus, depending on the transcriptional environment or the availability of transcription factors at any given moment, p52 seems to interact with multiple partners to function as a coactivator and to regulate gene expression.

Antibodies also allowed us to corroborate the identity and presence of RAP74 as another protein that binds to the HPFH −198 mutant 3 gyr-globin promoter. RAP74 is the largest subunit of TFIIF and associates with the smaller subunit RAP30 to form a tetramer. TFIIF plays important roles during transcription, recruiting RNA polymerase II to class II promoters, aiding in RNA polymerase II promoter escape, and stimulating elongation of transcription (15, 16, 42–44). Notably, RAP74 does not have DNA binding activity of its own, and to date, there are no reports indicating a TFIIF-independent function for this subunit, suggesting that the presence of RAP74 in our purification may not be specific. This could explain why we observed RAP74 eluting from both WT and HPFH −198 oligonucleotide affinity columns (Fig. 4). However, because we found that RAP74 co-eluted with DNMT1 and with the bulk of the HPFH −198 binding activity (Fig. 6) and because anti-RAP74 antibodies inhibited the formation of retarded bands in gel shift assays (Fig. 5), it is possible that part of RAP74 is brought specifically to the HPFH −198 mutation through interactions with other proteins.

The lack of commercially available antibodies for two other proteins identified, CDC5-like protein and an unnamed protein of unknown function, impeded us from characterizing them further and trying to link them to the regulation of gene expression in general and to the control of the globin gene in particular. However, despite its involvement in splicing (45), the human CDC5-like protein has been suggested to play a role as a transcription factor because it contains a DNA-binding domain with similarities to c-Myb (46, 47). This domain seems to be conserved, as the Arabidopsis thaliana homolog was shown to have sequence-specific DNA binding activity (48), potentially implicating CDC5-like protein in gene regulation.

We are in the process of generating transgenic mice carrying the HPFH −198 (T → C) mutation in the context of a 3 gyr yeast artificial chromosome construct. This will allow us to confirm the in vitro finding reported here in an effort to try to understand the changes that occur during the reactivation of 3 gyr-globin gene expression in an adult affected by non-deletion HPFH. Contributing to this understanding will be helpful in deciphering the complex mechanisms that regulate globin gene expression.

Acknowledgments—We thank the Mass Spectrometry Center, Department of Medicinal Chemistry, University of Washington, for skillful help in mass spectrometry and Dr. Patrick Navas for critical reading of the manuscript.

REFERENCES
1. Stamatoyannopoulos, G., and Grosveld, F. G. (2001) Hemoglobin Switching, 3rd Ed., Part II, W. B. Saunders Co., Philadelphia
2. Forget, B. G. (1998) Annu. N. Y. Acad. Sci. 850, 38–44
3. Ronchi, A., Nicolas, S., Santoro, C., and Ottolenghi, S. (1989) Nucleic Acids Res. 17, 10231–10241
4. Fischer, K. D., and Nowock, J. (1990) Nucleic Acids Res. 18, 5685–5693
5. Gmuoco, D. L., Rood, K. L., Blanchard-McQuate, K. L., Gray, T. A., Saulino, A., and Collins, F. S. (1991) Blood 78, 1853–1863
6. Berry, M., Grosveld, F., and Dillon, N. (1992) Nature 358, 499–502
7. Peterson, K. R., Li, Q. L., Clegg, C. H., Furukawa, T., Navas, P. A., Norton, E. J., Kimbrough, T. G., and Stamatoyannopoulos, G. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 5655–5659
8. Liu, L. R., Du, Z. W., Zhao, H. L., Liu, X. L., Huang, X. D., Shen, J., Li, L. M., Fang, F. D., and Zhang, J. (2005) J. Biol. Chem. 280, 7452–7459
9. Li, Q., Duan, Z. J., and Stamatoyannopoulos, G. (2001) EMBO J. 20, 157–164
10. Duan, Z., Stamatoyannopoulos, G., and Li, Q. (2001) Mol. Cell. Biol. 21, 3083–3095
11. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Anal. Chem. 68, 850–858
12. Wilm, M., Shevchenko, A., Houthaeve, T., Breit, S., Schweigerer, L., Fotsis, T., and Mann, M. (1996) Nature 379, 466–469
13. Robertson, K. D. (2002) Oncogene 21, 5361–5379
14. Bird, A. (2002) Genes Dev. 16, 6–21
15. Lei, L., Ren, D., Finkelstein, A., and Burton, Z. F. (1998) Mol. Cell. Biol. 18, 2130–2142
16. Flores, O., Maldonado, E., and Reinberg, D. (1989) J. Biol. Chem. 264, 8913–8921
17. Gotzmann, J., Gerner, C., Meissner, M., Holzmann, K., Grimm, R., Mikulits, W., and Sauermann, G. (2000) Exp. Cell Res. 261, 166–179
18. Ge, H., Si, Y., and Roeder, R. G. (1998) EMBO J. 17, 6723–6729
19. Singh, D. P., Kimura, A., Chylack, L. L. T., and Shinohara, T. (2000) Gene (Amst.) 242, 265–273
20. Muller, S., Scaffidi, P., Degryse, B., Bonalit, T., Ronfani, L., Agresti, A., Berlamte, M., and Bianchi, M. E. (2001) EMBO J. 20, 4337–4340
21. Tsai, S.-F., Martin, D. I., Zon, L. I., D’Andrea, A. D., Wong, G. G., and Orkin, S. H. (1989) Nature 339, 446–451
22. Corthesy, B., and Kao, P. N. (1994) J. Biol. Chem. 269, 20682–20690
23. Hansen, S. K., Baeuerle, P. A., and Blasi, F. (1994) Mol. Cell. Biol. 14, 2593–2603
24. Gadgil, H., Jurado, L. A., and Jarrett, H. W. (2001) Anal. Biochem. 290, 147–178
25. Rountree, M. R., Bachman, K. E., and Baylin, S. B. (2000) Nat. Genet. 25, 269–277
26. Fukis, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000) Nat. Genet. 24, 88–91
27. Fukis, F., Hurd, P. J., Deplus, R., and Kouzarides, T. (2003) Nucleic Acids Res. 31, 2305–2312
28. Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., and Wolfe, A. P. (2000) Nat. Genet. 25, 338–342
29. Macaluso, M., Cinti, C., Russo, G., Russo, A., and Giordano, A. (2003) Oncogene 22, 3511–3517
30. Pradhan, S., and Kim, G. D. (2002) EMBO J. 21, 779–788
31. Sun, Z., Pan, J., Hope, W. X., Cohen, S. N., and Balk, S. P. (1999) Cancer 86, 689–696
32. Kaczynzki, J., Cook, T., and Urrutia, R. (2003) Genome Biol. 4, 206
33. Kudo, S. (1998) Mol. Cell. Biol. 18, 5492–5499
34. Fujita, N., Watanabe, S., Ichimura, T., Ohkuma, Y., Chiba, T., Saya, H., and Nakao, M. (2003) Mol. Cell. Biol. 23, 2834–2843
35. Zhu, W. G., Srinivasan, K., Dui, Z., Duan, W., Drahun, L. J., Ding, H., Yee, L., Villalona-Calero, M. A., Plass, C., and Otterson, G. A. (2003) Mol. Cell. Biol. 23, 4056–4065
36. Robert, M. F., Morin, S., Beaulieu, N., Gauthier, F., Chute, I. C., Barsalou, A., and MacLeod, A. R. (2003) Nat. Genet. 33, 61–65
37. Fournel, M., Sapieha, P., Beaulieu, N., Besterman, J. M., and MacLeod, A. R. (1999) J. Biol. Chem. 274, 24250–24256
38. Milutinovic, S., Brown, S. E., Zhuang, Q., and Szfy, M. (2004) J. Biol. Chem. 279, 27915–27927
39. Tatematsu, K. I., Yamazaki, T., and Ishikawa, F. (2000) Genes Cells 5, 677–688
40. Brand, M., Ranish, J. A., Kummer, N. T., Hamilton, J., Igarashi, K., Frankel, C., Chi, T. H., Crabtree, G. R., Aebersold, R., and Groudine, M.
Proteins That Bind to the γ-Globin HPFH – 198 Mutation

(2004) Nat. Struct. Mol. Biol. 11, 73–80
41. Ge, H., Si, Y., and Wolffe, A. P. (1998) Mol. Cell 2, 751–759
42. Chang, C., Kostrub, C. F., and Burton, Z. F. (1993) J. Biol. Chem. 268, 20482–20489
43. Olave, I., Drapkin, R., and Reinberg, D. (1997) in mRNA Metabolism & Post-Transcriptional Gene Regulation (Harford, J. B., and Morris, D. R., eds) pp. 23–42, Wiley-Liss Inc., New York
44. Reinberg, D., Orphanides, G., Ebright, R., Akoulitchev, S., Carcamo, J., Cho, H., Cortes, P., Drapkin, R., Flores, O., Ha, I., Inostroza, J. A., Kim, S., Kim, T. K., Kumar, P., Lagrange, T., LeRoy, G., Lu, H., Ma, D. M., Maldonado, E., Merino, A., Mermelstein, F., Olave, I., Sheldon, M., Shiekhattar, R., and Zawel, L. (1998) Cold Spring Harb. Symp. Quant. Biol. 63, 83–103
45. Ajuh, P., Kuster, B., Panov, K., Zomerdijk, J., Mann, M., and Lamond, A. I. (2000) EMBO J. 19, 6569–6581
46. Ohi, R., McCollum, D., Hirani, B., Den Haese, G. J., Zhang, X., Burke, J. D., Turner, K., and Gould, K. L. (1994) EMBO J. 13, 471–483
47. Bernstein, H. S., and Coughlin, S. R. (1997) J. Biol. Chem. 272, 5833–5837
48. Hirayama, T., and Shinozaki, K. (1996) Proc. Natl. Acad. Sci. U. S. A. 93, 13371–13376
49. Kerrigan, L. A., and Kadonaga, J. T. (1996) in Current Protocols in Molecular Biology (Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds) pp. 12.10.1–12.10.18, John Wiley & Sons, Inc., New York