MOLECULAR AND CELLULAR BIOLOGY, Feb. 1990, p. 634–642
0270-7306/90/020634-09$02.00/0

Vol. 10, No. 2

The *GLI* Gene Encodes a Nuclear Protein Which Binds Specific Sequences in the Human Genome

KENNETH W. KINZLER AND BERT VOGELSTEIN*

Johns Hopkins Oncology Center, 424 North Bond Street, Baltimore, Maryland 21231

Received 28 August 1989/Accepted 1 November 1989

The *GLI* gene is amplified in a subset of human tumors and encodes a protein product with five zinc finger DNA-binding motifs. In this study, we show that the *GLI* gene product has a predominantly nuclear localization and binds DNA in a sequence-specific fashion. Three GLI binding sites were identified by using a novel procedure in which total human DNA was bound to a GLI recombinant fusion protein, and the polymerase chain reaction was used to amplify and recover the bound sequences. The GLI protein protected a 23- to 24-base region within all three binding sites, and the protected region in each case included the 9-base-pair sequence 5′-GACCACCCA-3′. One of the binding sites was contained within a 63-base-pair repeat of the variable number of tandem repeat type, whereas the other two sites were represented once in the genome. The approach used here to identify GLI binding sites should be applicable to the characterization of other zinc finger proteins.

One of the major challenges in cell and tumor biology is to define the interactions between a given protein and other cellular components. In this report, we describe the first steps at elucidating these interactions for a human zinc finger protein, GLI. The *GLI* gene was originally identified and cloned by virtue of its amplification and high levels of expression in a human glioblastoma (8). Subsequently, the *GLI* gene has been found to be amplified in other human glioblastomas and in some liposarcomas, osteosarcomas, and teratocarcinomas (20, 30; unpublished observations). The *GLI* gene locus has been localized to 12q13.3 to 14.1, a region frequently translocated in human lipomas, myxoid liposarcomas, uterine leiomyomas, and pleomorphic adenomas of the salivary gland (1, 8).

Analysis of the *GLI* gene sequence indicates that it encodes a 118-kilodalton (kDa) protein that contains five repeats of a zinc finger DNA-binding motif (9). The zinc finger motif was first recognized in *Xenopus* transcription factor IIIA (TFIIIA), which subsequently became the prototype for a large family of zinc finger DNA-binding proteins (3, 15). The GLI zinc fingers also have the H-C link first described in the *Drosophila* Kruppel gene (4, 22). In this subfamily of zinc finger genes, the C-terminal histidine from one finger is linked to the next finger through the sequence HTGEKP. The *Drosophila* Kruppel gene is a segmentation gene of the gap class which regulates embryonic development (19). This similarity to the Kruppel gene is particularly interesting because of the expression pattern of *GLI*. *GLI* is not expressed in most adult tissues but is expressed in embryonal carcinoma cell lines and testes. The *GLI* gene is itself a prototype for a family of human zinc finger genes; a probe for the *GLI* zinc finger region was used to identify six other zinc finger genes in the human genome (21). This family of human zinc finger genes could be divided into two groups by greater similarity to either the *GLI* (*GLI*, *GLI2*, or *GLI3*) or Kruppel (*HKR1*, *HKR2*, *HKR3*, or *HKR4*) gene. By analogy to other zinc finger proteins, we suspected that the *GLI* gene also encoded a protein that could bind to DNA in a sequence-specific manner. This hypothesis was tested in the experiments described below. Specifically, a combination of immunohistochemical experiments and a novel procedure for isolating DNA sequences bound by proteins has shown that the *GLI* gene product is a nuclear protein which can bind to specific sequences within the human genome.

MATERIALS AND METHODS

Construction of a bacterial fusion protein. A full-length *GLI* cDNA clone (pGLIK12) containing GLI nucleotides −79 to 3522 was constructed in pBluescript from partial cDNA clones previously described (9). The CXY fusion protein containing just the carboxy terminus of GLI was constructed by cloning a *Stu*I-*Xba*I restriction fragment of pGLIK12 into the *Sma*I and *Xba*I sites of the expression vector pEX-1 (26). *Stu*I cleaves GLI at nucleotide 2308, and *Xba*I cleaves the vector sequences flanking the 3′ end of GLI. This construct resulted in the fusion of the 117-kDa amino terminus of the Cro–β-galactosidase hybrid protein to the carboxy terminus of GLI (amino acids 771 to 1106). Similarly, the ZF fusion protein containing the zinc fingers was constructed by cloning a *Hin*cII-*Xba*I restriction fragment of pGLIK12 into the *Sma*I and *Xba*I sites of pEX-2. This construction resulted in the fusion of the 117-kDa amino terminus of the Cro–β-galactosidase hybrid protein to amino acids 78 to 1106 of GLI. The GLI zinc fingers are between amino acids 235 and 393. Fusion proteins were produced in *Escherichia coli* pop2136, in which expression of the pEX Cro–β-galactosidase hybrid gene is regulated by the temperature-sensitive λ repressor *c*Its857 (26). For production of proteins, cultures of pEX vector-transformed cells were grown at 30°C to an optical density of 0.5 at 550 nM. Fusion protein production was induced by shifting the bacteria to 37°C for 1.5 h. The bacteria were then harvested by centrifugation and suspended in 1/100 volume of TEN (50 mM Tris hydrochloride [pH 8.0], 1 mM EDTA, 50 mM NaCl), and kept on ice. Cells were disrupted by sonication, and insoluble protein was pelleted by centrifugation (10,000 × *g* for 10 min at 4°C). The insoluble fusion pellets were washed with TEN, and the pellets were stored at −70°C. The protein pellets retained DNA-binding activity when thawed.

* Corresponding author.

Antibody production. The CXY fusion protein was purified by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (SDS-PAGE). Insoluble fusion pellets were solubilized by boiling in SDS-PAGE loading buffer (63 mM Tris hydrochloride [pH 6.8], 10% glycerol, 5% 2-mercaptoethanol, 2% SDS, 0.025% bromophenol blue). The proteins were separated by electrophoresis on a 8% polyacrylamide gel, and the fusion protein was excised and homogenized in adjuvant. A female New Zealand rabbit was immunized with a gel slice containing approximately 200 μg of fusion protein in Freund complete adjuvant. The rabbit was given booster injections 14, 51, and 85 days postimmunization with gel slices containing approximately 200 μg of fusion protein in Freund incomplete adjuvant. Antisera for this study were taken 13 days after the third booster injection. Antibodies for Western blot (immunoblot) analysis and immunohistochemistry were affinity purified by binding to fusion proteins (7). Antibodies reactive against the Cro–β-galactosidase portion of the fusion protein were substracted by incubating 1 ml of antisera with approximately 300 μg of insoluble Cro–β-galactosidase (isolated from 30-ml culture). After 1 h on ice, the insoluble protein and Cro-β-galactosidase-reactive antibodies were removed by centrifugation (10,000 × *g* for 5 min at 4°C). The substracted antisera was then incubated with approximately 300 μg of insoluble CXY fusion protein for 1 h on ice. The GLI-reactive antibodies were then recovered by centrifugation (10,000 × *g* for 5 min at 4°C). The GLI-reactive antibodies were released from the insoluble protein by incubation in 500 μl of 0.2 M glycine (pH 2.3) for 5 min on ice. After centrifugation, the supernatant was neutralized with 70 μl of 1 M Tris hydrochloride (pH 9.5) and a second round of binding to GLI fusion protein was performed. After elution from the insoluble pellet, the supernatant was neutralized as above and bovine serum albumin was added to a final concentration of 5 mg/ml to stabilize the protein.

Western blot (immunoblot) analysis. Total protein was prepared by homogenizing and boiling D-259 MG cells (2), Tera-1 cells (6), or adult human testis tissue, in SDS-PAGE loading buffer. After electrophoresis through an 8% SDS polyacrylamide gel, proteins were electrophoretically transferred to a polyvinyldifluoride membrane (Immobilon; Millipore). The membrane was blocked for 1 h in Blotto-TBS (5% nonfat dried milk, 0.1% sodium azide in 100 mM Tris hydrochloride [pH 7.5], 0.9% NaCl [TBS]) and incubated for 2 h with affinity-purified GLI antibody (20 μg/ml) in Blotto-TTBS (Blotto-TBS plus 0.1% Tween 20). Nonbound antibody was removed by washing in TTBS (TBS plus 0.1% Tween 20) for 40 min with four buffer changes. The membrane was then incubated with ^{125}I-labeled goat anti-rabbit antibody (3.6 μCi/μg; 0.25 μCi/ml; Dupont) in Blotto-TTBS for 1 h at room temperature. Nonbound antibody was removed by washing as described above. Autoradiography was performed with Kodak XAR-5 film and Dupont Lightning Plus intensifier screens at −70°C.

In vitro translation. *GLI* transcript was prepared by in vitro transcription of *Xba*I-cleaved pGLIK12 with T3 DNA-dependent RNA polymerase. Transcription was carried out as previously described, except that 1 mM of the cap analog G(5′)ppp(5′)G was included and GTP was reduced to 200 μM (17). Transcripts were translated by using nuclease-treated rabbit reticulocyte lysates (Promega) and [^{35}S]methionine following the conditions recommended by the manufacturer. Translation products were analyzed on a 8% SDS polyacrylamide gel. After electrophoresis, gels were fixed, impregnated with En^{3}Hance (Dupont), dried, and exposed to Kodak XAR film at −70°C.

Immunohistochemistry. For the D-259 MG xenograft, 6-μm-thick frozen sections were cut from tumor embedded in O.C.T. (Miles Laboratories, Inc.) and mounted on slides coated with Cell-Tak (Bio Polymers, Inc.). Tissue was immediately fixed for 10 min at room temperature in 1% (vol/vol) formaldehyde in Hanks balanced saline solution (HBSS) and washed with HBSS. D-259 MG cells in culture were grown on slides coated with Cell-Tak and fixed in 1% (vol/vol) formaldehyde in HBSS. Cells were permeabilized with 0.5% Triton X-100 in HBSS for 10 min and washed with HBSS. Slides were then incubated for 20 min in TBS and blocked for 1 h in Blotto-TBS. Sections were incubated with primary antibody (10 to 20 μg/ml) in Blotto-TTBS for 1 h and washed with TTBS for 15 min with three changes. Slides were then incubated with biotinylated goat anti-rabbit (1.5 μg/ml; Vector Laboratories) in TBS for 0.5 h and washed with TBS for 15 min with three changes. Immunoperoxidase staining was performed with the Vectastain Elite ABC System (Vector Laboratories). Complex formation and color development was performed following the protocol of the manufacturer, except that 0.04% nickel chloride was included in the substrate solution.

Isolation of DNA sequences which bind to GLI. Sonicated human DNA (200 to 300 base pairs [bp]) was converted to a form suitable for whole-genome polymerase chain reaction (PCR) by ligation to catch-linkers as previously described (10). First-round binding was performed in 30 μl of DNA-binding buffer (50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid), (pH 7.5); 50 mM KCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 1 mM dithiothreitol, 20% glycerol) with 3 μg of poly(dI-dC) as competitor and 300 ng of catch-linked DNA. A ZF fusion protein pellet (60 μg of total protein) was suspended in this mix and incubated for 20 min on ice. Bound DNA and fusion protein were recovered by centrifugation (10,000 × *g* for 3 min at 4°C). The fusion protein pellet was then washed twice by suspending it in 1.5 ml of wash buffer (50 mM HEPES [pH 7.5], 150 mM KCl, 5 mM $MgCl_2$, 10 μM $ZnSO_4$, 1% Triton X-100, 0.05%

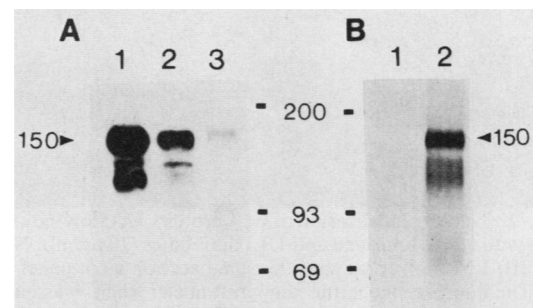

FIG. 1. The *GLI* gene codes for a 150-kDa protein. (A) Detection of GLI protein by Western blot analysis. Lanes were loaded with 100 μg of total protein from D-259 MG cells (lane 1), Tera-1 cells (lane 2), or human testes (lane 3). An affinity-purified rabbit antibody against GLI was used in conjunction with radiolabeled goat anti-rabbit antibodies for this detection. The major polypeptide identified was 150 kDa (lanes 1 and 2). Marker sizes (in kilodaltons) are indicated to the right of the gel. (B) In vitro-translated GLI migrates at the same position as in vivo-translated GLI. Nuclease-treated rabbit reticulocyte lysates were programmed with either *GLI* transcript (lane 2) or no exogenous transcript (lane 1). Translation products were analyzed by SDS-PAGE and fluorography. Marker sizes (in kilodaltons) are indicated to the left of the gel.

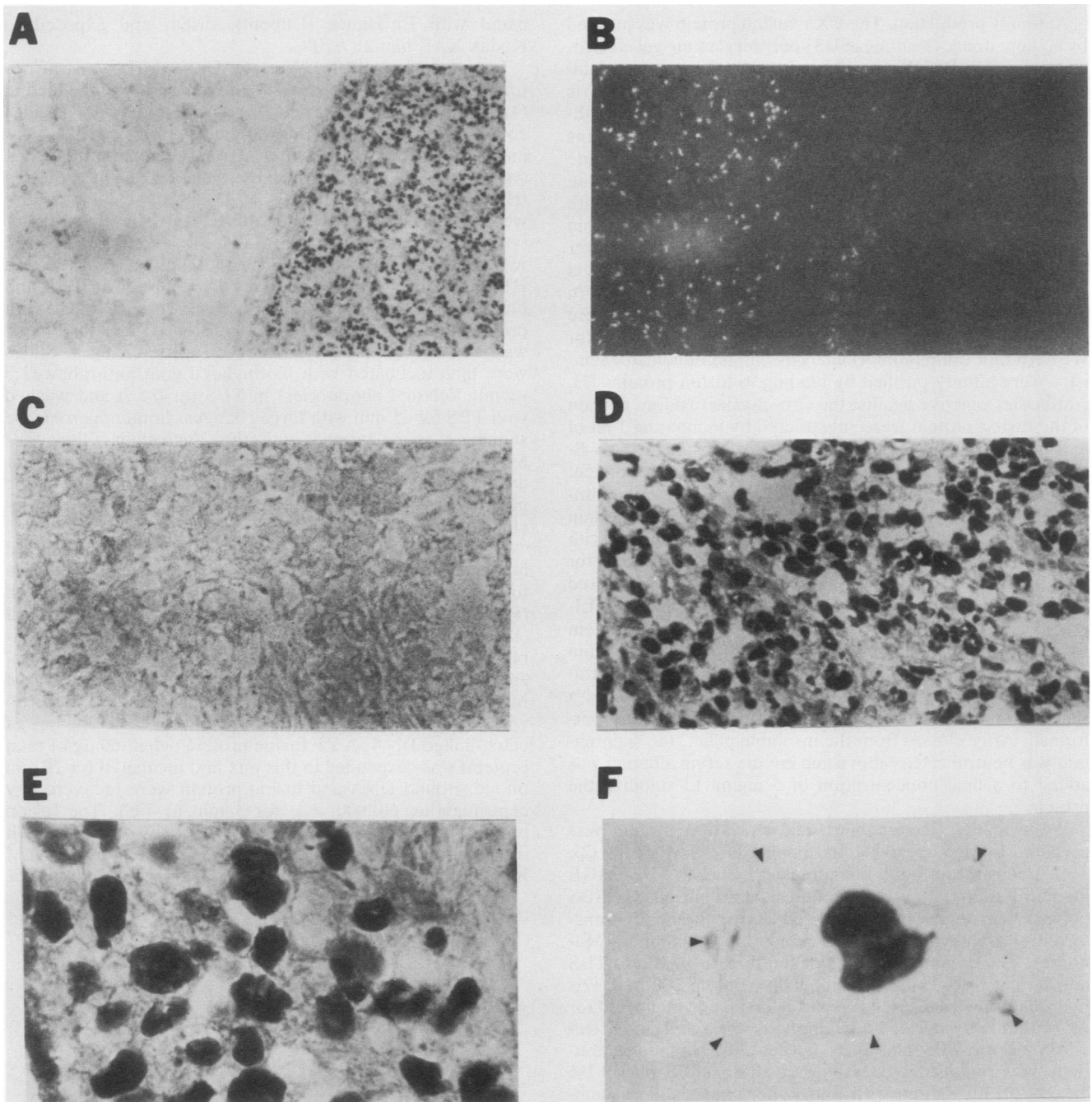

FIG. 2. Nuclear localization of GLI protein. (A) Low-power view of D-259 MG xenograft (right) and surrounding normal mouse cells (left) stained with affinity-purified anti-GLI antibodies (10 μg/ml). Note the staining of the nuclei in the D-259 MG xenograft but not in normal mouse tissue. (B) Low-power view of the same section as in panel A stained with the fluorescent dye 4,6-diamidino-2-phenylindole, which stains DNA. The fluorescence in the xenograft nuclei (right) was quenched by the immunoperoxidase product. (C) Medium-power view of a D-259 MG xenograft section stained with purified normal rabbit IgG (20 μg/ml). Little staining is evident on comparison to the adjacent section shown in panel D. (D) Medium-power view of D-259 MG xenograft stained with affinity-purified anti-GLI antibodies (10 μg/ml). (E) High-power view of same slide as in panel D. (F) High-power view of a D-259 MG cell in culture stained with affinity-purified anti-GLI antibodies (10 μg/ml). Arrowheads indicate the edge of the cell.

SDS), followed by centrifugation (10,000 × *g* for 5 min at 4°C). Bound DNA was freed from ZF fusion protein by digestion with proteinase K for 2 h at 50°C (500 μg of Proteinase K per ml, 500 mM Tris hydrochloride [pH 9.0], 20 mM EDTA, 10 mM NaCl, 1% SDS). DNA was further purified by extraction with buffered phenol-chloroform (PC9, 3 parts phenol: 2 parts chloroform: 2 parts 500 mM Tris hydrochloride [pH 9.0], 20 mM EDTA, 10 mM NaCl), followed by ethanol precipitation. Recovered DNA was then amplified by using 25 cycles of PCR with catch-linkers as primers. The bound and amplified DNA was then used as the substrate for additional rounds of binding to the GLI fusion protein. After three or four rounds of selection, libraries were constructed and screened as previously described (10).

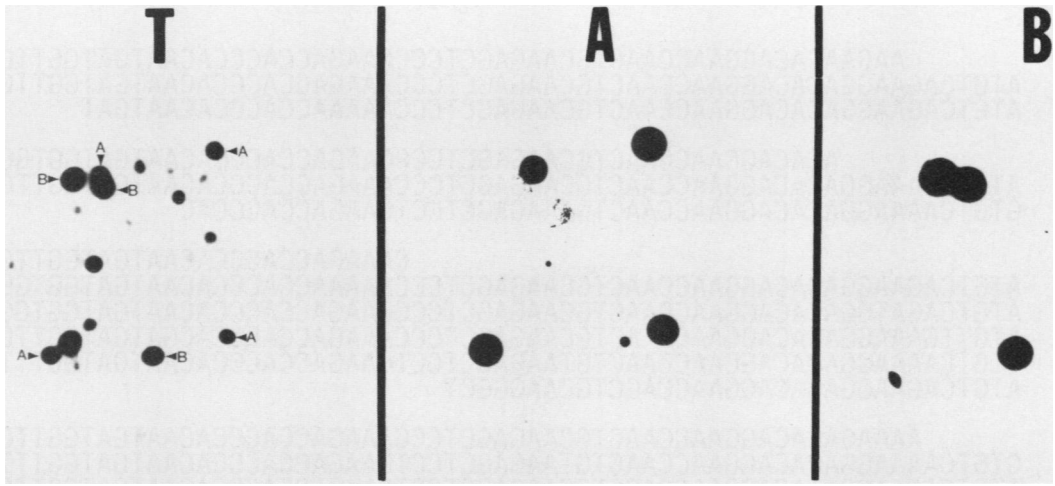

FIG. 3. Identification of potential GLI binding sites. (T) Results of screening a library constructed from DNA that had been selected and amplified four times. Approximately 300 recombinant plaques were represented on the membrane. The hybridization probe was the fourfold-amplified and selected DNA. (A) Results of hybridization of the same membrane with clone A1. (B) Results of hybridization obtained with clone B1. In panel T, the plaques hybridizing to either clones A1 or B1 are indicated by arrowheads.

Briefly, the amplified and selected sequences were cleaved with *Eco*RI, taking advantage of *Eco*RI sites engineered into the catch-linkers and cloned into λGT10. Sequences that were enriched in the library were identified through plaque screening with a radiolabeled probe prepared from the amplified and selected sequences. Repeated sequences were removed from the hybridization probe by preannealing with human DNA as previously described (23). *Eco*RI inserts from plaque purified phages were subcloned into pBluescript.

Individual *Eco*RI-cleaved plasmids were labeled for the binding assay by filling in the cohesive ends with [α-^{32}P]dATP and the Klenow fragment of DNA polymerase I. DNA binding was performed as described above, except that 10 μg of fusion protein and approximately 5 ng of labeled plasmid DNA were used. The final recovered DNA was analyzed by gel electrophoresis, followed by autoradiography.

Nucleotide sequencing. Nucleotide sequencing was performed on double-stranded plasmid templates as previously described (11).

Southern blot analysis. Total DNA prepared from normal human lymphocytes was cleaved with *Hae*III or *Eco*RI, separated by electrophoresis through an agarose gel, transferred to nylon membranes, and hybridized as previously described (21). *Eco*RI inserts isolated from agarose gels (28) were radiolabeled by the random primer method (5) and used as hybridization probes.

Preparation of labeled inserts for gel retardation and DNase footprinting. Inserts labeled on only one strand were prepared by utilizing unique *Xba*I and *Hin*dIII vector sites that flanked the insert on either side. Plasmid DNA was cleaved with *Xba*I and labeled by filling in with the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP. Labeling was terminated by extraction with PC9 and followed by ethanol precipitation. The insert was then freed from vector by digestion with *Hin*dIII. Labeling the opposite strand was similar, except that the *Hin*dIII end was labeled and the insert was freed with *Xba*I. For gel retardation and the negative control for DNase footprinting, the inserts were purified from a 5% polyacrylamide gel as previously described (13). Inserts prepared from clone C1 in this manner contained a dimer of the original binding site, because the plasmid subclone had a double insert oriented as a direct repeat.

DNase footprinting. DNA binding with approximately 100 ng of end-labeled plasmid per insert and 30 μg of GLI fusion protein was performed as described above. After being washed, the pellets containing bound DNA and GLI zinc finger fusion protein were incubated for 1.5 min at room temperature in 30 μl of DNA-binding buffer containing 5, 25, or 75 ng of DNase I per ml. DNase digestion was stopped by the addition of 400 μl of a solution containing 500 μg of proteinase K per ml, 500 mM Tris hydrochloride (pH 9.0), 20 mM EDTA, 10 mM NaCl, and 1% SDS. DNA was purified by extraction with PC9 and precipitated with ethanol as described above. Samples were dissolved in sample buffer

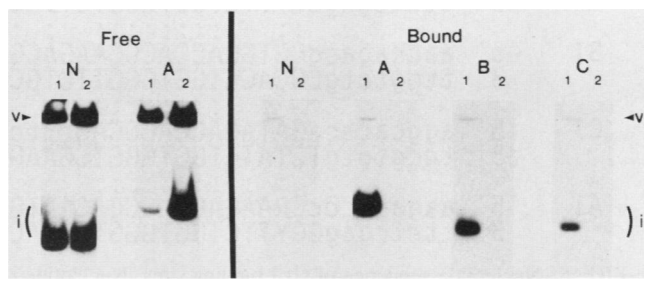

FIG. 4. The cloned sequences bind GLI zinc finger fusion protein specifically. Clones derived from the screening protocol shown in Fig. 3 were tested for binding to ZF fusion protein. Labeled DNA from these clones was incubated with insoluble fusion protein and centrifuged. DNA recovered from the supernatants (free) or the fusion protein pellet (bound) were separated on an agarose gel and visualized by autoradiography. Assays were performed with either the ZF fusion protein (lanes 1) or the CXY fusion protein which was devoid of the zinc finger region (lanes 2). The assay was performed on *Eco*RI-cleaved plasmids containing either a human sequence bound by *Xenopus laevis* TFIIIA (lanes marked N; 10) or with clone A1 (A lanes), clone B1 (B lanes), or clone C1 (C lanes). Vector bands are denoted by v (2.9 kilobases) and insert bands are denoted by i (100 to 300 bp).

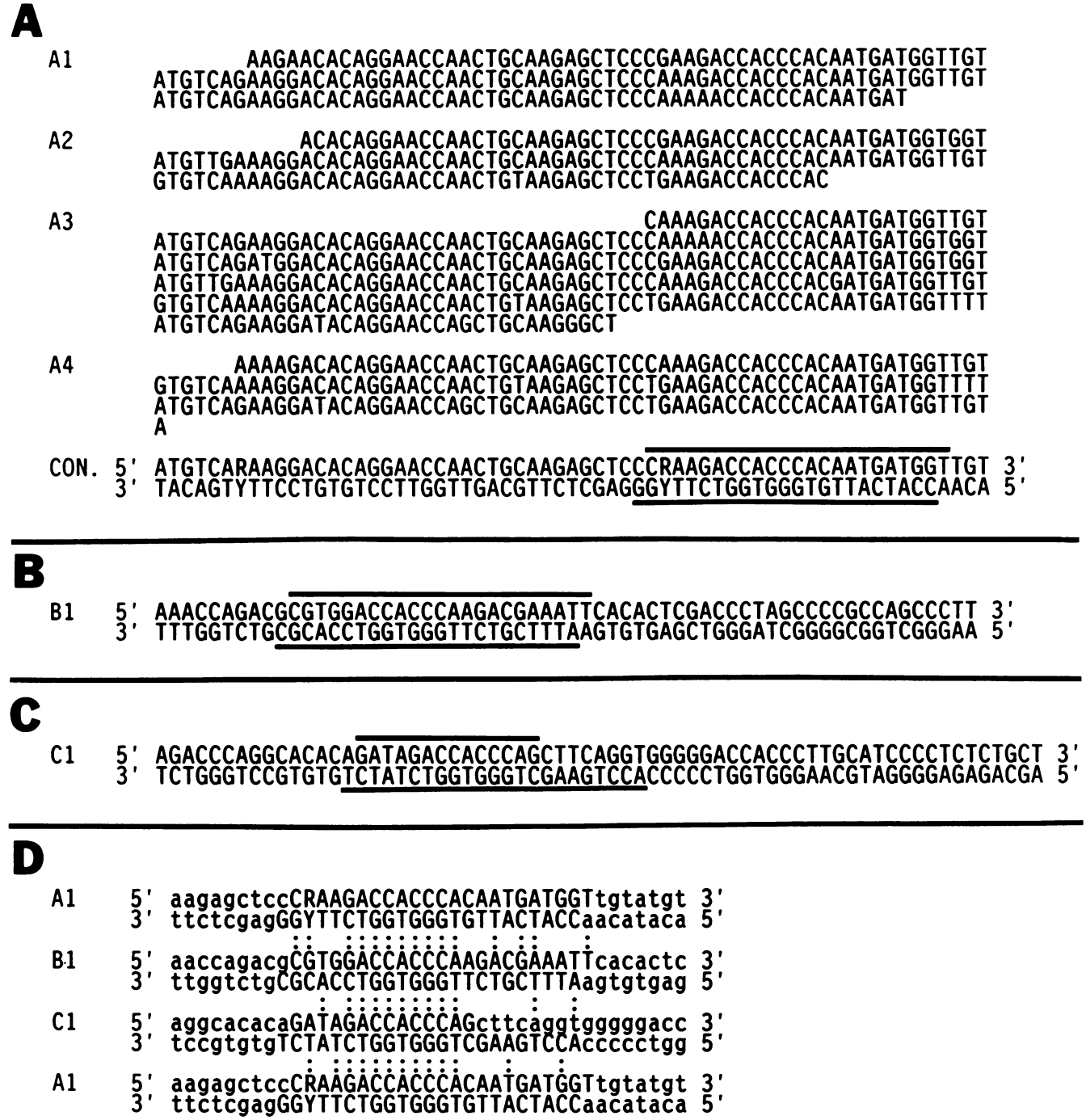

FIG. 5. Nucleotide sequence of GLI binding sites. Nucleotide sequences of inserts from clones A1, A2, A3, and A4 (176, 166, 313, and 184 bp, respectively) are shown in panel A. A consensus for the 63-bp repeat (CON.) was derived from the four clones. Each nucleotide specified was present in greater than 75% of the repeats. At two positions, a specific nucleotide could not be specified because the position was equally split between A and G (R = purine; Y = pyrmidine). Also shown are the nucleotide sequences of clones B1 (B) and C1 (C), which were 62 and 67 bp, respectively. In panels A, B, and C, the underlined and overlined sequences represent the regions on each strand which were protected in DNase footprinting experiments. The three binding sites are aligned with one another in panel D. The dots between sequences indicate identical nucleotides. The uppercase letters indicate the regions on each strand which were protected in DNase footprinting experiments.

(40% formamide, 8 mM EDTA, 0.02% bromophenol blue, 0.02% xylene cyanol) and analyzed in a 6% acrylamide–8 M urea sequencing gel. As a negative control, 2 ng of end-labeled insert was treated with DNase I as described above, except that no fusion protein was present. Markers consisted of the Maxam and Gilbert reaction for A and G performed on the same labeled inserts (14).

Gel retardation assay. End-labeled inserts (0.1 to 0.5 ng) were incubated for 20 min at room temperature in 20 μl of DNA-binding buffer containing 2 μg of poly(dI-dC) and 4 μl

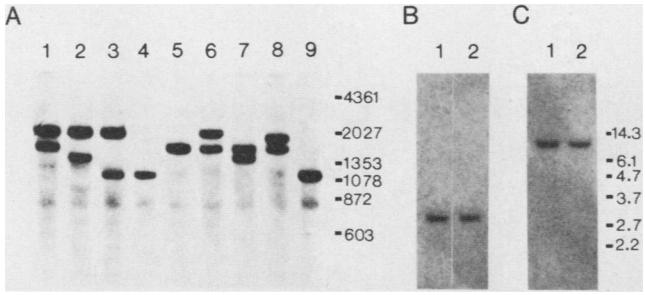

FIG. 6. Clone A contains VNTR sequences, whereas clones B and C are represented at low-copy numbers within the human genome. Southern blot analysis were performed by using clone A1 (A), clone B1 (B), or clone C1 (C) as a hybridization probe. In panel A, each lane contained 5 μg of DNA from normal human lymphocytes cleaved with *Hae*III (lanes 1 through 4, male; lanes 5 through 9, female). The markers were *Hin*dIII-cleaved λ phage DNA and *Hae*III-cleaved φX phage DNA. For panels B and C, each lane contained 5 μg of DNA from normal human lymphocytes cleaved with *Eco*RI (lane 1, male; lane 2, female). The marker was *Bam*HI-*Eco*RI-cleaved adenovirus II DNA (IBI). Marker sizes (in base pairs for panel A and in kilobases for panels B and C) are indicated to the right of the gels.

of rabbit reticulocyte lysate translation reaction. Translation reactions with and without *GLI* transcript were performed as described above, except that [^{35}S]methionine was not used. For competition experiments, approximately 0.5 μg of *Eco*RI-cleaved plasmid was added before adding the translation reaction mixture. Complexes were analyzed by electrophoresis on a 6% glycerol–4.5% acrylamide gel as previously described (25). After electrophoresis (80V for 3 h), gels were fixed in 5% methanol–5% acetic acid, dried, and exposed to film.

RESULTS

Detection of GLI protein. As a first step in the analysis of the GLI protein, a polyclonal antisera to GLI was generated. A fusion protein (CXY fusion protein) containing 117-kDa Cro–β-galactosidase hybrid protein fused to the carboxy terminus of GLI (amino acids 771 to 1106) was produced in *E. coli* and used to immunize a rabbit. The resultant antibodies were affinity purified and used for Western blot analysis of proteins isolated from D259 MG cells, Tera-1 cells, and testes (Fig. 1A). The D259 MG cell line is derived from a glioblastoma multiforme, contains a 75-fold amplification of the *GLI* gene, and expresses high levels of *GLI* transcript (8). Tera-1 is an embryonal carcinoma cell line which expresses intermediate amounts of *GLI* transcript (approximately 5 to 10 copies per cell), and testes express only 1 to 2 *GLI* transcripts per cell, as judged by RNase protection analysis (data not shown). In D259 MG cells, the affinity-purified antisera to GLI detected one major protein with a relative migration of 150 kDa and several smaller polypeptide bands, which were probably degradation products (Fig. 1A, lane 1). GLI protein was also detectable in Tera-1 (lane 2) and testes (lane 3). The levels of GLI protein detected in Western blot analysis correlated well with the levels of RNA transcripts in these cell types.

Because the observed migration of GLI protein was significantly different from that expected based on its calculated molecular mass of 118 kDa, we determined the relative migration of an in vitro-translated GLI protein. *GLI* transcript was prepared from a full-length *GLI* cDNA and then translated in vitro with rabbit reticulocyte lysates. The major GLI protein synthesized in vitro also migrated with an apparent molecular mass of 150 kDa (Fig. 1B, lane 2). Thus, the discrepancy between the observed migration of GLI and the calculated molecular mass was probably not due to posttranslational modification but was intrinsic to its amino acid sequence.

Intracellular localization of GLI. The GLI antibodies were used to determine the subcellular localization of GLI proteins in D259 MG cells by immunohistochemistry. Antibodies to GLI specifically stained D259 MG cells grown as a nude mouse xenograft but did not stain surrounding normal mouse cells (Fig. 2A). No staining above background was detected when normal rabbit immunoglobulin G (Fig. 2C) was used in place of the anti-GLI antibody (Fig. 2D). In individual D-259 MG cells grown in vivo (Fig. 2E) or in vitro (Fig. 2F), the vast majority of GLI protein was found in the nucleus. This localization was consistent with GLI being a DNA-binding protein.

Identification of GLI binding sites. To identify potential GLI binding sites, we modified a recently described technique in which whole-genome PCR is used to amplify sequences selected by binding to a specific protein (10). Briefly, total genomic DNA is sonicated to an average size of a few hundred base pairs and the DNA fragments ligated to catch-linkers consisting of a 20-bp DNA fragment synthesized in vitro. Each catch-linker was engineered with half of an *Xho*I site at its termini and an internal *Eco*RI site. The ligation product was then cleaved with *Xho*I to separate linkers ligated to themselves. The linked DNA was then in a form that could be amplified by the PCR, using the catch-linkers as primers. The DNA was then selected by binding to an insoluble GLI fusion protein and the bound fragments were purified. The small amount of DNA specifically bound was amplified by using PCR, and these two steps (GLI binding followed by amplification) were repeated to achieve the desired purity. A fusion protein was chosen for a source of GLI protein because a large quantity could be easily isolated from bacteria and the insoluble nature of the protein made it possible to easily separate bound DNA from free DNA by centrifugation. Furthermore, we suspected that a fusion protein would maintain legitimate DNA-binding activity based on previous experiments with other β-galactosidase fusion proteins which retain such activity (for a recent review, see reference 24). The GLI zinc finger fusion protein (ZF fusion protein) was produced in bacteria by using a pEX vector construct containing a Cro–β-galactosidase hybrid protein fused to amino acid 78 of GLI and extending to the C terminus.

After a total of three cycles of binding to ZF fusion protein and PCR, the amplified DNA was cleaved with *Eco*RI and used to construct a library in λGT10. Approximately 10,000 recombinant plaques were screened with radiolabeled threefold-selected sequences to identify sequences enriched in the library. Approximately 100 plaques were weakly positive, and four of these were chosen for further characterization. Subsequently all four clones (A1, A2, A3, and A4) were shown to represent the same GLI binding site (class A clones; see below). To identify additional binding sites, a fourth round of binding and PCR was performed. A new library was constructed from this DNA and screened with fourfold-selected sequences. Approximately 5% of the clones hybridized to the amplified DNA (Fig. 3T) and five additional hybridizing clones which did not hybridize to the

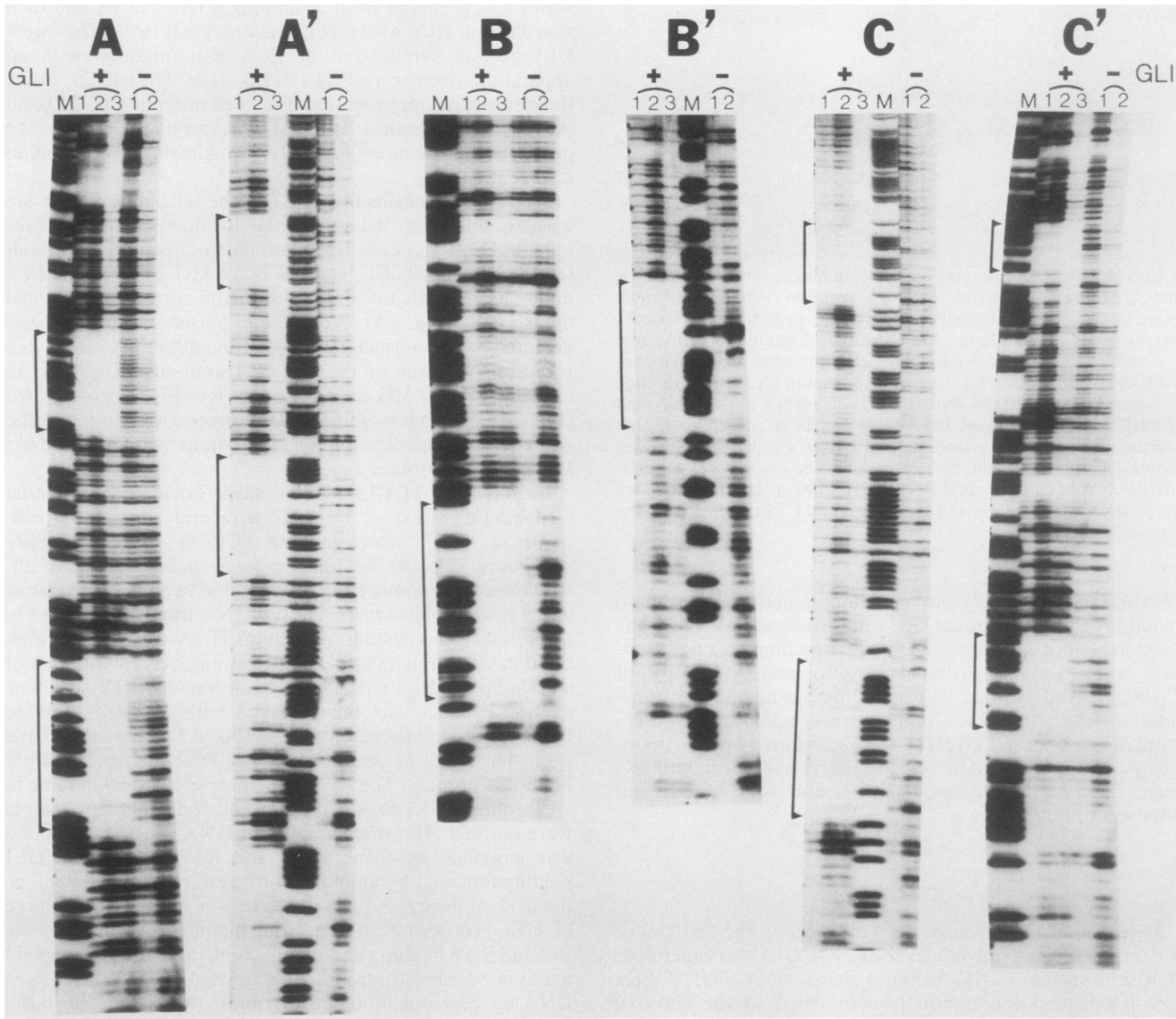

FIG. 7. DNase footprinting of GLI binding sites. DNase footprinting was performed with clone A1 (A and A'), clone B1 (B and B'), and clone C1 (C and C'). The regions protected in panels A, B, and C are from the lower strand of sequence presented in Fig. 5, whereas panels A', B', and C' illustrate protections of the upper strand. Inserts bound in the presence (+) or absence (−) of GLI fusion protein were treated with DNase I (lane 1, 5 ng/μl; lane 2, 25 ng/μl; lane 3, 75 ng/μl). Markers (lanes M) were the same inserts labeled at one end and subjected to the Maxam and Gilbert reactions for A and G. The brackets indicate the protected regions in each panel.

previously isolated binding site (Fig. 3A) were chosen for detailed analysis.

To determine whether the hybridizing clones actually bound to GLI protein, the inserts from the bacteriophage clones were subcloned, radioactively labeled, and incubated with the insoluble ZF fusion protein. Bound and free DNA were separated by centrifugation, purified, and analyzed by electrophoresis. Bindings were performed with the ZF fusion protein containing the zinc finger domain (Fig. 4, lanes marked 1) or with the CXY fusion protein containing only the carboxy terminus of GLI (no zinc fingers) as a negative control (lanes marked 2). Examples of the results of such an experiment are shown in Fig. 4. The majority of insert fragment from clone A1 was bound by ZF fusion protein, whereas no insert fragment was bound by CXY fusion protein (Fig. 4, A lanes). In contrast, very little of the vector fragment was bound by the ZF fusion protein. A clone with a human TFIIIA binding site was tested and did not bind to either the ZF or CXY fusion protein (Fig. 4, N lanes). Also shown is the strong binding of clone B1 (Fig. 4, B lanes) and the weaker binding of clone C1 (Fig. 4, C lanes). Of the nine clones tested in this assay, seven bound specifically to the GLI fusion protein containing the zinc fingers.

We next determined the nucleotide sequence of these seven clones and found that the seven clones represented three sequence classes. Clones A1, A2, A3, and A4 represented independent clones of a 63-bp tandem repeat (Fig. 5A). Clones B1 and B2 were identical and contained a 62-bp insert, which was distinct from those of the class A clones. Clone C1 contained a 67-bp insert, which was different from those of both class A and B clones. A search of a nucleotide sequence data base (Genbank R59, April 1989) indicated that

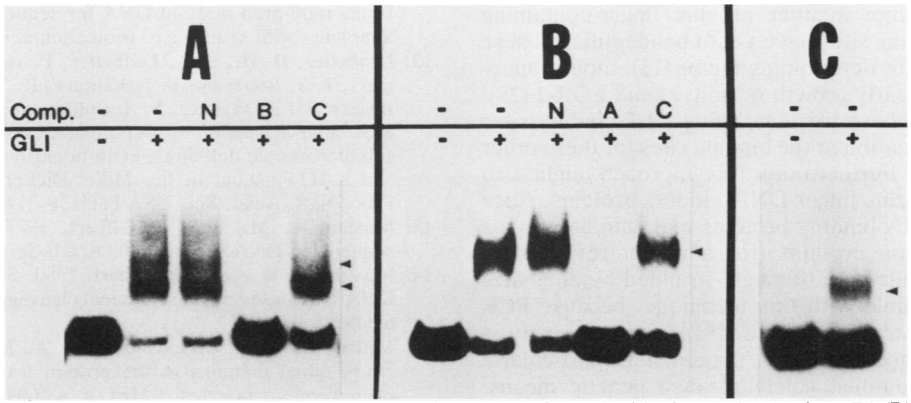

FIG. 8. Intact GLI protein binds to the three clones. Gel retardation was performed with clone A1 (A), clone B1 (B), or clone C1 (C). The protein used was a rabbit reticulocyte lysate programmed with *GLI* transcript (GLI +) or without exogenous transcript (GLI −). The competitors (COMP.) used were *Eco*RI-cleaved plasmids containing either an irrelevant nonbinding insert (sequences bound by TFIIIA [N lanes]), clone A1 (A lanes), clone B1 (B lanes), or clone C1 (C lanes. Arrowheads indicate the position of DNA protein complexes.

these sequences have not been previously reported. Southern blot analysis was performed by using a representative from each clone class. The analysis with clone A1 indicated that the number of repeats of the 63-bp sequence was highly variable among the human DNA samples tested (Fig. 6A). In the DNA from most individuals, two *Hae*III fragments were detected with this probe, although no *Hae*III sites were contained within the probe. Similarly, two variably sized fragments were observed with other enzymes including *Hin*fI and *Eco*RI. Thus, the 63-bp tandem repeat is a member of the variable number of tandem repeat (VNTR) family, and the alleles containing it are polymorphic. Southern blot analysis with clones B1 and C1 indicated they were represented only once (or a small number of times) in the genome; only one fragment was detected in Southern blots when human DNA was digested with *Eco*RI or other enzymes and the intensity of hybridization was that expected for a single-copy gene (Fig. 6B and C).

To determine the sequence elements responsible for GLI binding in these clones, we performed DNase footprinting. DNase footprinting was performed in the presence of the ZF fusion protein, as described in Materials and Methods. As a control, end-labeled insert from each clone was isolated from a gel and treated with DNase in a similar manner but without binding to a fusion protein. Figure 7 shows the results of DNase footprinting for both strands performed with a representative clone of each class. Clone A1 exhibited two protected regions of 24 bp in both strands (Fig. 7A and A'). The two protected regions corresponded to the same regions of the two tandem repeats contained within this clone. Clone B1 had a single protected region of 24 bp in both strands (Fig. 7B and B'). Clone C1 had a protected region of 23 bp on one strand but only 14 bp on the other strand (Fig. 7C and C'). The region which was protected by GLI are indicated for each clone in Fig. 5. The protected regions from clones A and B were 63% identical (Fig. 5D), whereas the protected region in clone C was 50 and 54% identical to the binding sites of A and B, respectively. All the binding sites had an invariant stretch of 9 nucleotides (5'-GACCACCCA-3'), which coincided with the regions of greatest similarity among the three clones.

To determine whether the binding of GLI zinc finger fusion protein was representative of the intact protein, we performed gel retardation with in vitro-translated GLI protein. The insert from clone A1 was almost completely retarded by the in vitro-translated GLI protein, but not by rabbit reticulocyte lysates without GLI protein (Fig. 8A). Furthermore, because of the repeated nature of this binding site, a second retarded band could be observed, presumably due to binding of a second GLI molecule. Competition experiments demonstrated that a 100-fold excess of a nonbinding clone had no effect on retardation of 2B (Fig. 8A, competitor N), whereas clone B1 (Fig. 8A, competitor B) totally blocked binding and clone C1 partially blocked binding (Fig. 8A, competitor C). Similarly, the insert from clone B1 was efficiently retarded by GLI protein. The results of competition experiments were similar to those observed with clone A1; a 100-fold excess of a nonbinding clone had no effect on retardation, whereas clone A1 totally blocked binding and clone C1 only partially blocked binding (Fig. 8B). As expected, only a small fraction of clone C1 was retarded (Fig. 8C). Thus, binding observed with the ZF fusion protein was representative of the binding observed with the intact GLI protein.

DISCUSSION

The amplification of *GLI* in some human cancers and its specific expression patterns suggest that *GLI* plays a role in human neoplasia and normal development. The presence of five zinc fingers in GLI suggests that GLI might act by binding to DNA. The nuclear localization of GLI and the identification of sequences to which GLI can bind add definitive support to this conjecture.

The three GLI binding sites identified had several notable characteristics. First, all three clones had protected regions that extended over 23 bp on at least one of the two strands. Second, the three different binding sites had significant similarity within the binding site, with an absolute conservation of the 9-bp sequence 5'-GACCACCCA-3'. This 9-bp region was not the sole determinant of binding, because the strength of binding varied considerably between clone C1 and the other clones. Third, the clone A type binding site was within a VNTR, resulting in a large but variable number of potential binding sites at this locus. The physiological significance of this observation is unclear, but VNTR sequences are thought to be hot spots for recombinational processes and are often in close proximity to expressed gene sequences (12). No proteins which can bind to VNTR sequences have been reported previously, but sites for other regulatory proteins are sometimes repeated within promotor regions (29).

There exists a large number of zinc finger-containing proteins whose binding sites have yet to be identified. These proteins include testes determining factor (18), mouse oncogene *evi-1* (16), and early growth response gene EGR-1 (27). The technique used here for identifying GLI binding sites should allow identification of the binding sites for these other zinc finger proteins. Furthermore, this approach might also be extended to non-zinc finger DNA-binding proteins, since other classes of DNA binding proteins maintain activity as β-galactosidase fusion proteins (for a recent review, see reference 24). Identification of single-stranded binding sites should also be possible with this technique, because PCR can be used to recover single-stranded sequences as easily as double-stranded sequences. This approach is particularly useful for genes identified solely through genetic means, since isolation of the binding site does not require purified proteins or specific antibodies. Moreover, these results demonstrate that it is possible to detect bound sequences which are present at low copy number within a complex genome such as that of the human. Although the delineation of specific sequences which bind to given regulatory proteins represents only the first step in the complex analysis of biological function, this step is an important one.

ACKNOWLEDGMENTS

We thank B. D. Nelkin and J. M. Ruppert for critical reading of this manuscript and Tina A. Gwiazda for its preparation.

This work was supported by Public Health Service training grant CA-09243 from the National Cancer Institute and Public Health Service grants CA-43460 and CA-35494 from the National Institutes of Health.

LITERATURE CITED

1. **Arheden, K., M. Ronne, N. Mandahl, S. Heim, K. W. Kinzler, B. Vogelstein, and F. Mitelman.** 1989. In situ hybridization localizes the human putative oncogene *GLI* to chromosome subbands 12q13.3-14.1. Hum. Genet. **82**:1–2.
2. **Bigner, S. H., J. Mark, and D. D. Bigner.** 1987. Chromosomal progression of malignant human gliomas from biopsy to establishment as permanent lines in vitro. Cancer Genet. Cytogenet. **24**:163–176.
3. **Brown, R. S., C. Sander, and P. Argos.** 1985. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. **186**:271–274.
4. **Chowdhury, K., U. Deutsch, and P. Gruss.** 1987. A multigene family encoding several "finger" structures is present and differentially active in mammalian genomes. Cell **48**:771–778.
5. **Feinberg, A. P., and B. Vogelstein.** 1984. Radiolabelling DNA fragments directly in agarose gels. Anal. Biochem. **137**:266–267.
6. **Fogh, J.** 1978. Cultivation, characterization, and identification of human tumor cells with emphasis on kidney, testis, and bladder tumors. Natl. Cancer Inst. Monogr. **49**:5–9.
7. **Hoffman, E. P., R. H. Brown, and L. M. Kunkel.** 1987. Dystrophin: the protein product of the Duchenne muscular dystrophy locus. Cell **51**:919–928.
8. **Kinzler, K. W., S. H. Bigner, D. D. Bigner, J. M. Trent, M. L. Law, S. J. O'Brien, A. J. Wong, and B. Vogelstein.** 1987. Identification of an amplified, highly expressed gene in a human glioma. Science **236**:70–73.
9. **Kinzler, K. W., J. M. Ruppert, S. H. Bigner, and B. Vogelstein.** 1988. The *GLI* gene is a member of the *Kruppel* family of zinc finger proteins. Nature (London) **332**:371–374.
10. **Kinzler, K. W., and B. Vogelstein.** 1989. Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. Nucleic Acids Res. **17**:3645–3653.
11. **Kraft, R., J. Tardiff, K. S. Krauter, and L. A. Leinwand.** 1988. Using mini-prep plasmid DNA for sequencing double stranded templates with sequenase. Biotechniques **6**:544–549.
12. **Ledbetter, D. H., S. A. Ledbetter, P. vanTuinen, K. M. Summers, T. J. Robinson, Y. Nakamura, R. Wolff, R. White, D. F. Barker, M. R. Wallace, F. S. Collins, and W. B. Dobyns.** 1989. Molecular dissection of a contiguous gene syndrome: frequent submicroscopic deletions, evolutionarily conserved sequences, and a HTF island in the Miller-Dieker chromosome region. Proc. Natl. Acad. Sci. USA **86**:5136–5140.
13. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
14. **Maxam, A. W., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
16. **Morishita, K., D. S. Parker, M. L. Mucenski, N. A. Jenkins, N. G. Copeland, and J. N. Ihle.** 1988. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3 dependent myeloid leukemia cell lines. Cell **54**:831–840.
17. **Nielsen, D. A., and D. J. Shapiro.** 1986. Preparation of capped RNA transcripts using T7 RNA polymerase. Nucleic Acids Res. **14**:5936.
18. **Page, D. C., R. Mosher, E. M. Simpson, E. M. C. Fisher, G. Mardon, J. Pollack, B. McGillivray, A. de la Chapelle, and L. G. Brown.** 1987. The sex-determining region of the human Y chromosome encodes a finger protein. Cell **51**:1091–1104.
19. **Preiss, A., U. B. Rosenberg, A. Kienlin, E. Seifert, and H. Jackle.** 1985. Molecular genetics of *Kruppel*, a gene required for segmentation of the *Drosophila* embryo. Nature (London) **313**: 27–32.
20. **Roberts, W. M., E. C. Douglass, S. C. Peiper, P. J. Houghton, and A. T. Look.** 1989. Amplification of the *GLI* gene in childhood sarcomas. Cancer Res. **49**:5407–5413.
21. **Ruppert, J. M., K. W. Kinzler, A. J. Wong, S. H. Bigner, F.-T. Kao, M. L. Law, H. N. Seuanez, S. J. O'Brien, and B. Vogelstein.** 1988. The *GLI*-Kruppel family of human genes. Mol. Cell. Biol. **8**:3104–3113.
22. **Schuh, R., W. Alcher, U. Gaul, S. Cote, A. Preiss, D. Maier, E. Seifert, U. Nauber, C. Schroder, R. Kemier, and H. Jackle.** 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *Kruppel*, a Drosophila segmentation gene. Cell **47**:1025–1032.
23. **Sealey, P. G., P. A. Whittaker, and E. M. Southern.** 1985. Removal of repeated sequences from hybridization probes. NAR **13**:1905–1922.
24. **Singh, H., R. G. Clerc, and J. H. LeBowitz.** 1989. Molecular cloning of sequence-specific DNA binding proteins using recognition site probes. Biotechniques **7**:252–261.
25. **Smith, D. R., I. J. Jackson, and D. D. Brown.** 1984. Domains of the positive transcription factor specific for the Xenopus 5S RNA gene. Cell **37**:645–652.
26. **Stanley, K. K., and J. P. Luzio.** 1984. Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. EMBO J. **3**:1429–1434.
27. **Sukhatme, V. P., X. Cao, L. C. Chang, C.-H. Tsai-Morris, D. Stamenkovich, P. C. Ferreira, D. R. Cohen, S. A. Edwards, T. B. Shows, T. Curran, M. M. Le Beau, and E. D. Adamson.** 1988. A zinc finger-encoding gene coregulated with c-*fos* during growth and differentiation, and after cellular depolarization. Cell **53**: 37–43.
28. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.
29. **Wingender, E.** 1988. Compilation of transcription regulating proteins. Nucleic Acids Res. **16**:1879–1902.
30. **Wong, A. J., S. H. Bigner, D. D. Bigner, K. W. Kinzler, S. R. Hamilton, and B. Vogelstein.** 1987. Increased expression of the epidermal growth factor receptor gene in malignant gliomas is invariably associated with gene amplification. Proc. Natl. Acad. Sci. USA **84**:6899–6903.