Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 668–672, January 1990
Biochemistry

The major human erythroid DNA-binding protein (GF-1): Primary sequence and localization of the gene to the X chromosome

(transcription factor/NF-E1/Eryf 1/erythrocytes/hereditary persistence of fetal hemoglobin)

LEONARD I. ZON*†, SHIH-FENG TSAI*†, SHAWN BURGESS*†, PAUL MATSUDAIRA‡, GAIL A. P. BRUNS§, AND STUART H. ORKIN*†

*Division of Hematology-Oncology, Children's Hospital, the Dana Farber Cancer Institute, and Department of Pediatrics, Harvard Medical School, Boston, MA 02115; †Howard Hughes Medical Institute, Boston, MA 02115; ‡Whitehead Institute for Biomedical Research, Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and §Division of Genetics, Children's Hospital, and Department of Pediatrics, Harvard Medical School, Boston, MA 02115

Communicated by William S. Sly, October 12, 1989 (received for review August 28, 1989)

ABSTRACT **Genes expressed in erythroid cells contain binding sites for a cell-specific nuclear factor, GF-1 (NF-E1, Eryf 1), believed to be an important transcriptional regulator. Previously we characterized murine GF-1 as a 413-amino acid polypeptide containing two cysteine-cysteine regions reminiscent of zinc-finger DNA-binding domains. By cross-hybridization to the finger domain of murine GF-1 we have isolated cDNA encoding the human homolog. Peptide sequencing of purified human GF-1 confirmed the authenticity of the human cDNA. The predicted primary sequence of human GF-1 is highly similar to that of murine GF-1, particularly in the DNA-binding region. Although the DNA-binding domains of human, murine, and chicken proteins are remarkably conserved, the mammalian polypeptides are strikingly divergent from the avian counterpart in other regions, most likely those responsible for transcriptional activation. By hybridization to panels of human–rodent DNAs we have assigned the human GF-1 locus to Xp21-11. The localization of the gene to the X chromosome has important implications for hereditary persistence of fetal hemoglobin syndromes unlinked to the β-globin cluster and for genetic experiments designed to test the role of the factor in erythroid cell gene expression.**

Human globins are encoded by structural genes residing in two distinct clusters, each organized 5′ to 3′ corresponding to their temporal expression during development (see ref. 1). In the α-globin cluster on chromosome 16, ≈30 kilobases (kb) of DNA contains the embryonic α-like globin (ζ) and duplicated adult α-globin genes. Within the 60 kb of the β-like cluster on chromosome 11, five active genes (ε, duplicated γ, δ, β) are situated 5′-ε-Gγ-Aγ-δ-β-3′. The different globin genes in man, as in other vertebrates, are expressed in a developmental-stage and tissue-specific manner. For example, the ε gene is expressed in yolk sac, the fetal Gγ and Aγ genes are expressed in fetal liver, and adult δ and β genes are expressed primarily in bone marrow. Furthermore, globin genes are transcriptionally active only within erythroid progenitor cells. Clinically significant anemias resulting from mutations affecting either globin structure (e.g., sickle cell anemia) or synthesis (e.g., thalassemias) are frequent and well understood at the molecular level (2). Much less common are syndromes associated with hereditary persistence of fetal hemoglobin expression into adult life (the HPFH syndromes). The molecular bases of these conditions are heterogeneous and include deletions in the β-globin gene cluster, single base substitutions in the γ-globin gene promoters, and mutations unlinked to the β-globin cluster itself (see ref. 1).

Although the molecular mechanisms responsible for controlling the developmental-stage and tissue-specific expression of globin genes are as yet incompletely understood, considerable progress has been made in defining important cis-regulatory elements. Experiments with transfected cultured cells and transgenic mice have identified cell-specific cis-acting DNA sequences located 5′ and 3′ to the transcription initiation sites of globin genes (3–6). Moreover, a dominant control region residing within a segment of DNase I hypersensitivity upstream of the embryonic ε-globin gene promotes high-level erythroid-specific expression of globin or heterologous genes (7–9).

It is widely believed that the developmental-stage and tissue-specific expression of globin genes reflects the interaction of the relevant cis-acting elements with nuclear regulatory proteins, some of which are cell- and/or developmental-stage specific in their activity or expression. In the past year we and others detected a sequence-specific DNA-binding activity in erythroid cells that appears to serve as the major regulator of erythroid-specific gene expression. This nuclear factor, termed GF-1, Eryf 1, and NF-E1, has been detected in extracts of human, mouse, and chicken erythroid cells and binds a core DNA sequence of the general form [(A/T)GATA(A/G)] that is found in the promoter and enhancer regions of numerous globin genes (10–13), at multiple sites within the dominant control region (8, 13), and within the promoters of several other erythroid-expressed nonglobin genes (14). Through expression cloning we recently isolated cDNA encoding murine GF-1 (13). Sequence analysis revealed a predicted polypeptide of 413 amino acids bearing two Cys-Cys domains reminiscent of zinc-finger structures. In addition, we presented preliminary evidence suggesting close similarity between murine and human homologs.

To initiate studies of the potential role of human GF-1 in globin gene expression and hemoglobin switching in man, we sought to isolate and characterize cDNA from human erythroid cells. Here we describe the primary sequence of human GF-1 cDNA, the correspondence of portions of the deduced protein with microsequenced protein, and assignment of the locus to the short arm of the X chromosome.¶

METHODS

cDNA Cloning and Characterization. A λgt11 cDNA library prepared from mRNA isolated from hemin-treated human erythroleukemia K562 cells was constructed by standard

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PCR, polymerase chain reaction; HPFH, hereditary persistence of fetal hemoglobin.

¶The sequence reported in this paper has been deposited in the GenBank data base (accession no. M30601).

procedures (15). The library, containing >1.5 × 10^6 independent recombinants, was screened with a portion of murine GF-1 cDNA that spanned the putative zinc-finger domain. The 468-base pair (bp) probe was generated by the polymerase chain reaction (PCR) using primers spanning amino acids 139–147 and 289–294 of the murine sequence (16). After hybridization, the filters were washed at reduced stringency (0.15 M NaCl/15 mM sodium citrate at 52°C). After two rounds of plaque purification, cDNA inserts were recovered by PCR using λgt11 sequencing primers and subcloned into PUC-13 for DNA sequencing (17). Seven independent cDNA clones were initially obtained and fully characterized. Northern blot analysis was performed by standard methods (15).

Protein Purification and Peptide Sequencing. Human GF-1 protein was purified by oligonucleotide affinity chromatography from nuclear extracts of K562 cells as described (13). The final preparation contained three major polypeptides (50, 38, and 20 kDa) that bound the human Aγ promoter in a sequence-specific manner as well as a contaminating 115-kDa protein. The 38- and 20-kDa species are believed to represent proteolytic fragments of a native 50-kDa protein. Approximately 5 μg of the affinity-purified protein was digested with endoproteinase Lys-C (Boehringer Mannheim). The resulting peptides were resolved by HPLC on a Brownlee Aquapore RP300 column. Peptides from selected peak fractions were adsorbed onto Polybrene-coated glass fiber discs and sequenced in an Applied Biosystems model 475 sequencer equipped with on-line phenylthiohydantoin amino acid analysis. The amount of sequenceable peptide varied from 10 to 30 pmol.

Chromosomal Localization. DNAs from a panel of 13 rodent–human hybrids containing various human chromosomes were digested with *Hin*dIII and subjected to Southern blot analysis (18). For initial localization, human GF-1 cDNA spanning the Cys-Cys finger region was used as probe. In subsequent analysis (see *Results*) of a second panel of hybrids that had retained subregions of the X chromosome by virtue of X/autosome translocations (19–21), a genomic fragment encompassing the terminal exon of the human gene was employed. This fragment was isolated from a genomic library of normal female DNA in the bacteriophage λEMBL3 (unpublished data).

RESULTS

Cloning of Human GF-1 cDNA. Previously we described weak cross-hybridization of murine GF-1 cDNA to an erythroid-specific 1.8-kb mRNA in human erythroleukemia K562 or HEL cells (13). Screening of the K562 cDNA library with full-length murine GF-1 cDNA under reduced stringency yielded cDNA clones bearing no apparent relationship to the murine sequence (not shown). As we suspected that the putative DNA-binding domains of the murine and human proteins might be highly conserved, we subsequently screened the K562 cDNA library with a PCR-generated fragment specific for the two Cys-Cys finger domains (see *Materials and Methods*). By using this probe, strong cross-hybridization of human mRNA was observed (not shown) and homologous cDNA clones were readily obtained.

The DNA sequence of the largest clone (designated K101) is displayed in Fig. 1. This cDNA encodes the entire translated portion of human GF-1 mRNA and predicts a polypeptide identical in size (413 amino acids; predicted 42 kDa) to the murine homolog. As shown in Fig. 2, the peptide sequence encompassing the two Cys-Cys fingers is remarkably conserved between human and mouse; only a single difference is predicted in the 78 amino acids between the Cys-Cys pairs. Strong conservation of peptide sequences is also evident C-terminal to the second finger domain (amino acids

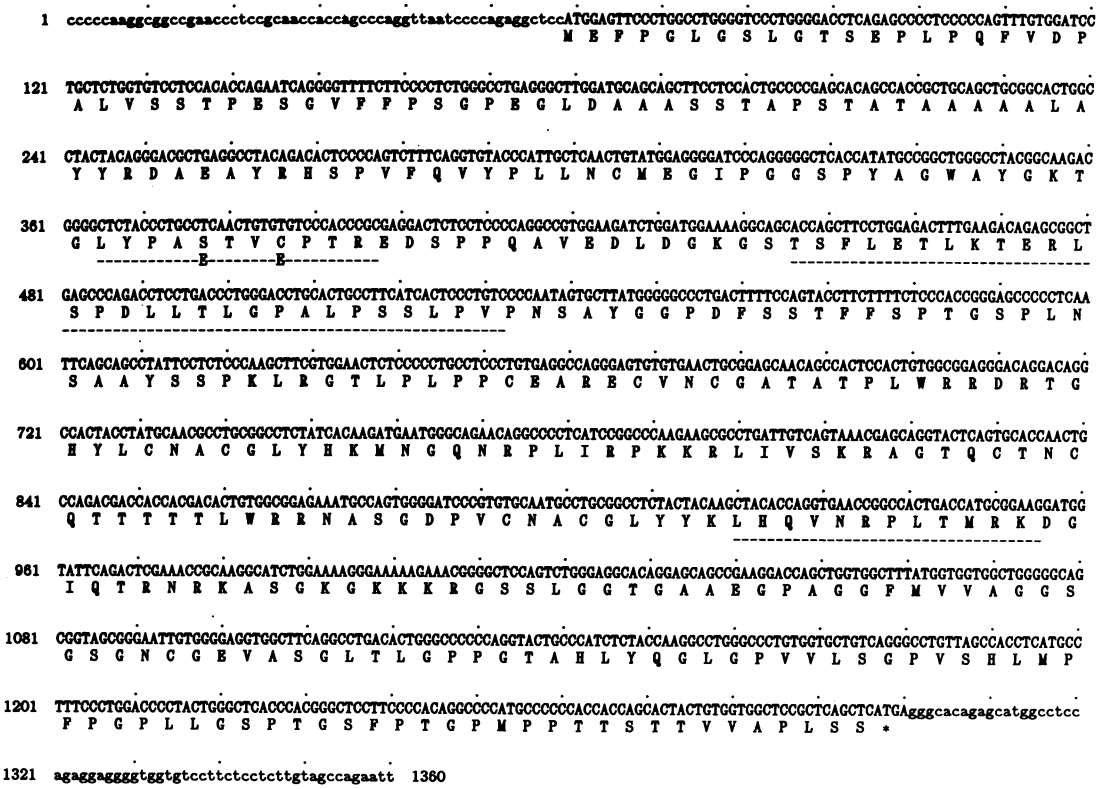


FIG. 1. DNA sequence of human GF-1 cDNA (clone K101) and its predicted protein. The dashed lines (---) denote regions for which direct peptide sequences were obtained (see Fig. 4*b*). Amino acids that differ from the predicted sequence are indicated below the corresponding predicted amino acid. An *Eco*RI site at the 3′ extent of clone K101 is also present at the same relative position in a human genomic GF-1 clone (unpublished data).

```
                1
HUMAN  GF1      MEFPGLGSLGTSEPLPQFVDPALVSSTPESGVFFPSGPEGLDAAASSTAPSTATAAAAALAYYRDAEAYRHSPVFQVYPLLNCMEGIPGGSPYAGWAYGK
MOUSE  GF1      MDFPGLGALGTSEPLPQFVDSALVSSPSDSTGFFSSGPEGLDAASSTSPNAATAAASALAYYREAEAYRHSPVFQVYPLLNSMEGIPGGSPYASWAYGK
CHICKEN ERYF1   MEFVALG..G.PDAGSPTPFPDEAGA.....FLGLGGGE.....................................................RTEAGGL

              101
HUMAN  GF1      TGLYPASTVCPTREDSPP......QAVEDLDGKGSTSFLETLKTERLSPDLLTLGPALPSSLPVPNSAYGGPDFSSTFFSPTGSPLNSAAYSSPKLRGTL
MOUSE  GF1      TALYPASTVCPSHEDAPS......QALEDQEGKSNNTFLDTLKTERLSPDLLTLGTALPASLPVTGSAYGGADFPSPFFSPTGSPLSSAAYSSPKFHGSL
CHICKEN ERYF1   LASYPPSGRVSLVPWADTGTLGTPQWVPPATQMEPPHYLELLQPPRGSPPH.......PSSGPLL.....................PLSSGP..........

              201
HUMAN  GF1      PLPPCEARECVNCGATATPLWRRDRTGHYLCNACGLYHKMNGQNRPLIRPKKRLIVSKRAGTQCTNCQTTTTTLWRRNASGDPVCNACGLYYKLHQVNRP
MOUSE  GF1      PLAPCEARECVNCGATATPLWRRDRTGHYLCNACGLYHKMNGQNRPLIRPKKRMIVSKRAGTQCTNCQTTTTTLWRRNASGDPVCNACGLYFKLHQVNRP
CHICKEN ERYF1   ..PPCEARECVNCGATATPLWRRDGTGHYLCNACGLYHRLNGQNRPLIRPKKRLLVSKRAGTVCSNCQTSTTTLWRRSPMGDPVCNACGLYYKLHQVNRP

              301
HUMAN  GF1      LTMRKDGIQTRNRKASGKGKKKRGSSLGGTGAAEGPAGGFMVVAGGSGSGNCGEVASGLTLGPPGTAHLYQGLGPVVLSGPVSHLMPFPGPLLGSPTGSF
MOUSE  GF1      LTMRKDGIQTRNRKASGKGKKKRGSNLAGAGAAEGPAGGFMVVAGSSSSGNCGEVASGLALGTAGTAHLYQGLGPVVLSGPVSHLMPFPGPLLGSPTTSF
CHICKEN ERYF1   LTMRKDGIQTRNRKVSSKGKKRRPP...GGGNPSATAGGGAPM.GGGGDPSMPPPPPPPAAAPPQSDALY.ALGPVVLSG...HFLPF.GNSGGFFGGGA

              401
HUMAN  GF1      PTGPMPPTTSTTVVAPLSS
MOUSE  GF1      PTGPAPTTSSTSVIAPLSS
CHICKEN ERYF1   GGYTAPPGLSPQI......
```

FIG. 2. Species comparison of the predicted amino acid sequences of human, mouse, and chicken GF-1s. The highly conserved finger region is underlined. Other areas of similarity between species are boxed.

288–324) in a region that is likely to contribute to sequence-specific DNA binding (22).

The remarkable conservation of the Cys-Cys finger domains is apparent by comparison of the two mammalian sequences with that of the analogous chicken protein (designated Eryf 1), recently described by Evans and Felsenfeld (Fig. 2, underlined) (23). Outside the finger domains the human and mouse polypeptides are 83% identical and yet divergent from the predicted chicken protein except for small segments (Fig. 2, boxed).

Human GF-1 cDNA strongly hybridizes to the 1.8-kb K562 and HEL mRNA species that was weakly detected previously with the entire murine cDNA as probe (Fig. 3). The mRNA is absent in non-erythroid cell lines. This tissue-restricted pattern of GF-1 mRNA is identical to that observed for the murine homolog (13).

Microsequencing of Human GF-1. As different proteins, often containing related DNA-binding domains, may recognize the same target DNA sequence, we sought to demonstrate that the cDNAs we obtained for human (and mouse) GF-1 encode the purified erythroid nuclear factor rather than another protein. Previously, we presented evidence that one peptide sequence derived from purified human K562 cell GF-1 (peptide 7 in Fig. 4) was identical to the predicted sequence of murine GF-1 just carboxyl to the putative zinc-finger domain. As this correspondence does not formally exclude a highly homologous DNA-binding domain shared between two different proteins, analogous to the relationship of the octamer-binding proteins known as Oct-1 and Oct-2 (24), we examined additional peptides of purified human GF-1. These peptides, designated nos. 41 and 44 (Fig. 4), provide more definitive evidence that the human GF-1 polypeptide sequence derived from cDNA corresponds to that of authentic GF-1. Specifically, the 20-amino acid peptide no. 41

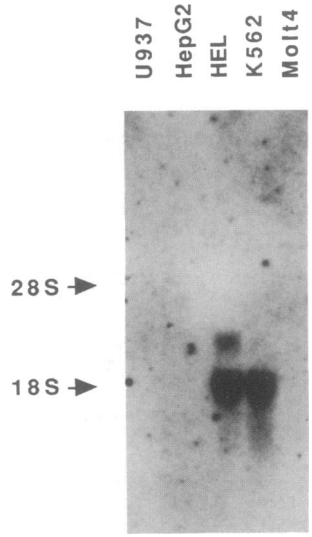
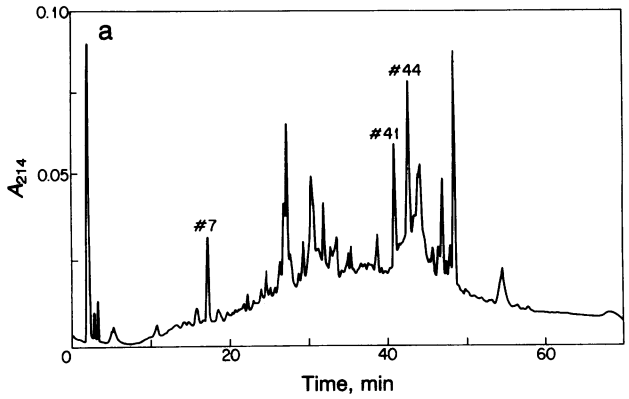


FIG. 3. Cell-specific expression of human GF-1 mRNA. Total cell RNAs (10 μg) of the indicated cell lines were subjected to Northern blot analysis and hybridized with the complete human GF-1 cDNA. U937, monocytic leukemia; HepG2, hepatoma; Molt4, T-cell leukemia. The slightly larger, additional GF-1 mRNA species evident in the HEL cell sample is also seen in phorbol 12-myristate 13-acetate-induced K562 cells (not shown). HeLa and myeloid (HL60) cell lines lack detectable GF-1 RNA (not shown).



b

Peptide #7

L H Q V N R P L T M R K

Peptide #41

E R L S P D L L T L G P A L P S S L P V

Peptide #44

L Y P A (E) T V (E) P T R
T S F L E T L K T E R

FIG. 4. Microsequencing of human GF-1 peptides. (*a*) Peptides generated by endoproteinase Lys-C digestion were resolved on a C_8 reversed-phase column. A 60-min linear gradient of 0–80% acetonitrile was used at 0.4 ml/min flow rate, and peptide concentration was monitored by absorbance at 214 and 280 nm. (*b*) Amino acid sequences of human GF-1 peptides. Fraction 44 contained two major peptides present in 10-pmol amounts. The sequence of each peptide was deduced from comparisons with the protein sequence of the cDNA clone. Residues differing from predicted sequences are in parentheses.

matched the predicted human protein sequence in a distant region. The second peptide (no. 44) could be deduced to be a mixture of two peptides positioned just N-terminal to peptide no. 41. On the basis of these data, we conclude that the human GF-1 cDNA (and hence the previously isolated murine counterpart) encodes authentic GF-1 protein and not an unrelated protein bearing a closely related DNA-binding domain.

Assignment of the GF-1 Locus to Xp. As the chromosomal position of the human GF-1 locus might provide insights into the basis of inherited variation in hemoglobin F production not due to mutations within the β-globin locus, we determined its location by Southern blot hybridization of human GF-1 cDNA to a panel of rodent–human hybrid cell DNAs. A provisional assignment to the X chromosome was achieved (discordancy values: X chromosome = 0.00; autosomes and Y chromosome = 0.25–0.77). To explore this further, we examined a mapping panel designed to further localize the gene on the X chromosome. To facilitate mapping a genomic fragment derived from the 3′ portion of the human GF-1 gene was used as probe. Consistent with assignment to the X chromosome, the intensity of hybridization was greater with normal female than male DNA (Fig. 5*a*, lanes 1 and 2). Positive hybridization was obtained only for those hybrids retaining Xp21-11 (Fig. 5*a*, lanes 3–5, 7, and 8), as summarized in Fig. 5*b*.

DISCUSSION

In this paper we have described the primary structure of human GF-1, a tissue-specific DNA-binding protein that recognizes a sequence motif widely distributed in the promoters and enhancers of numerous erythroid-expressed genes. On the basis of its presence at all stages of erythroid cell development and evidence implicating its role in the function of the chicken and human 3′ β-enhancers and in promoter activity of the human γ-globin and murine α-globin genes (10–14), GF-1 is an excellent candidate to serve as the major regulator of gene expression in the erythroid lineage. In the studies presented here we have isolated human cDNA by cross-hybridization to the DNA-binding domain of the murine homolog and used direct peptide sequencing of purified human material to authenticate our characterization of the relevant protein.

Several features of the primary sequence deserve particular comment. Overall, the murine and human polypeptides are 89% identical. The conservation is extraordinary within the region encompassed by the duplicated Cys-Cys fingers and just carboxyl to it (Fig. 2). In other studies (D. I. K. Martin and S.H.O., unpublished data) we have shown that this segment of the murine protein is required and sufficient for sequence-specific DNA binding. In accord with this, this region displays remarkable conservation in the avian homolog Eryf 1. Moreover, a *Xenopus* homolog we have cloned also bears comparable conservation in the DNA-binding region (L.I.Z., R. Harland, S.H.O., unpublished data). As the transcriptional activator function of the GF-1 molecule lies outside the Cys-Cys finger region (D. I. K. Martin and S.H.O., unpublished data), it is equally notable that the human and murine proteins are highly similar outside this domain but only distantly related to chicken Eryf 1. Results from the expression of human globin genes introduced into cultured mouse erythroid cells or the mouse germ line, and cross-species trans-activation of globin genes following heterokaryon formation, support the existence of regulatory factors that are closely related in structure and function between mouse and man. Further study of the divergent regions of the mammalian, avian, and amphibian GF-1 molecules is likely to be relevant to an understanding of the differences in regulation of specific globin genes among these species. Moreover, analysis of those segments that retain structural similarity (Fig. 2, boxed) may provide clues to functions in common among GF-1s of diverse species.

In addition to its broad role in regulation and coordination of gene expression in erythroid cells, the potential involvement of GF-1 in the switching of globin genes during development is unknown. We have proposed previously that the widespread distribution of GF-1-binding sites throughout the human β-globin gene cluster, including the dominant control region and the promoters and enhancers of the β- and γ-globin genes, suggests a possible role for GF-1 in hemoglobin switching (13), in accord with models in which promoters and enhancers are envisioned to compete and/or interact physically by means of protein–protein interactions (25). The assignment of the human GF-1 locus to Xp21-11 is therefore provocative in view of the recent description of chromosome X-linked dominant inheritance of the Swiss form of HPFH in the Japanese population (26). Inherited differences in the structure or expression of a regulatory factor, such as GF-1, could account for this or other varieties of HPFH (27) that are genetically unlinked to the β-globin locus. Linkage analysis is needed to assess any relationship between the GF-1 gene and chromosome X-linked forms of HPFH. As we have not observed restriction fragment length polymorphism of the human GF-1 gene (unpublished data), linkage analysis using other Xp markers will be required. In

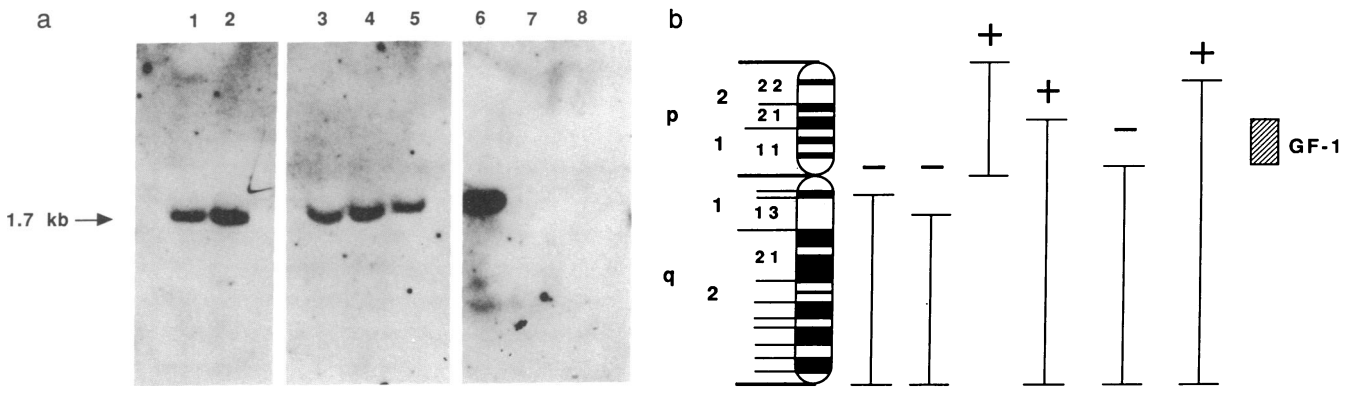


FIG. 5. Assignment of the GF-1 locus to Xp. (*a*) Southern blot hybridization of human and human–rodent hybrid cell DNAs with a genomic fragment of the GF-1 locus. Lanes: 1, normal male; 2, normal female; 3–5, hybrids retaining Xp21.2-qter, Xpter-cen., Xp22.3-qter, respectively; 6, normal female DNA; 7 and 8, hybrids retaining Xp11-qter and Xq12-qter, respectively. (*b*) Summary of assignment of human GF-1 locus to Xp21-11. Extents of the portions of the X chromosome retained in rodent–human hybrid cells are displayed to the right of a schematic representation of the X chromosome. +, Presence of the GF-1 locus; −, absence of the GF-1 locus. The minimal area of overlap is shown to the far right.

the mouse GF-1 is also encoded on the X chromosome (V. Chapman and S.H.O., unpublished data).

Apart from its potential relevance to HPFH states, the assignment of the GF-1 locus to the X chromosome has important implications for strategies designed to evaluate the role of the protein in the development of erythroid cells and coordination of tissue-specific gene expression. Specifically, hemizygosity for the GF-1 gene should facilitate experiments designed to test the consequences of site-specific disruption or modification of the locus in cultured erythroid cells and mouse embryonic stem cells (28).

We thank Todd Evans and Gary Felsenfeld for communicating the Eryf 1 sequence prior to publication. S.H.O. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by grants from the National Institutes of Health.

1. Stamatoyannopoulos, G. & Nienhuis, A. W. (1987) in *The Molecular Basis of Blood Diseases*, eds. Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. & Majerus, P. (Saunders, Philadelphia), pp. 66–105.
2. Orkin, S. H. (1987) in *The Molecular Basis of Blood Diseases*, eds. Stamatoyannopoulos, G., Nienhuis, A. W., Leder, P. & Majerus, P. (Saunders, Philadelphia), pp. 106–126.
3. Kollias, G., Hurst, J., deBoer, E. & Grosveld, F. (1987) *Nucleic Acids Res.* **15,** 5739–5747.
4. Townes, T. M., Lingrel, J. B., Chen, H. Y., Brinster, R. L. & Palmiter, R. D. (1985) *EMBO J.* **4,** 1715–1723.
5. Hess, J. E., Nickol, J. M., Lieber, M. R. & Felsenfeld, G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4312–4316.
6. Behringer, R., Hammer, R. E., Brinster, R. L., Palmiter, R. D. & Townes, T. M. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7056–7060.
7. Grosveld, F., van Assendelft, G. B., Greaves, D. R. & Kollias, B. (1987) *Cell* **51,** 975–985.
8. Forrester, W. C., Novak, U., Gelinas, R. & Groudine, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5439–5443.
9. Ryan, T. M., Behringer, R. R., Townes, T. M., Palmiter, R. D. & Brinster, R. L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 37–41.
10. Evans, T., Reitman, M. & Felsenfeld, G. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 5976–5980.
11. Martin, D. I. K., Tsai, S.-F. & Orkin, S. H. (1989) *Nature (London)* **338,** 435–438.
12. Wall, L., deBoer, E. & Grosveld, F. (1988) *Genes Dev.* **2,** 1089–1100.
13. Tsai, S. F., Martin, D. I., Zon, L. I., D'Andrea, A. D., Wong, G. G. & Orkin, S. H. (1989) *Nature (London)* **339,** 446–451.
14. Plumb, M., Frampton, J., Wainwright, H., Walker, M., Macleod, K., Goodwin, G. & Harrison, P. (1989) *Nucleic Acids Res.* **17,** 73–92.
15. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
16. Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1986) *Nature (London)* **324,** 163–166.
17. Dorfman, D. M., Zon, L. I. & Orkin, S. H. (1989) *BioTechniques* **7,** 568–570.
18. Bruns, G., Stroh, H., Veldman, G. M., Latt, S. A. & Floros, J. (1987) *Hum. Genet.* **76,** 58–62.
19. Wieacker, P., Davies, K. E., Cooke, H. J., Pearson, P. L., Williamson, R., Bhattacharya, S., Zimmer, J. & Ropers, H. H. (1984) *Am. J. Hum. Genet.* **36,** 265–276.
20. Mohandas, T., Shapiro, L. J., Sparkes, R. S. & Sparkes, M. C. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5779–5783.
21. Kunkel, L. M., Monaco, A. P., Middlesworth, W., Ochs, H. D. & Latt, S. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4778–4782.
22. Pfeifer, K., Kim, K. S., Kogan, S. & Guarente, L. (1989) *Cell* **56,** 291–301.
23. Evans, T. & Felsenfeld, G. (1989) *Cell* **58,** 877–885.
24. Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G. & Horvitz, H. R. (1988) *Genes Dev.* **2,** 1513–1516.
25. Choi, O.-R. & Engel, J. D. (1988) *Cell* **55,** 17–26.
26. Miyoshi, K., Kaneto, Y., Kawai, H., Ohchi, H., Niki, S., Hasegawa, K., Shirakami, A. & Yamano, T. (1988) *Blood* **72,** 1854–1860.
27. Gianni, A. M., Bregni, M., Cappellini, M. D., Fiorelli, G., Taramelli, R., Giglioni, B., Comi, P. & Ottolenghi, S. (1983) *EMBO J.* **2,** 921–925.
28. Capecchi, M. R. (1989) *Science* **244,** 1288–1292.