J. Biochem. 124, 389-395 (1998)

Isolation and Initial Characterization of GBF, a Novel DNA-Binding Zinc Finger Protein That Binds to the GC-Rich Binding Sites of the HIV-1 Promoter[1]

Toru Suzuki,*,†,2 Tohru Yamamoto,* Masahiko Kurabayashi,‡ Ryozo Nagai,‡ Yoshio Yazaki,† and Masami Horikoshi*,§,3

*Department of Developmental Biology, Department of Cellular Biology, Institute of Molecular and Cellular Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-0032; †Third Department of Internal Medicine, Faculty of Medicine, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8655; ‡Second Department of Internal Medicine, Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi, Gunma 371-0034; and §Horikoshi Gene Selector Project, Exploratory Research for Advanced Technology, Japan Science and Technology Corporation, 5-9-6 Tokodai, Tsukuba, Ibaraki 300-2635

Received for publication, March 10, 1998

Human immunodeficiency virus (HIV) and its clinical syndrome, acquired immune deficiency syndrome (AIDS), are one of the world's most prominent health problems. To understand the mechanisms underlying HIV transcription and thereby its propagation, we have focused on the molecular interactions at the GC-rich binding sites of the HIV-1 core promoter, a region important for HIV-1 transcription. Previous biochemical studies have shown that Sp1, a zinc finger transcription factor initially isolated as a cellular factor binding that binds to the SV40 early promoter GC-rich sequence, binds to the HIV-1 GC-rich binding sites due to sequence similarities. However, the HIV-1 GC-rich binding sites are considerably different from the Sp1 consensus binding sequence, and recent genetic studies have shown the lack of regulation by Sp1 in numerous genes thought to be regulated by that factor in the past. We reasoned that other factors may bind to the HIV-1 GC-rich binding sites. Using the native HIV-1 GC-rich binding sequence as the bait, genetic screening for interacting factors was performed by the yeast one-hybrid method. A cDNA encoding a novel zinc finger protein named GBF, GC-rich sites binding factor, was isolated from a human peripheral blood leukocyte library. Primary structure analysis of GBF revealed a C_2H_2 Krüppel-type zinc finger at its C-terminus, and putative acidic and proline-rich domains at its N-terminus. We also show that GBF belongs to a subgroup of Krüppel-type zinc fingers distinct from Sp1. By directly addressing interactions at the HIV-1 GC-rich binding sites, our present study sheds new light on molecular interactions at the HIV-1 promoter.

Key words: DNA binding protein, DNA-protein interaction, GC-box, HIV, zinc finger.

Human immunodeficiency virus (HIV), the causative agent of the global epidemic acquired immune deficiency syndrome (AIDS), is a prime focus of medical research as its pathogenic mechanisms are not yet comprehensively understood (*1*). We have focused on understanding the mechanisms underlying HIV transcription, as regulation is primarily controlled at the transcription level (*2*).

Transcription involves a milieu of DNA-protein/protein-protein interactions assembling multi-protein complexes at enhancer/core promoter elements which act in concert to activate the basal RNA polymerase machinery. Additional mechanisms of alteration of the chromatin structure are seen in eukaryotes. Viruses, which have conserved basic elements of eukaryotic transcriptional mechanisms, are well suited for investigations because of their simplicity. The HIV-1 core promoter is compact, containing a canonical TATA element, GC-rich binding sites and NFκB binding sites located within the first 100 bp upstream of the transcription initiation site (*3*); these DNA binding elements are conserved across species, and are found in both human and simian immunodeficiency viruses.

Numerous past studies have concerned the mechanisms of transcriptional regulation at these DNA binding elements. Biochemical studies on the GC-rich binding sites have shown the binding of cellular zinc-finger transcription factor Sp1, which was initially isolated as an activity that

[1] This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Science, Sports and Culture of Japan, and the Exploratory Research for Advanced Technology (ERATO) of the Japan Science and Technology Corporation (JST), in addition to grants from the Asahi Glass Foundation, the Naito Foundation, the New Energy and Industrial Technology Development Organization (NEDO), and the Toray Science Foundation.
[2] T.S. is a Research Fellow of the Japanese Society for the Promotion of Science, and was funded in part by the Kowa Life Science Foundation.
[3] To whom correspondence should be addressed. Tel: +81-3-5802-3388, Fax: +81-3-5684-8341, E-mail: horikosh@imcbns.iam.u-tokyo.ac.jp

stimulated activation of the SV40 early promoter, to the HIV-1 GC-rich sites owing to sequence similarities (*4-8*). We reasoned that other factors may bind to the HIV-1 GC-rich binding sites; (i) the HIV-1 GC-rich binding sites are considerably different from the consensus binding sequence of Sp1, suggesting the possibility of binding (*e.g.* differential affinity) of other sequence-specific DNA-binding factors (see Fig. 1a for a schematic diagram); (ii) recent studies have shown that similar yet functionally distinct zinc finger transcription factors bind to similar if not identical GC-rich binding sites (*9-11*); and (iii) genetic loss-of-function studies in mice showed the lack of regulation by Sp1 in numerous genes thought to be regulated by that factor in the past (*12*).

To understand the precise molecular interactions at the HIV-1 GC-rich binding sites, genetic screening involving the native HIV-1 GC-rich binding sequence as the bait was performed to identify interacting factors by the yeast one-hybrid method. A cDNA encoding a factor named GBF, GC-rich sites binding factor, was isolated. GBF is a novel Krüppel-type zinc finger protein which belongs to a subgroup of Krüppel-type zinc fingers distinct from Sp1. Our findings shed new light on the mechanisms underlying gene regulation of HIV-1, and on the DNA-protein interactions caused by zinc finger factors.

METHODS

Yeast One-Hybrid Screening—The DNA sequence spanning the GC-rich binding sites of HIV was synthesized with *Eco*RI/*Sma*I sites at the 5′ and 3′ ends (sense, 5′-AATCCG-AGGCGTGGCCTGGGCGGGACTGGGGAGTGGCCCC-3′; antisense, 5′-GGGGCCACTCCCCAGTCCCGCCCAGGCC-ACGCCTCG-3′), annealed, and then subcloned into the *Eco*RI/*Sma*I sites of vectors pHISi and pLacZi (both from Clontech), which contained the *HIS3* and *lacZ* genes downstream of the inserted binding site, respectively. pHISi and pLacZi also harbored the *HIS3* and *URA3* genes, respectively, to facilitate site-selective homologous recombination. In the presence of activating factors, *HI3* and *lacZ* gene expression is increased, allowing for dual selection as to 3-amino-triazole (3-AT) and β-gal activity, respectively. Reporters were transformed into yeast strain YM4271 (*MATa, ura3-52, his3-200, ade2-101, lys2-801, leu2-3, 112, trp1-903, tyr1-501*) by the lithium acetate method (*13*). The *HIS3* and *lacZ* reporters were integrated selectively into the genome at the *his* and *ura* loci of YM4271, respectively. YM4271 strains harboring the *HIS3* and *lacZ* reporters for HIV and p53 binding sites (p53 constructs supplied by the manufacturer) are referred to as YM4271/HIV and YM4271/p53, respectively. Preparation of the transformant with the *HIS3* and *lacZ* reporters was confirmed by the prototrophy to histidine (His) and uracil (Ura) on SC plates lacking His and Ura, respectively. Site-specific homologous recombination was confirmed by means of the polymerase chain reaction using primers specific to flanking genomic sequences. After determining the background level of the expression of YM4271/HIV, a human peripheral blood leukocyte library in the λ-ACT vector, which harbors the GAL4 activation domain fused N-terminal to the clone (kindly provided by S. Elledge), was screened. The plasmid library was transformed into YM4271/HIV, and positive clones were selected on SC plates lacking leucine but containing 50 mM 3-AT. Positive clones were further selected as to β-galactosidase activity essentially according to the method described previously (*14*). In each case, p53 was used as a control. Plasmids containing positive clones were isolated from yeast using the glass-beads technique and transformed into *Escherichia coli* strain DH5α. Plasmid DNA was further isolated using standard techniques, and the nucleotide sequence of the clone was determined by the dye terminator method (Applied Biosystems Model 377 sequencer). Nucleotide sequences were confirmed for both strands.

Co-transfection assays were performed using essentially the same techniques. The plasmid containing the positive clone was transformed into yeast and selected as to prototrophy for leucine on SC plates lacking leucine, and then subjected to the β-galactosidase assay.

Isolation of the Full Length cDNA Clone—To isolate the full length clone, plaque hybridization using the peripheral blood leukocyte λ-act library was performed. After infection of *E. coli* strain LE 392, plaques were transferred to Hybond N nylon filters (Amersham), fixed with alkali, and then UV-crosslinked. One million clones were screened under hybridization conditions of 50% formamide, 500 mM sodium-phosphate buffer (pH 7.0), 0.1% SDS, 1 M NaCl, 5×Denhardt's solution, and 50 μg/ml denatured salmon sperm DNA. After initial preincubation for 2 h at 42°C, a random-primed ^{32}P-labeled 1.3 kb *Xho*I insert of the positive clone was added as a probe and the filters were hybridized at 42°C for 16 h. The filters were washed in 2× SSC/0.05% SDS at room temperature, followed by further washing with 0.5×SSC/0.05% SDS at 42°C. The filters were then subjected to autoradiography for 24 h. Following isolation of single plaques by two repeated screenings, *E. coli* strain BNN132 was infected with the phage supernatant, and after 30 min incubation at 30°C, the plasmid was recovered. The plasmid DNA was further transformed into *E. coli* strain DH5α and isolated using standard techniques. The nucleotide sequence of the positive clone was determined using the dye terminator method (Applied Biosystems Model 377 sequencer). Nucleotide sequences were confirmed for both strands.

Northern Blot Analysis—A filter with mRNA from various cell lines was obtained from Clontech. Following preincubation of the filter in ExbressHyb hybridization solution (Clontech) at 68°C for 30 min, a random-primed ^{32}P-labeled 1.3 kb *Xho*I insert, as a probe, was added, followed by further incubation for 1 h at 68°C. The filter was washed with 2×SSC/0.5% SDS followed with 0.1× SSC/0.5% SDS at 68°C. The filter was then subjected to autoradiography for 48 h. The GAPDH probe was used as a control for expression.

Analysis of the Intracellular Localization of the Isolated Factor—The 1.3 kb *Xho*I fragment was subcloned in-frame into pEGFP-C1 (Clontech) containing a red-shifted green fluorescent protein fused N-terminal to the clone. HeLa cells were cultured to subconfluency in DMEM (Nissui) supplemented with 5% FCS (Gibco). pEGFP-C1 (1 μg) with the 1.3 kb insert was tranfected using Tfx-50 reagent (Promega), and cells were observed by microscopy at 24 h. Hoechst nuclear staining was performed as a control for localization of the nucleus.

Expression of the Recombinant Protein—The zinc finger putative DNA binding domain with the flanking regions

(from 17 amino acids N-terminal of the initial Cys residue of the first zinc finger) of GBF was subcloned into vector pGEX4T-1 (Pharmacia Biotech) at the *Bam*HI site following PCR-mediated mutagenesis using the following primers to introduce *Bam*HI restriction enzyme sites into the flanking regions (sense, 5′-CGGGATCCCATATGGACAAGGGAAATGGCGATGC-3′; antisense, 5′-GCCTACAGGATCCACCTCTC-3′). The GST-fusion construct was transfected into *E. coli* strain BL21(DE3)pLysS, and grown in M9 medium containing 100 μM $ZnSO_4$ at 27°C until the OD_{595} reached 0.6-0.8, at which time IPTG (0.4 mM) was added. The cells were harvested after further growth at 27°C for 3 h.

The cells were washed in 20 mM Tris Cl (pH 7.9)/200 mM NaCl, and then suspended in Buffer A [20 mM Tris Cl (pH 7.4), 10% glycerol, 500 mM KCl, 0.01% NP-40, and 50 mM β-mercaptoethanol]. Following sonication, NP-40 was added to 0.1%, and the cells were centrifuged at 18.5 krpm for 30 min. Glutathione-Sepharose resin (Pharmacia Biotech) was added to the lysate, followed by incubation for 2 h. After repeated washings with Buffer A, the recombinant protein was eluted using Buffer A containing 20 mM reduced glutathione (Wako).

Products were analyzed by 12% SDS-PAGE by staining with Coomassie blue. Protein concentrations were estimated using the Bio-Rad protein assay reagent (Bio-rad). All procedures concerning proteins were performed at 4°C.

Gel Shift Assay—The double stranded sequence of the bait for the one-hybrid screening was used as the probe for gel shift analysis of DNA-protein binding (top 5′-TTTCCAGGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGTCCCTCA-3′). The consensus p53 sequence was used as a negative control in binding reactions (5′-GGCATGCCTAGGCATGCCTAGGCATGCCTAGGCATGCC-3′). Hybridized double-stranded DNA was gel-purified prior to use. Following incubation of the recombinant protein in the presence or absence of cold self or nonself-competitor at 4°C for 15 min in binding buffer consisting of 20 mM HEPES (pH 7.6), 20% glycerol, 0.2 mM EDTA, 0.1% Triton X-100, 100 mM NaCl, and 100 μM $ZnSO_4$, the ^{32}P kinase-labeled probe (1 ng, 1.0×10^4 cpm) was added, followed by incubation at 30°C for 15 min. The mixture was then electrophoresed on 6% polyacrylamide gels. The gels were then subjected to autoradiography.

Sequence Analysis—Primary structure alignment and phylogenetic tree analysis (NJ plot) were performed using Clustal W (EMBL) (*15*). Edmundson Wheels hydrophobic plot analysis (*16*) was performed with Genetyx-Mac 8.0 software (Software Development). Homology searches were performed with online BLAST software (GenBank, NCBI) (*17*). Domain analysis was performed with online PROSITE software (GenomeNet WWW server, Human Genome Center, Institute of Medical Science, The University of Tokyo) (*18*). Secondary structure prediction analysis was performed with online Predict-Protein software (PHD, EMBL) (*19*).

RESULTS

Isolation of the Putative Clone Encoding the Factor Binding to the GC-Rich Binding Sites by Yeast One-Hybrid Screening—On yeast one-hybrid genetic screening (scheme shown in Fig. 1b), we isolated a cDNA encoding a factor which interacts with the HIV-1 GC-rich binding sites from a peripheral blood leukocyte library. Of the 7.0×10^6 clones screened, 30 met criteria for a dual reporter system (*e.g.* HIS/3AT and βgal). Two independent clones encoded the novel protein. The Sp1-encoding clone was not isolated. Initial analysis of the nucleotide sequence showed this factor contained a C_2H_2-type zinc finger domain, as assessed by PROSITE motif analysis. This factor was named GBF, GC-rich site binding factor. Its intracellular localization was analyzed using a GFP-fusion construct (see Fig. 2), which showed localization to the nucleus. This is supported by the finding that GBF contains a stretch of basic amino acids N-terminal of the zinc finger, which has been implied to be a nuclear localizing signal in similar zinc finger proteins (*20*). With initial data supporting that we had obtained a novel zinc finger nuclear factor, further studies were performed to isolate the entire coding region and to ascertain its characteristics.

Isolation of the Entire Coding Region of the cDNA—The

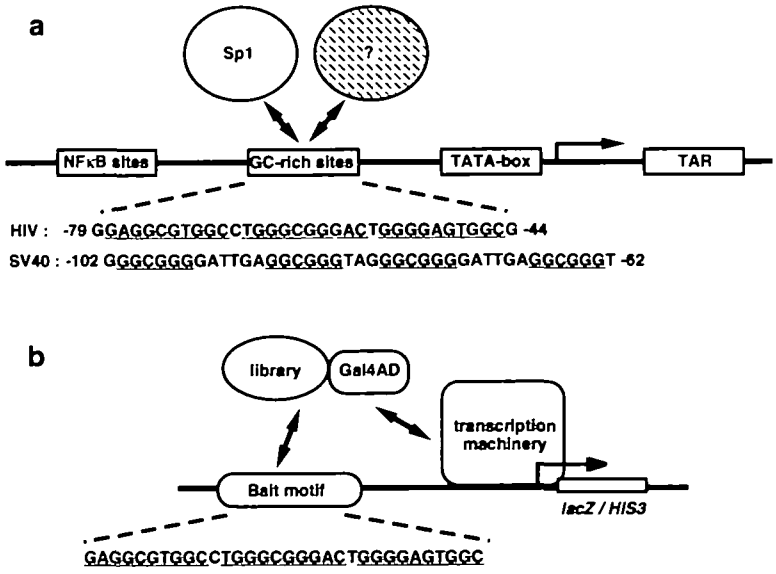

Fig. 1. **(a) Schematic diagram of the DNA binding elements at the HIV-1 core promoter and interactions with DNA-binding factors.** Note the divergence in the GC-rich sequence between the SV40 and HIV-1 promoter regions. The underlined HIV-1 GC-rich sequences are regions protected by Sp1 as seen in DNaseI footprint studies (*7*), and the underlined regions in SV40 are identically repeating GC-rich sequences. Note that only the middle underlined region of the HIV-1 GC-rich sequence is identical to that of SV40. **(b) Schematic diagram of the yeast one-hybrid assay.**

initially obtained clone contained an in-frame stop codon but lacked a putative translation initiation codon. Phage hybridization screening of a peripheral blood leukocyte library was performed to determine the 5′ end. The most distal clone, resulting in extension 130 bp upstream, contained a putative translation initiation site encoding a methionine with a flanking consensus Kozak sequence (*21*), and was denoted as the putative translation initiation codon (see Fig. 3a). Further upstream of the putative translation initiation site (*e.g.* the first methionine residue) we located an in-frame stop codon which additionally supports this methionine is the *bona-fide* translation initiation site. RT-

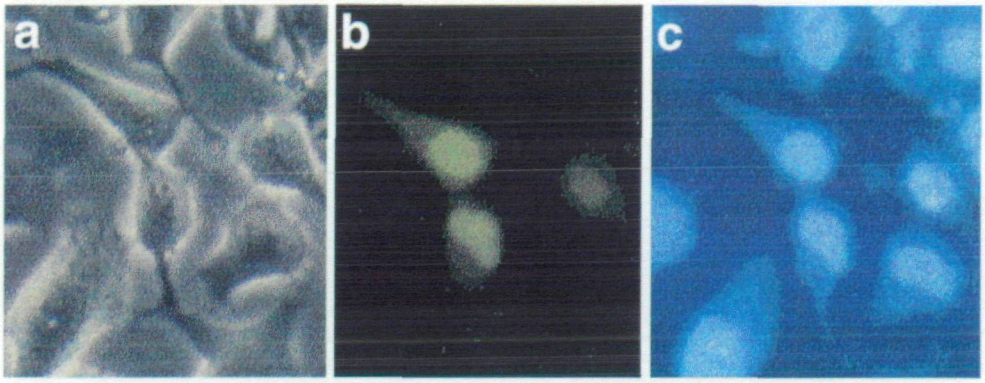

Fig. 2. **Intracellular localization analysis of GBF.** (a) Phase contrast microscopy, (b) green fluorescent protein (GFP) fusion construct, (c) Hoechst nuclear staining.

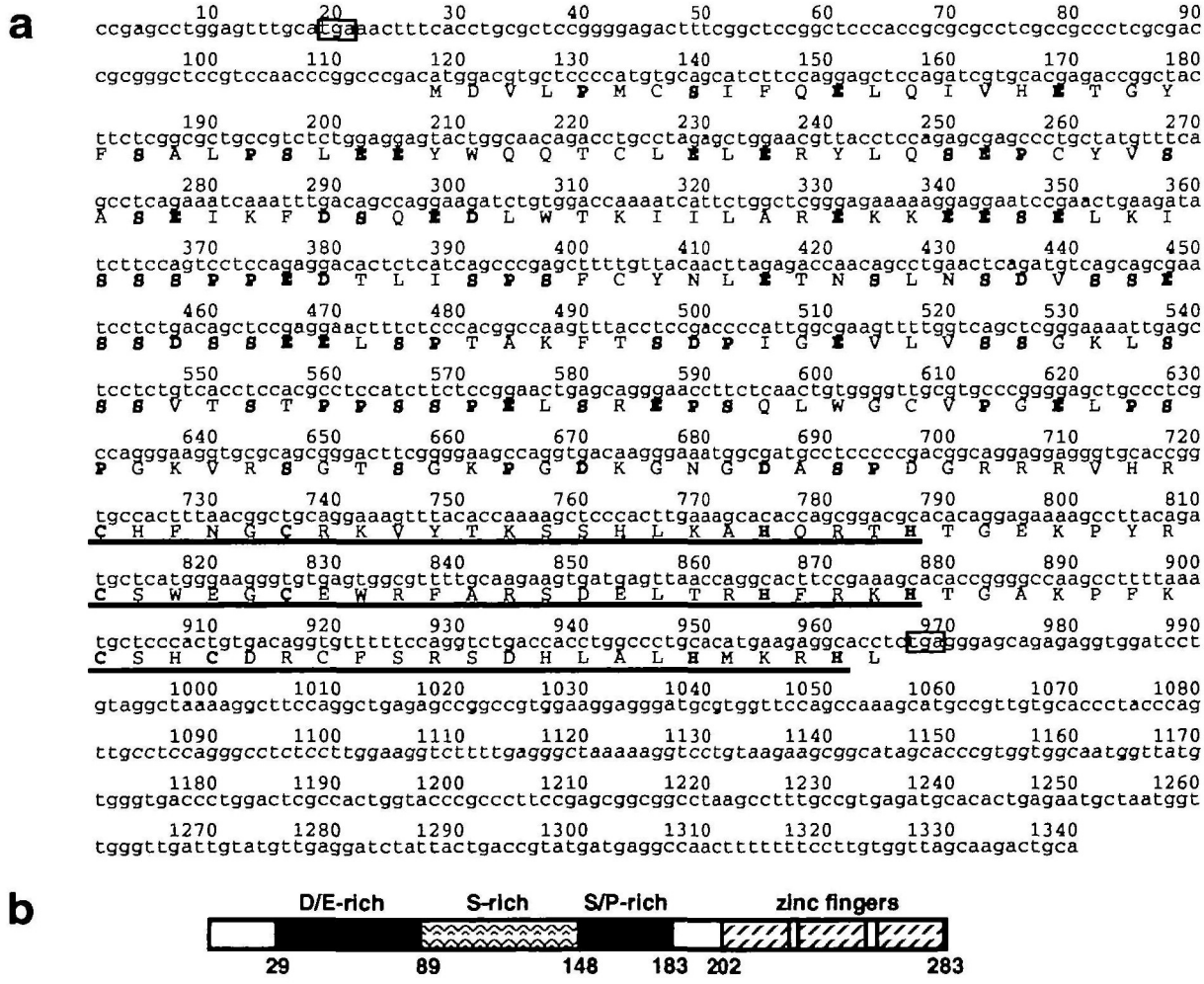

Fig. 3. **Primary structure of GBF.** (a) Nucleotide and amino acid structures of GBF. The deduced amino acid sequences are shown below the nucleotide sequences. Putative zinc finger regions are underlined with the anchoring cysteine and histidine amino acids shown in bold. The in-frame stop codon upstream of the first methionine and the termination codon are boxed. Major amino acids in the non-zinc finger region are colored: D or E, red; S, blue; P, green. (b) Schematic diagram showing the putative domain structures of GBF.

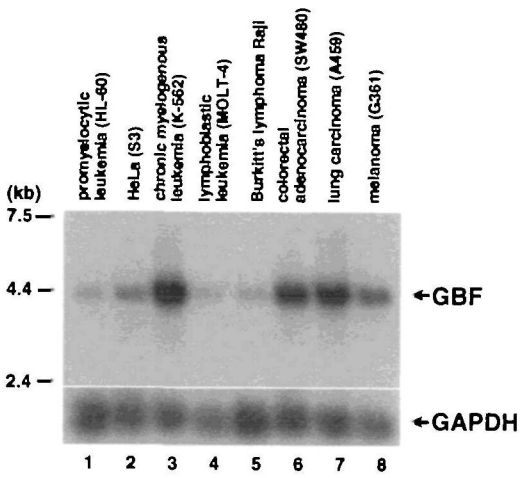

Fig. 4. **Multi-cell Northern blot analysis of GBF mRNA.** A single band at 4.4 kb (GBF) can be seen. GAPDH is shown as an expression control. The investigated cell lines are shown at the top of the figure.

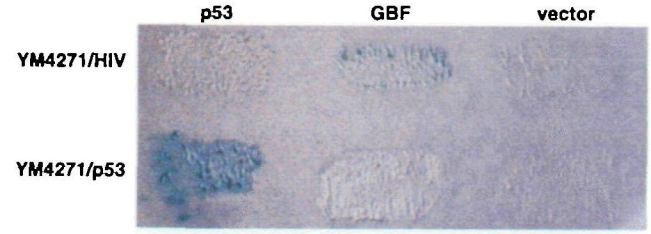

Fig. 5. **Assaying of specific binding of GBF to the GC-rich binding sites of the HIV-1 promoter on co-transfection of the denoted plasmids into the denoted reporter yeast strain shown on the left.** The columns show transfected effectors. Note that only the combination of YM4271/HIV with GBF, and the p53 control show a positive interaction, as shown by blue discoloration on β-gal assaying.

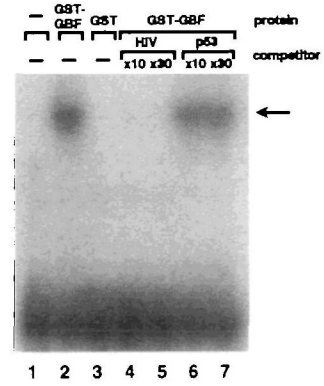

Fig. 6. **Gel shift analysis showing direct sequence-specific DNA-binding of the GST-fusion GBF zinc finger recombinant protein (GST-GBF) and the HIV-1 GC-rich binding sites.** GST-GBF was competed out by increasing amounts of cold probe oligos but not an equimolar amount of a non-specific cold competitor (*e.g.* p53 consensus binding site). The arrow indicates the shifted DNA-protein complex.

PCR involving HeLa mRNA was also performed to independently confirm the sequence of the 5′ untranslated region (data not shown).

Characteristics of the Primary Structure of GBF—Blast analysis of GBF showed the protein is similar to known Krüppel-type DNA-binding zinc finger factors (*e.g.* BTEB2 and EKLF). Primary structure analysis of GBF revealed a 60 amino acid stretch at the N-terminus (aa 29-88) rich in acidic amino acids (D/E, 25%). Secondary structure prediction with Predict-Protein and hydrophobicity plotting by Edmundson Wheels analysis further revealed this region is predominantly helical with an amphipathic distribution (data not shown). Given that the amphipathic helix/acidic stretch has been implicated as a transactivation domain (*22*), these findings strongly suggest this region is the putative activation domain. Furthermore, a region rich in serine residues (31%) (aa 89-148), and a serine- and proline-rich region (S, 19%; P, 22%) (aa 149-183) were found in the mid-region of the protein. The proline-rich domain may also be a potential activation or repressor domain. The serine-rich region may be a potential site for additional regulatory modification (*e.g.* phosphorylation) (see Fig. 3b for a schematic diagram of the GBF primary structure).

Expression Profile of GBF—Analysis of expression by Northern blotting using a multiple cell line blot showed that GBF is ubiquitously expressed, and importantly that it is expressed in T-cells, which are susceptible to infection by HIV (*e.g.* MOLT-4) (see Fig. 4). A single transcript of approximately 4.4 kb, suggesting the transcript consists of approximately 3.5 kb non-coding and 0.9 kb coding regions, was obtained. Studies involving multiple tissue Northern blotting also showed GBF is expressed ubiquitously, with marginally differing transcript levels among tissue types (data not shown).

Specific Binding of GBF to the HIV GC-Rich Binding Sites—Initial yeast selection and co-transfection analysis showed the interaction between GBF and the HIV-1 GC-rich binding sites is specific with p53 as a control (see Fig. 5). Additionally, BLAST analysis had shown GBF is similar to known DNA-binding zinc finger factors. Studies were further conducted to show GBF is a direct DNA-binding factor. Gel shift analysis was performed to confirm the DNA binding and sequence-specific binding properties. The GST-fusion GBF zinc finger recombinant protein bound to the HIV-1 GC-rich sites, which was not competed for by the p53 consensus binding sequence in an equimolar amount (see Fig. 6). These findings support the sequence-specific DNA-binding properties of GBF to the HIV-1 GC-rich binding sites.

DISCUSSION

Elucidation of the nature of DNA-protein interactions is an important initial step for understanding the mechanisms underlying regulation of gene expression by transcriptional processes, as this is a decisive step in the regulation of transcription (*23-25*). To better understand the transcriptional mechanisms governing HIV transcription, we initially focused on the interaction between DNA binding factors and their binding elements, notably on the interactions at the GC-rich binding sites of the HIV-1 core promoter. In contrast to previous studies, which had shown that zinc finger factor Sp1 interacts with the GC-rich

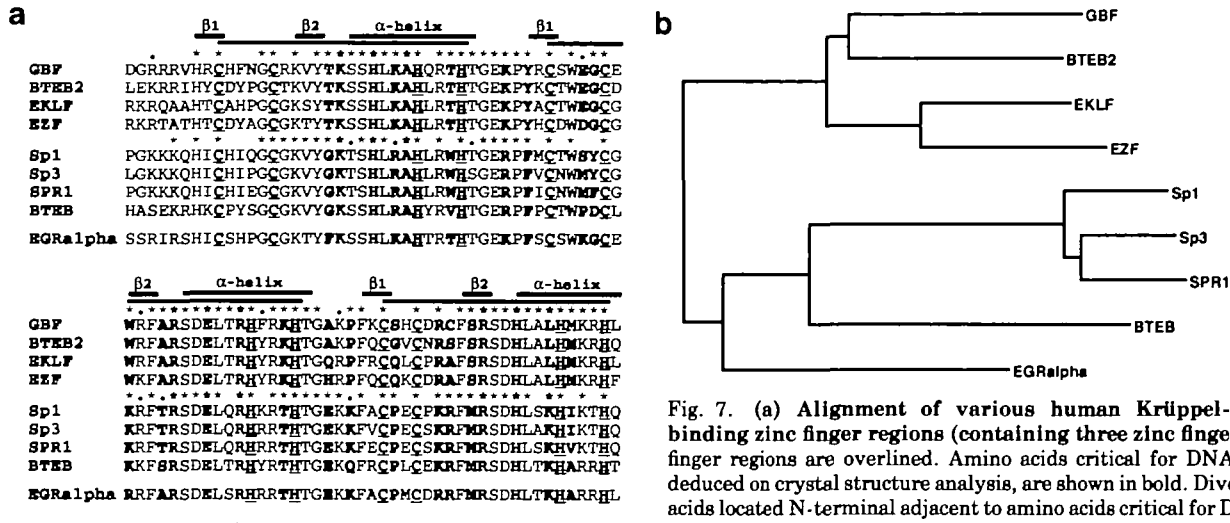

Fig. 7. **(a) Alignment of various human Krüppel-type DNA-binding zinc finger regions (containing three zinc fingers).** The zinc finger regions are overlined. Amino acids critical for DNA binding, as deduced on crystal structure analysis, are shown in bold. Divergent amino acids located N-terminal adjacent to amino acids critical for DNA binding, and the third amino acid of the third zinc finger critical for DNA binding are shown in blue. These amino acids share common positions relative to the α-helix. All other amino acids showing similarity in groups to the two upper groups are shown in green. Cysteine and histidine residues contacting the zinc are in bold and underlined. * denotes identity, · denotes similarity. β1, β2, and α-helix represent the putative first and second β sheets, and the α-helix, respectively, as deduced on past crystal structure analyses of similar zinc finger proteins (*e.g.* Zif-268). The reference accession numbers are as follows: BTEB2 (D14520), EKLF (U65404), EZF (U70663), Sp1(J03133), Sp3 (M97191), SPR1 (X68561), BTEB (D31716), EGRalpha (S81439). **(b) Phylogenetic tree analysis of these zinc finger clones (Clustal W, NJ plot) showing distinct subfamilies exist among these factors.**

binding sites of the HIV-1 LTR, as previously mentioned we had reasoned that other factors may interact with this site under physiologic conditions given the divergence of the DNA sequence from the consensus binding sequence of Sp1, suggesting possible binding of other sequence-specific DNA-binding factors and/or functionally divergent members of the zinc finger factor family. More importantly, during the course of the present study genetic experiments in mice (*e.g.* Sp1 knockout mice) showed that numerous genes thought to be regulated by Sp1 based on *in vitro* functional data were not regulated by Sp1 *in vivo*, thus further supporting our initial hypothesis that the criteria in past studies on the relationship between DNA binding and functional activity do not accurately reflect the actual qualitative and quantitative conditions in physiologic processes.

Indeed, the validation of a *bona-fide* molecular interaction is a difficult task. One can only improve the scientific criteria on which the evidence is based. Biochemical studies can only attempt to mimic physiological conditions. Established experimental assays may not reflect actual physiological conditions. Yeast one-hybrid screening was employed as a model system most likely reflecting the eukaryotic transcriptional processes under physiological conditions, as the intracellular environment is maintained and compartmentation is controlled. Stoichiometry (*e.g.* overexpression) and conservation of the qualitative nature of the transcriptional environment (*e.g.* cofactors and basal machinery), however, are problems which remain. We reasoned that fundamental viral transcriptional processes, given their simplicity, are likely conserved within this easily manipulated eukaryotic organism. Genomic integration of the reporters was also performed to better reflect physiological conditions. Initial studies to confirm DNA binding were performed in the present study. Further studies are necessary to validate our claims (*e.g.* gene knockout studies, *etc.*).

A cDNA encoding a novel zinc finger protein, GBF, containing three zinc fingers at its C-terminal end, and putative N-terminal acidic and proline-rich domains was isolated through our yeast genetic screening for factors interacting with the GC-rich binding sites of the HIV-1 core promoter. Yeast co-transfection assays and gel-shift analysis further showed that GBF directly binds to the HIV-1 GC-rich binding sites in a sequence-specific manner. The DNA-binding characteristics of zinc finger factors have been extensively studied. Crystal structure analyses of DNA-binding zinc finger transcription factors have allowed the prediction of the amino acids critical for DNA binding from the primary structure (*26, 27*) (see Fig. 7a). These amino acids in contact with DNA reside in the α-helical region of the zinc finger. As these critical amino acids are highly conserved in GBF and other Krüppel-type zinc finger transcription factors, it is tempting to assume that these Krüppel-type zinc finger factors likely share similar DNA binding properties. Closer examination of this zinc finger region, however, shows discrete yet distinct differences. Notable differences are seen in the third amino acid critical for DNA binding of the third zinc finger, and in the amino acids N-terminal adjacent to the first amino acid critical for DNA binding and the third amino acid critical for DNA binding in each of the zinc fingers (see Fig. 7a). Based on these findings, it is tempting to speculate that they may have a stochastic effect on DNA binding properties. Numerous other differences are evident, as shown in the figure (see Fig. 7a). Aside from DNA-binding properties, these differences in primary structure and quite possibly in the overall conformation of the folded protein may have a profound effect on protein-protein interactions, as the zinc finger is a known interface for protein-protein interactions including homo- and hetero-dimer formation.

Overall, based on the results of our analysis of the primary structure, it is likely that distinct subgroups exist within the Krüppel-type zinc finger types. Distinguishing

characteristics in the zinc-finger region may have a possible effect on DNA-binding properties as well as on additional protein-protein interactions. Given that it has been estimated that nearly 1% of the human genome encodes zinc finger proteins (*28*), it is likely that multiple regulatory and interacting domains are necessary for individual functional diversity.

In closing, we have isolated a novel factor which interacts with the GC-rich binding sites of the HIV-1 LTR by means of methods for specific isolation of factors which interact with the native sites. Our findings suggest that our present knowledge of the mechanisms underlying gene regulation of HIV-1 may still be quite limited, and also suggest the need to further understanding of the mechanisms underlying HIV-1 transcriptional regulation. Additionally, the understanding of functional discrimination of zinc finger factors at similar GC-rich DNA binding sites will also provide a viable model for elucidation of the physiological dissection of transcriptional regulatory mechanisms.

Note Added in Proof: During the course of the present work, the authors learned that Koritschoner *et al.* (*J. Biol. Chem.* 1997; **272**, 9573) had isolated a partial cDNA identical in part with GBF in their study to isolate DNA-binding transcription factors interacting with the core element of the promoter of a pregnancy-specific glycoprotein gene.

The authors wish to thank S. Elledge for the human peripheral blood leukocyte library. The authors are also grateful to S. Ijuin and M. Susa for the helpful discussion regarding crystal structure analysis, T. Kuzuhara for the helpful discussions and technical assistance, and Y. Ogata and Y. Araki for the helpful discussions.

REFERENCES

1. Gallo, R.C., Salahuddin, S.Z., Popovic, M., Shearer, G.M., Kaplan, M., Haynes, B.F., Palker, T.J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P.D. (1984) Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500-503
2. Verdin, E., Paras, P. Jr., and Van Lint, C. (1993) Chromatin disruption in the promoter of human immunodeficiency virus type 1 during transcriptional activation. *EMBO J.* **12**, 3249-3259
3. Jones, K.A. and Peterlin, B.M. (1994) Control of RNA initiation and elongation at the HIV-1 promoter. *Annu. Rev. Biochem.* **63**, 717-743
4. Dynan, W.S. and Tjian, R. (1983) The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**, 79-87
5. Courey, A.J. and Tjian, R. (1988) Analysis of Sp1 *in vivo* reveals multiple transcription domains, including a novel glutamine-rich activation motif. *Cell* **55**, 887-898
6. Kadonaga, J.T., Courey, A.J., Ladika, J., and Tjian, R. (1988) Distinct regions of Sp1 modulate DNA binding and transcriptional activation. *Science* **242**, 1566-1570
7. Jones, K.A., Kadonaga, J.T., Luciw, P.A., and Tjian, R. (1986) Activation of the AIDS retrovirus promoter by the cellular transcription factor, Sp1. *Science* **232**, 755-759
8. Imataka, H., Sogawa, K., Yasumoto, K., Kikuchi, Y., Sasano, K., Kobayashi, A., Hayami, M., and Fujii-Kuriyama, Y. (1992) Two regulatory proteins that bind to the basic transcription element (BTE), a GC box sequence in the promoter region of the rat P-4501A1 gene. *EMBO J.* **11**, 3663-3671
9. Sogawa, K., Imataka, H., Yamasaki, Y., Kusume, H., Abe, H., and Fujii-Kuriyama, Y. (1993) cDNA cloning and transcription properties of a novel GC box-binding protein, BTEB2. *Nucleic Acids Res.* **21**, 1527-1532
10. Sogawa, K., Kikuchi, Y., Imataka, H., and Fujii-Kuriyama, Y. (1993) Comparison of DNA-binding properties between BTEB and Sp1. *J. Biochem.* **114**, 605-609
11. Majello, B., DeLuca, P., Hagen, D., Suske, G., and Lania, L. (1994) Different members of the Sp1 multigene family exert opposite transcriptional regulation of the long terminal repeat of HIV-1. *Nucleic Acids Res.* **22**, 4914-4921
12. Marin, M., Karis, A., Visser, P., Grosveld, F., and Philipsen, S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell* **89**, 619-628
13. Ito, H., Fukuda, Y., Murata, K., and Kimura, A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* **153**, 163-168
14. Choi, B.I., Bando, M., Hasegawa, S., and Horikoshi, M. (1996) Isolation and characterization of a cDNA encoding a novel human transcription factor TFIID subunit containing similarities with histones H2B and H3. *Gene* **169**, 263-267
15. Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680
16. Schiffer, M. and Edmundson, A.B. (1967) Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* **7**, 121-135
17. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
18. Bairoch, A., Bucher, P., and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**, 217-221
19. Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584-599
20. Shields, J.M. and Yang, V.W. (1997) Two potent nuclear localization signals in the gut-enriched Krüppel-like factor define a subfamily of closely related Krüppel proteins. *J. Biol. Chem.* **272**, 18504-18507
21. Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res.* **12**, 857-872
22. Giniger, E. and Ptashne, M. (1987) Transcription in yeast activated by a putative amphipathic alpha helix linked to a DNA binding unit. *Nature* **330**, 670-672
23. Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318-356
24. Ptashne, M. (1967) Isolation of the λ phage repressor. *Proc. Natl. Acad. Sci. USA* **57**, 306-313
25. Gilbert, W. and Muller-Hill, B. (1967) The *lac* operator is DNA. *Proc. Natl. Acad. Sci. USA* **58**, 2415-2421
26. Klevit, R.E. (1991) Recognition of DNA by Cys2, His2 zinc fingers. *Science* **253**, 1367
27. Suzuki, M., Gerstein, M., and Yagi, N. (1994) Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res.* **22**, 3397-3405
28. Klug, A. and Schwabe, J.W. (1995) Protein motifs 5. Zinc fingers. *FASEB J.* **9**, 597-604