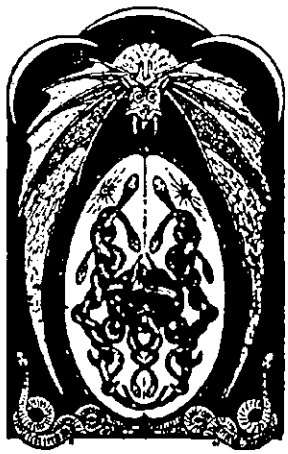
**H, Scheer M, Thiele S, and Wingender E.** TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31: 374–378, 2003.

12. **Rhodes DR, Barrette TR, Rubin MA, Ghosh D, and Chinnaiyan AM.** Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62: 4427–4433, 2002.
13. **Saban MR, Hellmich H, Nguyen NB, Winston J, Hammond TG, and Saban R.** Time course of LPS-induced gene expression in a mouse model of genitourinary inflammation. *Physiol Genomics* 5: 147–160, 2001.
14. **Seo J and Shneiderman B.** Interactively exploring hierarchical clustering results. *IEEE Computer* 35: 80–86, 2002.
15. **Shiffman D, Mikita T, Tai JT, Wade DP, Porter JG, Seilhamer JJ, Somogyi R, Liang S, and Lawn RM.** Large scale gene expression analysis of cholesterol-loaded macrophages. *J Biol Chem* 275: 37324–37332, 2000.
16. **Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, and Golub TR.** Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 96: 2907–2912, 1999.
17. **Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, and Church GM.** Systematic determination of genetic network architecture. *Nat Genet* 22: 281–285, 1999.
18. **Wang GL, Jiang BH, Rue EA, and Semenza GL.** Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular $O_2$ tension. *Proc Natl Acad Sci USA* 92: 5510–5514, 1995.
19. **Yan SF, Lu J, Zou YS, Soh-Won J, Cohen DM, Buttrick PM, Cooper DR, Steinberg SF, Mackman N, Pinsky DJ, and Stern DM.** Hypoxia-associated induction of early growth response-1 gene expression. *J Biol Chem* 274: 15030–15040, 1999.





Physiological Genomics

Downloaded from physiolgenomics.physiology.org on May 11, 2005

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 34, Issue of August 20, pp. 35377–35383, 2004
Printed in U.S.A.
© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

# Cruciform DNA Structure Underlies the Etiology for Palindrome-mediated Human Chromosomal Translocations*

Received for publication, January 13, 2004, and in revised form, June 10, 2004
Published, JBC Papers in Press, June 20, 2004, DOI 10.1074/jbc.M400354200

**Hiroki Kurahashi‡§¶, Hidehito Inagaki‡, Kouji Yamada‡, Tamae Ohye‡, Mariko Taniguchi§, Beverly S. Emanuel‖**, and Tatsushi Toda§**

*From the ‡Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi 470-1192, Japan, the §Division of Functional Genomics, Department of Post-Genomics and Diseases, Osaka University Graduate School of Medicine, Osaka 565-0871, Japan, the ‖Division of Human Genetics and Molecular Biology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104, and the **Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104*

**There is accumulating evidence to suggest that palindromic AT-rich repeats (PATRRs) represent hot spots of double-strand breakage that lead to recurrent chromosomal translocations in humans. As a mechanism for such rearrangements, we proposed that the PATRR forms a cruciform structure that is the source of genomic instability. To test this hypothesis, we have investigated the tertiary structure of a cloned PATRR. We have observed that a plasmid containing this PATRR undergoesaconformationalchange,causingtemperature-dependent mobility changes upon agarose gel electrophoresis. The mobility shift is observed in physiologic salt concentrations and is most prominent when the plasmid DNA is incubated at room temperature prior to electrophoresis. Analysis using two-dimensional gel electrophoresis indicates that the mobility shift results from the formation of a cruciform structure. S1 nuclease and T7 endonuclease both cut the plasmid into a linear form, also suggesting cruciform formation. Furthermore, anti-cruciform DNA antibody reduces the electrophoretic mobility of the PATRR-containing fragment. Finally, we have directly visualized cruciform extrusions from the plasmid DNA with the size expected of hairpin arms using atomic force microscopy. Our data imply that for human chromosomes, translocation susceptibility is mediated by PATRRs and likely results from their unstable conformation.**

The constitutional t(11;22)(q23;q11) is the only known recurrent non-Robertsonian translocation in humans. Its recurrent nature implicates a specific genomic structure at the t(11;22) breakpoints. Analyses of numerous independent t(11;22) cases have localized the breakpoints within palindromic AT-rich repeats (PATRRs)[1] on 11q23 and 22q11 (1–4). Most 11;22 translocations show breakpoints at the center of the PATRRs, suggesting that the center of the palindrome is susceptible to double-strand breaks, leading to the translocation (5, 6). Indeed, translocation-specific PCR detects a high frequency of *de novo* t(11;22)s in normal sperm samples (7).

The breakpoints on 22q11 are located within one of the unclonable gaps in the human genome (8, 9). Extensive screening of YAC/BAC/PAC libraries has not been successful in cloning this breakpoint region. However, experimentally derived sequences from numerous t(11;22) junction fragments demonstrate that the 22q11 breakpoints reside within a larger PATRR. The breakpoints of a variety of translocations involving 22q11 cluster within this region, suggesting that the 22q11 PATRR is highly unstable (10–13). More recently, molecular cloning of translocation breakpoints has demonstrated similar palindromic sequences on partner chromosomes, such as 17q11, 4q35.1, and 1p21.2 (14–16).

It has been suggested that for palindromic sequences in double-stranded DNA, the interstrand base pairs might convert to intrastrand pairs, producing a set of hairpin structures described as a cruciform (see Fig. 1*A*). In appropriate conditions, these structures can be generated from palindromic DNA *in vitro* and visualized directly by electron microscopy (17). Under physiological conditions, cruciform extrusion is usually kinetically blocked; it appears to take place only in response to heating, suggesting that a considerable amount of energy is required for cruciform extrusion (18). However, under appropriate conditions, palindromic DNA prefers to extrude cruciform arms even at physiological temperatures (19). In a previous report, we demonstrated that the PATRRs on 11q23, 17q11, and 22q11 are all comprised of a long AT-rich region with relatively GC-rich ends (5, 14). We have proposed that the AT richness of the PATRRs contributes to strand separation at physiological temperatures, whereas the relatively GC-rich ends contribute to the stable intrastrand complementary interaction of the PATRR. These characteristics are likely to induce or favor the formation of a cruciform structure.

In this study, we have investigated the *in vitro* tertiary structure of the 11q23 PATRR and demonstrated that it forms a cruciform structure at physiological conditions. Our data suggest that PATRRs may form unstable structures that lead to chromosomal translocations in humans. These data also implicate a biological role for the temperature-sensitive conformational change characteristic of palindromic DNA.

## EXPERIMENTAL PROCEDURES

*PATRR Plasmid*—A plasmid containing the chromosome 11 PATRR (204 bp) was constructed by PCR and TA cloning as described previously (psPATRR11) (1). The control plasmid that deletes the PATRR region was constructed using BAC 442e11, which deletes almost the

* This work was supported by a grant-in-aid for scientific research from the Ministry of Education, Science, Sports and Culture of Japan (to H. K.) and Grants HD26079 and CA39926 from the Charles E. H. Upham chair in Pediatrics (to B. S. E.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed: Division of Molecular Genetics, Institute for Comprehensive Medical Science, Fujita Health University, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, Aichi 470-1192, Japan. Tel.: 81-562-93-9391; Fax: 81-562-93-8831; E-mail: kura@fujita-hu.ac.jp.

[1] The abbreviations used are: PATRR, palindromic AT-rich repeat; TBE, Tris/borate/EDTA buffer; EMSA, electrophoretic mobility shift assay; AFM, atomic force microscopy.

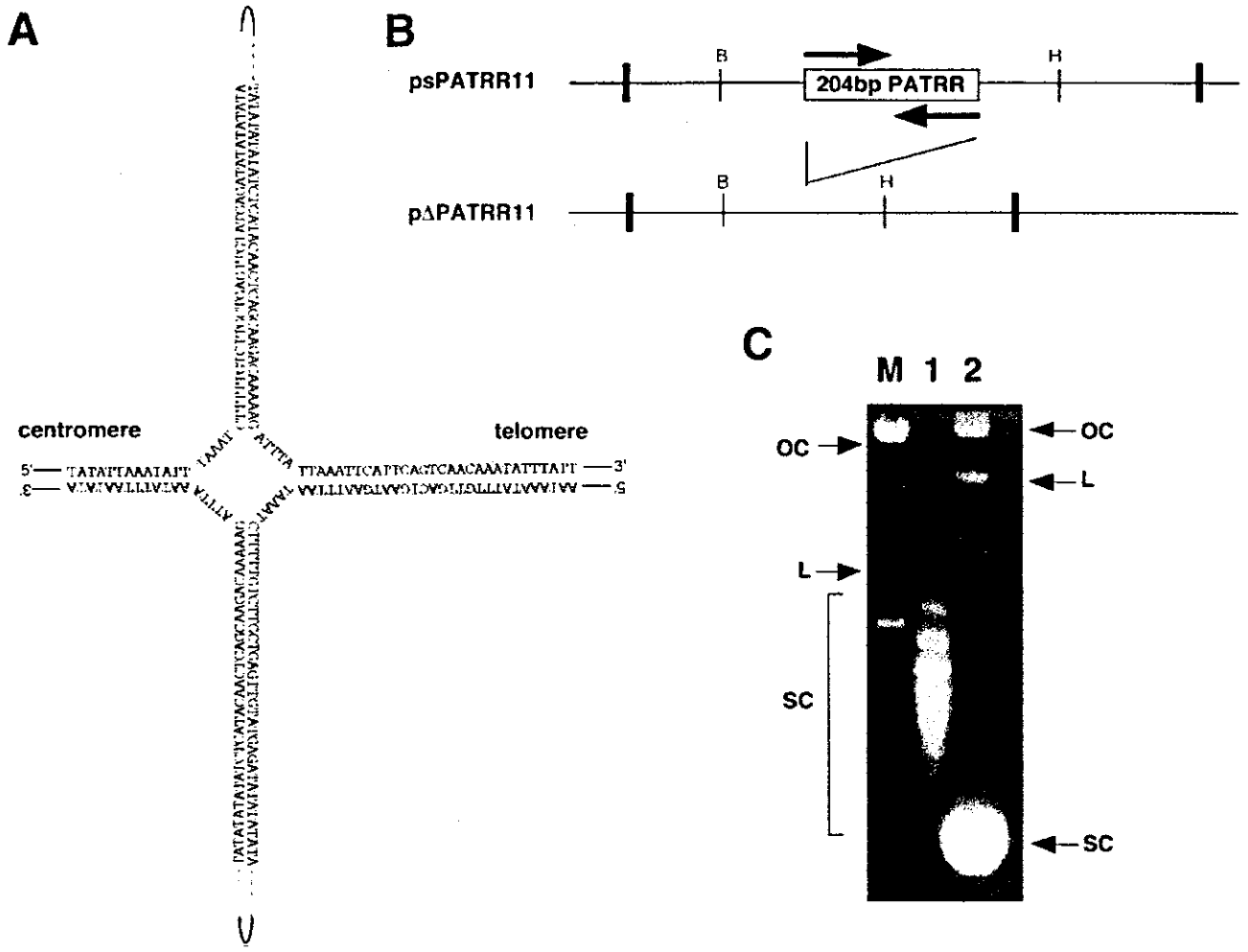



FIG. 1. **Agarose gel electrophoresis of the PATRR plasmid.** *A*, putative cruciform structure of the PATRR of chromosome 11. *B*, schematic representation of psPATRR11 and pΔPATRR11. The *box* indicates the inverted repeat region, whereas each *arrow* indicates the repeat unit. pΔPATRR11 deletes almost the entire 204-bp PATRR region. *Bold vertical lines* indicate the cloning sites of the plasmids. The *thin vertical lines* indicate restriction sites for the following enzymes: BamHI (*B*) and HincII (*H*). *C*, agarose gel electrophoresis of psPATRR11 and pΔPATRR11. *Lane M*, molecular size marker; *lane 1*, psPATRR11; *lane 2*, pΔPATRR11. *OC*, open circle; *L*, linear; *SC*, supercoiled circle.

entire PATRR region (pΔPATRR11). Plasmid DNA was isolated by means of alkaline lysis (denaturing) or Triton lysis (non-denaturing) methods and purified using ion-exchange columns (Qiagen) without the use of phenol. Isopropanol-precipitated DNA was dissolved in phosphate-buffered saline/1 mM EDTA. All of the procedures were performed at 4 °C in a cold room to avoid artifactual cruciform extrusion during the procedure. Isolated plasmid DNA was divided into small aliquots and immediately frozen until used in an experiment. The plasmid was quickly thawed, incubated at a given temperature, and then cooled on ice before electrophoresis.

DNA topoisomers containing different linking numbers were prepared by relaxation of the DNA with topoisomerase I (Invitrogen) in the presence of varied concentrations of ethidium bromide as described (20). Several samples with different ranges of linking numbers were pooled to produce a mixture with a broader distribution of topoisomers.

*Two-dimensional Gel Electrophoresis*—Standard agarose gel electrophoresis was performed in horizontal gel slabs immersed in 1× Tris/borate/EDTA (TBE) electrophoresis buffer. The 0.7% agarose gel was prepared in TBE in a square gel tray. Electrophoresis was performed at 45 V for 20 h at 4 °C. The gel was stained in water plus ethidium bromide at 1 μg/ml for 2 h followed by extensive destaining in several changes of water before photography. For two-dimensional gel electrophoresis, the sample was applied at the left upper corner. Electrophoresis in the first dimension was performed using the same conditions as for standard electrophoresis. The gel was then soaked in TBE plus chloroquine (0.75 μg/ml) (Sigma) for 8 h with gentle agitation for equilibration. After soaking, the gel was rotated clockwise 90° from the original position. Electrophoresis in the second dimension was performed in TBE plus chloroquine at 45 V for 20 h. The gel was then soaked in several changes of deionized water to remove the chloroquine. The gel was stained in water plus ethidium bromide in similar conditions as for standard agarose gel electrophoresis. To stabilize the cruciform structure, two-dimensional gel electrophoresis was also performed in $Mg^{2+}$-containing buffer (0.5× TBE, 10 mM $MgCl_2$) for the first dimension (21). Electrophoresis was carried out at 45 V for 30 h at 4 °C, replacing the electrophoresis buffer every 6 h (22).

*Nuclease Sensitivity Assay*—T7 endonuclease (New England BioLabs), S1 nuclease (Takara), and nuclease P1 (Roche Applied Science) digestions were performed at room temperature for 1 h in the appropriate buffers. 10 units of T7 endonuclease, 1 unit of S1 nuclease, and 0.1 units of nuclease P1 were used for digestion of 1 μg of plasmid. After digestion, the plasmid DNA was precipitated with ethanol and then subjected to 0.7% agarose gel electrophoresis.

For mapping of the cleavage sites, the nuclease-digested plasmids were purified, digested with EcoRI or BstXI, and then subjected to 2% agarose gel electrophoresis. The fragments were purified and sequenced using internal primers.

*Electrophoretic Mobility Shift Assay (EMSA)*—EMSA was performed as described previously (23), using a monoclonal antibody against cruciform DNA, 2D3 (24). To shorten the DNA fragment, psPATRR11 and pΔPATRR11 were digested with BamHI and HincII and subcloned into pT7Blue. 100 ng of supercoiled plasmid DNA was incubated at 4 °C for 1 h in 20 μl of binding buffer (10 mM Hepes, pH 7.8, 100 mM NaCl, 1 mM EDTA, and 5 mM $MgCl_2$) with or without 2D3. A second circular plasmid was added simultaneously into the reaction for nonspecific competition. After the addition of 2 μl of 10× restriction digestion buffer, the plasmid was digested for 30 min at room temperature with 5 units of both BamHI and HindIII. Subsequently, 1 unit of Klenow fragment and

10 mCi of [$\alpha$-$^{32}$P]dCTP were added to the reaction mixture, and incubation continued for 10 min. The fragments were resolved on 4% nondenaturing polyacrylamide gels at 4 °C at 5 V/cm. Gels were dried and exposed for an appropriate time.

*Atomic Force Microscopy (AFM) Experiments*—The plasmid DNA sample (10 $\mu$l) was dripped onto a mica substrate coated with spermine, rinsed with 1 ml of water, and blown with dry nitrogen gas. AFM analysis was carried out using a Nanoscope IIIa Multimode system (Digital Instruments Inc., Santa Barbara, CA) in the tapping mode at room temperature in air. Silicon cantilevers of 127 $\mu$m in length with a spring constant of 48 newtons/m were purchased from Nanosensors (Wetzlar-Blankenfeld). Typical resonant frequencies of these tips were 340 kHz. The 512 × 512 pixel images were collected at a rate of two scan lines/second. The contour lengths of the DNA were determined using NIH Image software.

To lock the PATRR plasmid into the cruciform structure, DNA cross-linking was performed by a method similar to that described previously (25). In brief, plasmid DNA was dissolved in a solution of 4,5',8-trimethylpsoralen (100 $\mu$g/ml) and exposed to UV light at 365 nm for 5 min. The DNA was digested with EcoRI, purified, and then subjected to AFM.

## RESULTS

*PATRR Plasmid Prefers to Be in Its Relaxed State*—We have hypothesized that the PATRR forms a cruciform structure even in physiological conditions, leading to genomic instability in humans. To test this hypothesis, we have examined the tertiary structure of a cloned 11q23 PATRR. This PATRR is too unstable to be cloned into a plasmid vector in its entirety. The 11q23 BAC including the t(11;22) breakpoint region inevitably deletes the PATRR during bacterial culture. YACs that span this region also delete the PATRR, suggesting that it is unstable even when yeast is used as the host. Previously, we identified a short polymorphic version of the PATRR (Fig. 1*B*) that can be stably expanded as a plasmid in *Escherichia coli* (5). We prepared the short PATRR-containing plasmid (psPATRR11), as well as a plasmid that deletes the PATRR (pΔPATRR11), in a buffer at physiological salt concentration (phosphate-buffered saline/EDTA) and used them for the following studies.

First, we analyzed the plasmids prepared by the standard alkaline lysis method. Upon standard agarose gel electrophoresis, a plasmid usually yields three bands, which correspond to supercoiled, linear, and open circular DNA. In addition to the expected bands, the psPATRR11 showed a ladder migrating at the position for supercoiled DNA. The pΔPATRR11 did not yield such a ladder (Fig. 1*C*). Each band of the ladder corresponds to a topoisomer containing different linking numbers. Relaxed topoisomers move slowly in the gel. The observed ladder suggests that the PATRR-containing plasmid, unlike the plasmid without the PATRR, prefers to be in its relaxed state in solution.

Cruciform formation involves unpairing of the complementary strands and thus reduces the interstrand twist number. For negatively supercoiled DNA, a reduction in twist number reduces the degree of supercoiling and thus would cause a reduction in electrophoretic mobility. The observed ladder is consistent with cruciform formation of the PATRR plasmid.

*Nuclease Sensitivity Assay*—Cruciform DNA is known to be digested by S1 nuclease, nuclease P1, or T7 endonuclease (26). To confirm cruciform extrusion of the psPATRR11 and exclude the possibility of other conformational variants, we examined the nuclease sensitivity of the PATRR plasmid prepared by the alkaline lysis method. S1 nuclease, which is known to cut single-stranded DNA at the tip of the cruciform structure, cleaved most of the supercoiled psPATRR11 into a linear form (Fig. 2, *lane 2*) but cleaved only a small amount of the pΔPATRR11 or TA cloning vector (Fig. 2, *lanes 6* and *10*). Since the pH for the S1 nuclease reaction buffer is so low, it may nick the plasmid and affect the results. Thus, we also incubated the plasmid with buffer, which resulted in no background linearization (data not shown). Nuclease P1, which can react at physiological pH, produced similar results (data not shown). To exclude the possibility that S1 nuclease cleaves single-stranded DNA of the psPATRR11 only because of the AT-rich nature of the palindrome, we used T7 endonuclease, which can cut the four-way junction of the cruciform structure. When incubated at room temperature, a considerable amount of the psPATRR11 was digested by T7 endonuclease into a linear form (Fig. 2, *lane 3*). In contrast, the endonuclease did not cut either pΔPATRR11 or the TA cloning vector (Fig. 2, *lanes 7* and *11*).

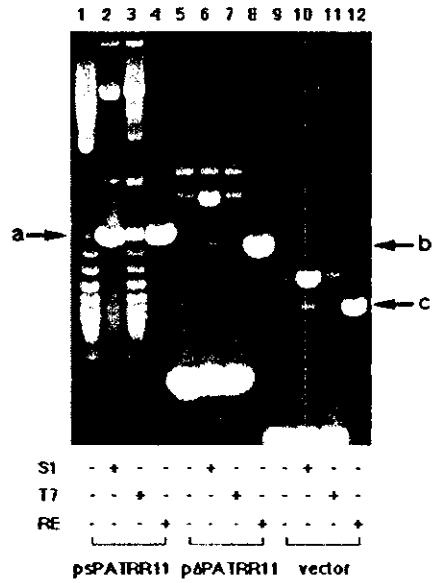



FIG. 2. **Nuclease sensitivity assay of the PATRR plasmid.** The original plasmid, plasmid with S1 nuclease digestion, plasmid with T7 endonuclease digestion, and plasmid with restriction enzyme digestion that cuts the plasmid only once were subjected to agarose gel electrophoresis in this order. *Lanes 1–4*, psPATRR11; *lanes 5–8*, pΔPATRR11; *lanes 9–12*, TA cloning vector. The *arrows* indicate the position of the linear form of psPATRR11 (*a*), pΔPATRR11 (*b*), and TA cloning vector (*c*). *S1*, S1 nuclease; *T7*, T7 endonuclease; *RE*, restriction enzyme.

*Two-dimensional Gel Electrophoresis*—To confirm cruciform extrusion of the psPATRR11 further, we performed two-dimensional gel electrophoresis. The first dimension, run in the absence of intercalative ligands, resolves the topoisomers on the basis of torsion in the backbone. Since structural transitions due to unwinding torsion are reversed in the presence of chloroquine, a second dimension run in the presence of this ligand resolves the topoisomers on the basis of linking differences.

Examination of psPATRR11 prepared by the standard alkaline lysis method reveals two non-continuous downward-sloping curves on the gel, suggestive of an abrupt structural transition (Fig. 3*A*). Next, we treated psPATRR11 with topoisomerase I in the presence of various amounts of ethidium bromide so as to prepare plasmids with broader topoisomer distribution. After topoisomerase I treatment, psPATRR11 yields an additional curve above the original two (Fig. 3*B*). This additional curve corresponds to more relaxed topoisomers generated by topoisomerase I treatment. In the first dimension, the electrophoretic mobilities of cruciform-containing topoisomers are retarded in comparison with the same topoisomers without the cruciform, whereas these two forms migrate identically in the second dimension. This is the result of the elimination of cruciform structures in the presence of chloroquine, which acts to reduce torsional stress. Since a greater degree of negative supercoiling is known to accelerate cruciform extrusion (17, 18), the two minor curves on this gel are likely to originate from cruciform DNA.

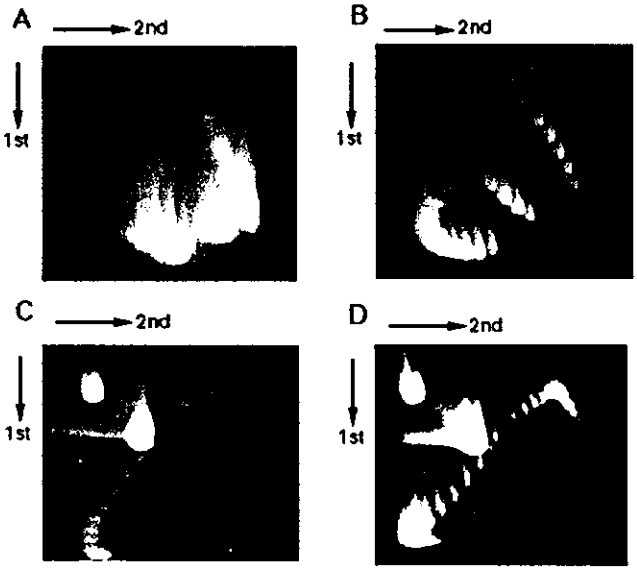



FIG. 3. **Two-dimensional gel electrophoresis of the PATRR plasmid.** *A*, psPATRR11 prepared by the alkaline lysis method. *B*, psPATRR11 treated with topoisomerase I in the presence of various amounts of ethidium bromide. *C*, topoisomerase I-treated psPATRR11 digested with T7 endonuclease. *D*, topoisomerase I-treated psPATRR11 digested with S1 nuclease. Two downward-sloping curves originating from cruciform extrusion are observed at the *lower right side* on both the *A* and the *B* gels, but neither are observed in the *C* nor in the *D* gel. *Spots* at the *upper left side* on the gels originate from open circular nicked plasmids, whereas *spots* near the *center* of the gels are from nuclease-cleaved linear plasmids.

We digested topoisomerase I-treated psPATRR11 with S1 nuclease or T7 endonuclease prior to the two-dimensional gel electrophoresis. As expected, the two minor curves virtually disappeared as a result of nuclease digestion, whereas the other curve remained unchanged (Fig. 3, *C* and *D*). Large spots indicative of open circles and linear DNA appeared instead. These results demonstrate that two minor arcs originate from cruciform DNA and that, combined with the data derived from undigested psPATRR11, the majority of psPATRR11, when isolated by the alkaline lysis method, forms a cruciform extrusion.

$Mg^{2+}$ is known to stabilize a cruciform structure (21, 22). We repeated the two-dimensional gel electrophoresis experiments in $Mg^{2+}$-containing buffer for the first dimension. The results were similar to those seen for standard TBE buffer (data not shown).

*Mapping of Nuclease Cleavage Sites in the PATRR*—The presence of two curves on the two-dimensional gel indicates that the psPATRR11 adopts two types of cruciform conformations. To localize the subregion of the PATRR responsible for cruciform extrusion, we mapped nuclease cleavage sites using restriction enzyme digestion (Fig. 4, *A* and *B*). Both T7 endonuclease and S1 nuclease yielded several fragments that were shorter than that of the PATRR-containing fragment, whereas the intensity of the original fragment was reduced by treatment with the nucleases. The pattern of cleavage provides additional support toward confirming cruciform extrusion of the PATRR region of the plasmid.

To further map the cleavage sites, we performed sequence analysis of each nuclease-cleaved fragment. Four EcoRI-T7 endonuclease fragments were sequenced, and we localized the cleavage sites at two positions (Fig. 4*C*). Thus, we conclude that psPATRR11 forms large and small cruciform extrusions as was predicted from the two-dimensional gel analysis. The small cruciform extrusion originates from the 36-bp region at the center of the PATRR. Further, the PATRR has a relatively GC-rich region at both ends (1, 5). Thus, the large cruciform originates from the region that excludes this GC-rich region. The cleavage sites of S1 nuclease were rather complicated. One of the cleavage sites colocalized with the precise center of the PATRR, the putative tip of the cruciform structure (Fig. 4*C*). The rest of the cleavage sites of S1 nuclease were mapped at the mismatch regions between the proximal and distal arms of the PATRR. These results suggest intrastrand annealing within the PATRR, which is also consistent with cruciform formation.

*Temperature-sensitive Structural Transition of the PATRR*—To determine whether the PATRR plasmid adopts cruciform conformation in *E. coli*, we have prepared the plasmid by a Triton lysis method (non-denaturing method) since denaturation of the PATRR can induce artifactual cruciform formation. Unlike plasmids isolated by the alkaline lysis method, psPATRR11 did not yield a characteristic ladder in standard agarose gel electrophoresis (Fig. 5*A*, *lane 1*). This suggests that the psPATRR11 does not form a cruciform in *E. coli*.

Next, we incubated the psPATRR11 at various temperatures for 1 h prior to agarose gel electrophoresis. Mobility retardation of the supercoiled plasmid was observed as a characteristic ladder in samples incubated at temperatures between 16 and 37 °C (Fig. 5*A*). This band shift must be the result of a transition from non-cruciform to a cruciform conformation. Unexpectedly, mobility retardation was most prominent in the plasmid incubated at room temperature. The ladder appeared as soon as 1 min after the start of incubation at room temperature, and the fraction of upward shifted bands expanded with increasing duration of incubation (Fig. 5*B*). Since the band pattern of the sample after 1 h of incubation was similar to that prepared by the alkaline lysis method, incubation at room temperature for 1 h appeared to be sufficient for a large fraction of the plasmid to form the cruciform extrusion. The sample incubated for 1 h at room temperature was subjected to two-dimensional gel analysis, producing similar results to that prepared by the alkaline lysis method (data not shown).

*EMSA*—We performed EMSA using the anti-cruciform antibody 2D3, which recognizes cruciform DNA (24). To reduce the size of the PATRR fragments, we subcloned BamHI/HincII fragments into a plasmid vector (Fig. 1*B*). The sizes of the subfragments from psPATRR11 and pΔPATRR11 are 375 and 162 bp, respectively. Plasmid DNA samples prepared by the Triton lysis method were used for EMSA. Without any preincubation of the plasmid, the addition of the 2D3 antibody did not result in a band shift for the subfragments from psPATRR11 (Fig. 6, *lanes 9* and *10*) or from pΔPATRR11 (data not shown). However, when the plasmids were incubated at room temperature for 7 days prior to the reaction to maximize the fraction adopting cruciform conformation, subfragments from psPATRR11 demonstrated a robust band shift in an antibody concentration-dependent manner (Fig. 6, *lanes 1–3*). The subfragment from pΔPATRR11 did not show a band shift even after incubation at room temperature for 7 days (Fig. 6, *lanes 4–6*). Since retardation of psPATRR11 was not observed with control monoclonal antibodies (Fig. 6, *lanes 7* and *8*), this band shift indicates that the 2D3 antibody bound specifically to the PATRR plasmid incubated at room temperature.

*AFM Study*—We used atomic force microscopy to examine the tertiary structure of the PATRR plasmid and to directly visualize the cruciform extrusion (27). We examined psPATRR11 prepared by the Triton lysis method. Initial examination revealed no overt cruciform extrusion (Fig. 7*A*). However, after incubation of the DNA samples at room temperature for 7 days, 37 of 45 plasmids analyzed (82.2%) showed cruci-

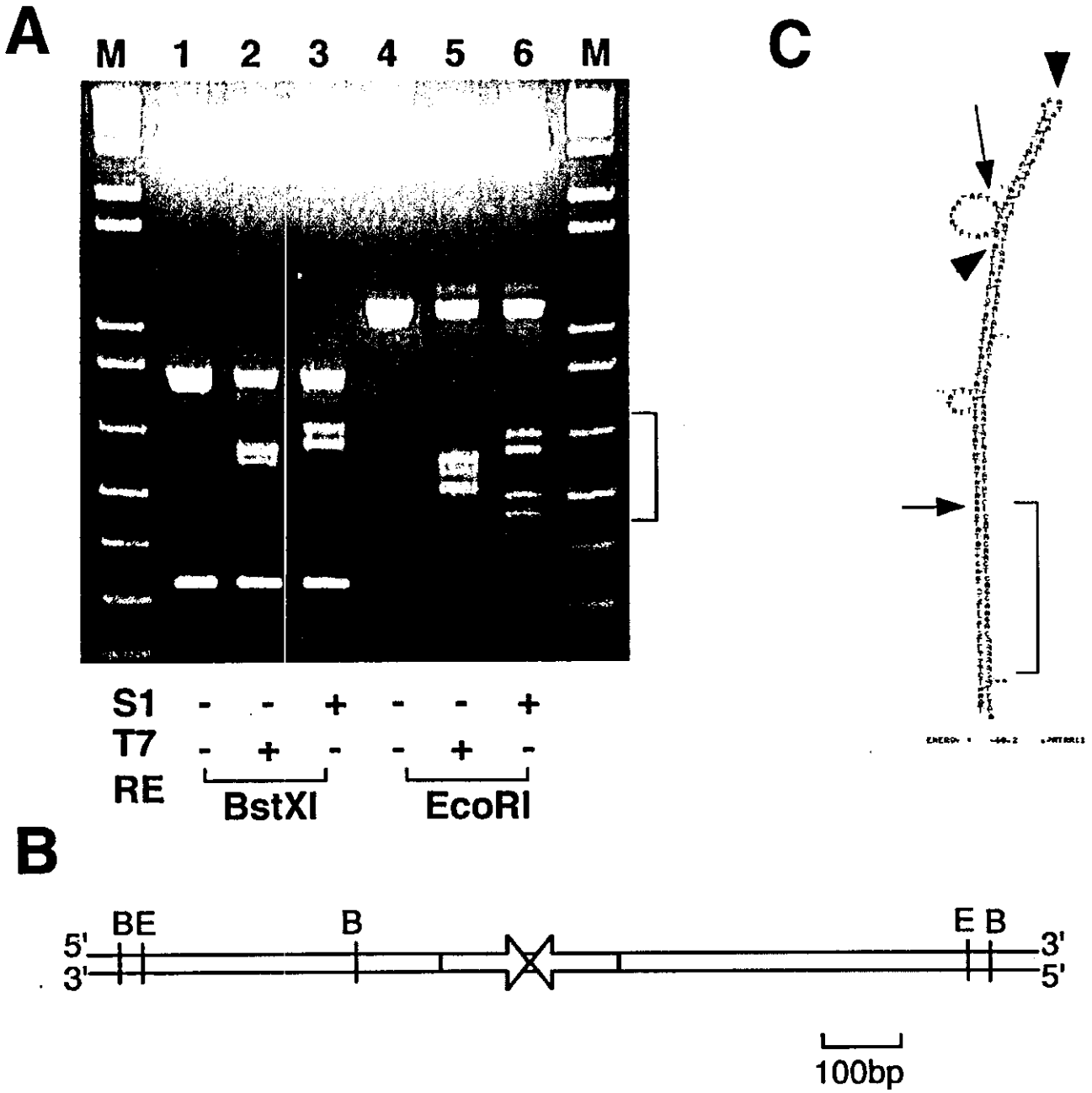



FIG. 4. **Mapping of nuclease cleavage sites in the PATRR.** *A*, mapping by digestion with restriction enzyme. *S1*, S1 nuclease; *T7*, T7 endonuclease; *RE*, restriction enzyme. *B*, restriction map of the PATRR-flanking region. *C*, mapping by sequencing. The entire PATRR is shown as a putative hairpin structure predicted by mfold software (mfold.burnet.edu.au/dna_form). The *arrows* indicate cleavage sites with T7 endonuclease, whereas the *arrowheads* indicate those with S1 nuclease. the relatively GC-rich region is indicated by a *bracket*.

form extrusions (Fig. 7, *B* and *C*). The mean length of extrusions was 23.2 ± 4.3 nm. Since the size of the PATRR is 204 bp, the predicted size for the cruciform arm is 34.68 nm. When the GC-rich region at both ends of the PATRR (31 bp each) was excluded, the predicted size of the cruciform arm is 24.14 nm, which nearly corresponds to the observed size. Two types of extrusions are expected; those with extended or those with acute geometry of the cruciform arms. However, all of the observed cruciforms had acute geometry (X-type), with a mean cruciform arm angle of 89.3 ± 35.2°. The small cruciform, the size of which was predicted to be 6.12 nm from its nucleotide size (18 bp), was too small to be identified by AFM.

To localize the cruciform extrusion of the psPATRR11 within the PATRR region, we separated the insert including the PATRR (1 kb) from the plasmid vector (4 kb). psPATRR11 was cross-linked with psoralen and UV prior to EcoRI digestion. Cruciform extrusion was observed at the center of the 1-kb fragment but not in the 4-kb fragment (Fig. 7*D*), demonstrating that the PATRR formed a cruciform extrusion.

## DISCUSSION

In this study, we have demonstrated that the 11q23 PATRR located at the t(11;22) breakpoint region can form a cruciform structure *in vitro*. Most individuals carry a 445-bp PATRR with an AT content of 93%. The homology between the proximal and distal arms is 98%, comprising a nearly complete palindromic structure (5). Previously, we identified a short polymorphic allele of the PATRR among a group of individuals. The total

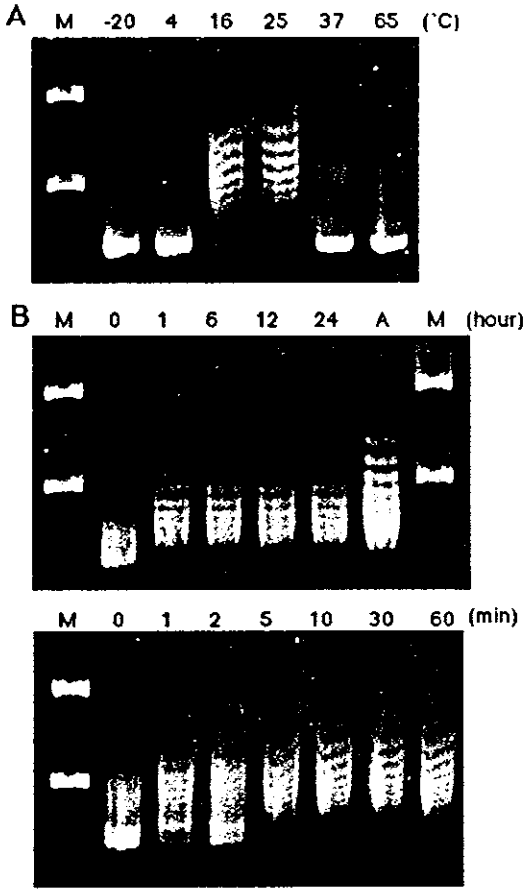



FIG. 5. **Temperature-sensitive structural transition of the PATRR.** *A*, temperature-dependent gel mobility shift of the PATRR plasmid. psPATRR11 prepared by the Triton lysis method was incubated at various temperatures shown on the *top*. *A*, psPATRR11 prepared by the alkaline lysis method. *M*, molecular size markers. *B*, a time course of incubation at room temperature. psPATRR11 was incubated at room temperature for periods indicated on the *top*.

size of the short PATRR is 204 bp, with an AT content of 88%. The homology between the proximal and distal arms is 88%, with a 34-bp asymmetric center. Since it is difficult to clone DNA containing the complete palindrome, the plasmid we used for structural analysis contained this shorter version of the 11q23 PATRR. Our results clearly demonstrate that the short 11q23 PATRR forms a cruciform configuration at physiological conditions. Although there is no direct evidence that the short PATRR is involved in genesis of the translocation, one exceptional t(11;22) case appears to have originated from a short PATRR (5). Thus, it seems likely that the short PATRR behaves similarly to the long PATRR. It is reasonable to predict that the long 445-bp PATRR, as well as other PATRRs involved in chromosomal translocations, may also form cruciform structures that induce genomic instability.

It has been suggested that palindromic DNA capable of forming unstable cruciform structures induces both homologous and illegitimate recombination, leading to translocations. The data presented here support this proposed mechanism. Although the mechanism of PATRR-mediated double-strand break induction remains unknown, it is possible that the cruciform structure blocks progression of the replication fork, generating free ends that must be repaired (28, 29). Alternatively, a conformation-specific endonuclease may cleave at the center of the PATRR. However, no key enzyme to catalyze this cleavage has been identified. Cruciform structures that comprise a four-way junction resemble Holliday junctions, which may be a target for

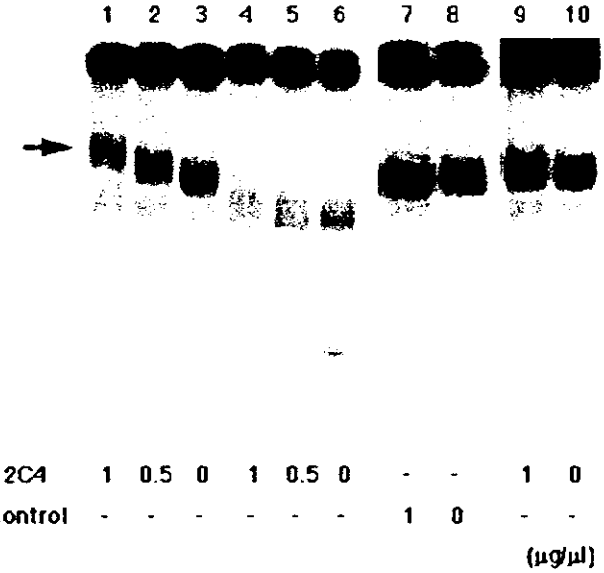



FIG. 6. **Gel electrophoresis mobility shift assay of the PATRR fragment.** psPATRR11 (*lanes 1–3*) or pΔPATRR11 (*lanes 4–6*) prepared by the Triton lysis method was incubated at room temperature for 7 days and then reacted with anti-cruciform antibody, 2D3. For psPATRR11, another monoclonal antibody was used as a control (*lanes 7–8*). psPATRR11 without incubation at room temperature was also reacted with 2D3 (*lanes 9–10*). The concentration of 2D3 antibody is indicated at the *bottom*. The *arrow* indicates the shifted band.

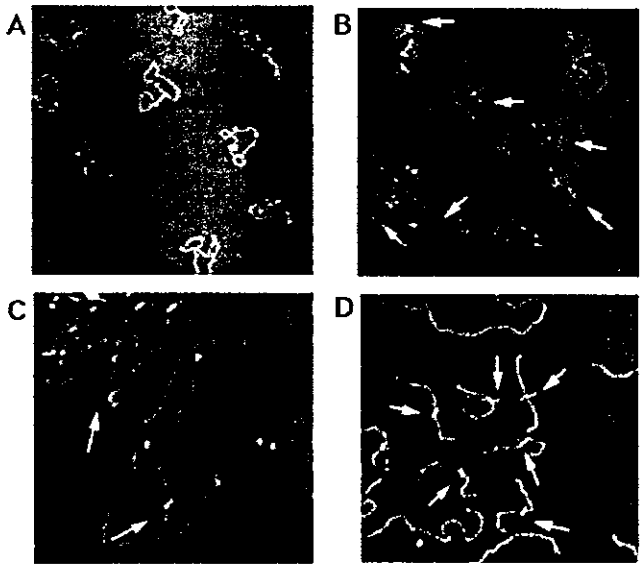



FIG. 7. **AFM image of the PATRR plasmid.** *A*, an AFM image of psPATRR11 prepared by the Triton lysis method. *B*, an AFM image of the psPATRR11 after incubation at room temperature. *C*, another image of psPATRR11 at a higher magnification. *D*, an AFM image of psPATRR11 after DNA cross-linking followed by restriction digestion. The *arrows* indicate the cruciform extrusion. Longer fragments without cruciform structure may originate from the cloning vector. The *scale bar* is indicated in μm.

Holliday junction resolvase. Such a break might be repaired by non-homologous end joining, for there are symmetrical bx; 1deletionsba at the center of the PATRR, and there is no homology between the sequences of the translocation breakpoint partners (16).

In this report, we have demonstrated that cruciform extrusion is temperature-sensitive. Interestingly, cruciform formation is most efficient at room temperature, consistent with previous observations by Panyutin *et al.* (19). We have often

observed strand separation of the PATRR fragment even at 37 °C.[2] The temperature-induced drop in cruciform formation might be attributed to the competition for superhelix energy release, taking place between cruciform and open regions within the premelting range. We previously reported that only meiotic cells undergo this translocation (7). This prompted the hypothesis that PATRR-mediated translocations may represent errors in meiotic recombination and that the double-strand break catalytic enzyme involved in meiotic recombination might recognize a specific DNA conformation. Male meiosis might be permissive to such temperature-sensitive cruciform extrusion, and we have previously identified frequent *de novo* t(11;22)s in male meiosis. Unfortunately, a similar opportunity does not exist to examine female meiotic cells. We have analyzed only one case of a *de novo* t(11;22), and it indicated an origin in the paternal germ line.[2] Thus, it is not unreasonable to imagine that PATRR-mediated translocations might occur only in male meiotic cells generated at a lower body temperature.

*In vivo* cruciform extrusion is still controversial. No report has directly demonstrated *in vivo* cruciform structures in palindromic DNA. Although previous studies addressed cruciform DNA in *E. coli*, others have argued against the phenomenon (18, 26, 30). Based on the fact that PATRR plasmids prepared by the non-denaturing method do not demonstrate cruciform extrusion, the favored PATRR cruciform conformation is likely to be inhibited in *E. coli*. Cruciform extrusion is known to require sufficient negative superhelicity. The natural negative superhelicity of a plasmid in *E. coli* may not afford it the opportunity to adopt a cruciform configuration. Since the replication of a plasmid may be too rapid for a transition to a cruciform arrangement to take place, this may be the reason why the observation of cruciform configurations is rare in bacteria. In mammals, the doubling time is sufficiently long to permit palindromic DNA to generate cruciform extrusion configurations. Thus, although a cruciform configuration is generally thought to be energetically unfavorable, our results indicate that the PATRR prefers to adopt a cruciform configuration even at physiological conditions. Further, recent studies have demonstrated that mammalian chromatin remodeling factors induce negative superhelical torsion of DNA, leading to cruciform extrusion of palindromic DNA (31). This suggests that palindromic DNA may adopt a cruciform conformation within the chromatin scaffold. The conformational status of genomic DNA affects many cellular functions through the regulation of DNA replication or transcription (23, 32). Thus, the elucidation of an *in vivo* cruciform structure may lead to a further understanding of the regulation of such cellular functions.

[2] B. S. Emanuel and H. Kurahashi, unpublished data.

*Acknowledgments*—The authors thank Drs. M. Zannis-Hadjopoulos and G. B. Price for 2D3 anti-cruciform DNA monoclonal antibody. AFM images were measured by R. Hirota and Y. Iwazaki of the Research Institute of Biomolecule Metrology, Tsukuba, Ibaraki, Japan.

REFERENCES

1. Kurahashi, H., Shaikh, T. H., Hu, P., Roe, B. A., Emanuel, B. S., and Budarf, M. L. (2000) *Hum. Mol. Genet.* **9,** 1665–1670
2. Kurahashi, H., Shaikh, T. H., Zackai, E. H., Celle, L., Driscoll, D. A., Budarf, M. L., and Emanuel, B. S. (2000) *Am. J. Hum. Genet.* **67,** 763–768
3. Edelmann, L., Spiteri, E., Koren, K., Pulijaal, V., Bialer, M. G., Shanske, A., Goldberg, R., and Morrow, B. E. (2001) *Am. J. Hum. Genet.* **68,** 1–13
4. Tapia-Paez, I., Kost-Alimova, M., Hu, P., Roe, B. A., Blennow, E., Fedorova, L., Imreh, S., and Dumanski, J. P. (2001) *Hum. Genet.* **109,** 167–177
5. Kurahashi, H., and Emanuel, B. S. (2001) *Hum. Mol. Genet.* **10,** 2605–2617
6. Emanuel, B. S., and Shaikh, T. H. (2001) *Nat. Rev. Genet.* **2,** 791–800
7. Kurahashi, H., and Emanuel, B. S. (2001) *Nat. Genet.* **29,** 139–140
8. Dunham, I., Shimizu, N., Roe, B. A., Chissoe, S., Hunt, A. R., Collins, J. E., Bruskiewich, R., Beare, D. M., Clamp, M., Smink, L. J., Ainscough, R., Almeida, J.P., Babbage, A., Bagguley, C., Bailey, J., Barlow, K., Bates, K. N., Beasley, O., Bird, C. P., Blakey, S., Bridgeman, A. M., Buck, D., Burgess, J., Burrill, W. D., O'Brien, K. P., *et al.* (1999) *Nature* **402,** 489–495
9. Shaikh, T. H., Kurahashi, H., Saitta, S. C., O'Hare, A. M., Hu, P., Roe, B. A., Driscoll, D. A., McDonald-McGinn, D. M., Zackai, E. H., Budarf, M. L., and Emanuel, B. S. (2000) *Hum. Mol. Genet.* **9,** 489–501
10. Kehrer-Sawatzki, H., Haussler, J., Krone, W., Bode, H., Jenne, D. E., Mehnert, K. U., Tummers, U., and Assum, G. (1997) *Hum. Genet.* **99,** 237–247
11. Rhodes, C. H., Call, K. M., Budarf, M. L., Barnoski, B. L., Bell, C. J., Emanuel, B. S., Bigner, S. H., Park, J. P., and Mohandas, T. K. (1997) *Cytogenet. Cell Genet.* **78,** 247–252
12. Debeer, P., Mols, R., Huysmans, C., Devriendt, K., Van de Ven, W. J., and Fryns, J. P. (2002) *Clin. Genet.* **62,** 410–414
13. Spiteri, E., Babcock, M., Kashork, C. D., Wakui, K., Gogineni, S., Lewis, D. A., Williams, K. M., Minoshima, S., Sasaki, T., Shimizu, N., Potocki, L., Pulijaal, V., Shanske, A., Shaffer, L. G., and Morrow, B. E. (2003) *Hum. Mol. Genet.* **12,** 1823–1837
14. Kurahashi, H., Shaikh, T., Takata, M., Toda, T., and Emanuel, B. S. (2003) *Am. J. Hum. Genet.* **72,** 733–738
15. Nimmakayalu, M. A., Gotter, A. L., Shaikh, T. H., and Emanuel, B. S. (2003) *Hum. Mol. Genet.* **12,** 2817–2825
16. Gotter, A. L., Shaikh, T. H., Chieffo, C., Budarf, M. L., Rhodes, C. H., and Emanuel, B. S. (2003) *Hum. Mol. Genet.* **13,** 103–115
17. Mizuuchi, K., Mizuuchi, M., and Gellert, M. (1982) *J. Mol. Biol.* **156,** 229–243
18. Courey, A. J., and Wang, J. C. (1983) *Cell* **33,** 817–829
19. Panyutin, I., Klishko, V., and Lyamichev, V. (1984) *J. Biomol. Struct. Dyn.* **1,** 1311–1324
20. Peck, L. J., and Wang, J. C. (1983) *Proc. Natl. Acad. Sci. U. S. A.* **80,** 6206–6210
21. Dai, X., Greizerstein, M. B., Nadas-Chinni, K., and Rothman-Denes, L.B. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94,** 2174–2179
22. Vologodskaia, M. Y., and Vologodskii, A. V. (1999) *J. Mol. Biol.* **289,** 851–859
23. Kim, E. L., Peng, H., Esparza, F. M., Maltchenko, S. Z., and Stachowiak, M. K. (1998) *Nucleic Acids Res.* **26,** 1793–1800
24. Frappier, L., Price, G. B., Martin, R. G., and Zannis-Hadjopoulos, M. (1987) *J. Mol. Biol.* **193,** 751–758
25. Bell, L. R., and Byers, B. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47,** 829–840
26. Haniford, D. B., and Pulleyblank, D. E. (1985) *Nucleic Acids Res.* **13,** 4343–4363
27. Shlyakhtenko, L. S., Potaman, V. N., Sinden, R. R., and Lyubchenko, Y. L. (1998) *J. Mol. Biol.* **280,** 61–72
28. Leach, D. R. (1994) *BioEssays* **16,** 893–900
29. Lobachev, K. S., Gordenin, D. A., and Resnick, M. A. (2002) *Cell* **108,** 183–193
30. Lyamichev, V., Panyutin, I., and Mirkin, S. (1984) *J. Biomol. Struct. Dyn.* **2,** 291–301
31. Havas, K., Flaus, A., Phelan, M., Kingston, R., Wade, P. A., Lilley, D. M., and Owen-Hughes, T. (2000) *Cell* **103,** 1133–1142
32. Novac, O., Alvarez, D., Pearson, C. E., Price, G. B., and Zannis-Hadjopoulos, M. (2002) *J. Biol. Chem.* **277,** 11174–11183

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 279, No. 20, Issue of May 14, pp. 20626–20635, 2004
Printed in U.S.A.
© 2004 by The American Society for Biochemistry and Molecular Biology, Inc.

# Interaction between Hex and GATA Transcription Factors in Vascular Endothelial Cells Inhibits *flk-1/KDR*-mediated Vascular Endothelial Growth Factor Signaling*

Received for publication, August 7, 2003, and in revised form, March 10, 2004
Published, JBC Papers in Press, March 10, 2004, DOI 10.1074/jbc.M308730200

**Takashi Minami‡§¶, Takeshi Murakami‡§, Keiko Horiuchi‡, Mai Miura‡, Tamio Noguchi‖, Jun-ichi Miyazaki**, Takao Hamakubo‡, William C. Aird‡‡§§, and Tatsuhiko Kodama‡**

*From the ‡Research Center for Advanced Science and Technology, the University of Tokyo, Tokyo, 153-8904 Japan, the ‖Department of Applied Molecular Biosciences, Graduate School of Bioagricultural Sciences, Nagoya University, Nagoya 464-8604 Japan, the **Division of Stem Cell Regulation Research, Osaka University Medical School, Osaka 565-0871 Japan, and the ‡‡Department of Molecular and Vascular Medicine, Beth Israel Deaconess Medical Center/Harvard Medical School, Boston, Massachusetts 02215*

**Recent evidence supports a role for GATA transcription factors as important signal intermediates in differentiated endothelial cells. The goal of this study was to identify proteins that interact with endothelial-derived GATA transcription factors. Using yeast two-hybrid screening, we identified hematopoietically expressed homeobox (Hex) as a GATA-binding partner in endothelial cells. The physical association between Hex and GATA was confirmed with immunoprecipitation in cultured cells. Hex overexpression resulted in decreased *flk-1/KDR* expression, both at the level of the promoter and the endogenous gene, and attenuated vascular endothelial growth factor-mediated tube formation in primary endothelial cell cultures. In electrophoretic mobility shift assays, Hex inhibited the binding of GATA-2 to the *flk-1/KDR* 5′-untranslated region GATA motif. Finally, in RNase protection assays, transforming growth factor β1, which has been previously shown to decrease *flk-1* expression by interfering with GATA binding activity, was shown to increase Hex expression in endothelial cells. Taken together, the present study provides evidence for a novel association between Hex and GATA and suggests that transforming growth factor β-mediated repression of *flk-1/KDR* and vascular endothelial growth factor signaling involves the inducible formation of inhibitory Hex-GATA complexes.**

The endothelium displays remarkable diversity in health and disease. Endothelial cell phenotypes are modulated in space and time by mechanisms that involve the highly coordinated combinatorial action of transcriptional modules (1). The GATA family of transcription factors consists of six members (GATA-1 to -6). Of the various GATA factors, GATA-2, -3, and -6 are expressed in vascular endothelial cells (2). Initially, these transacting factors were believed to play a singular role in the differentiation of endothelial cells during development (3–5). However, more recently, studies have provided evidence that GATA factors also function as "immediate early genes," coupling changes within the extracellular environment to changes in downstream target gene expression (6, 7). In addition to binding directly to their consensus motif (A/T)GATA(A/G), GATA factors have been shown to participate in a wide array of protein-protein interactions with Zinc finger type transcription factors, including SP1, EKLF, RBTN2, ER, and GATA itself (8–13); homeobox transcription factors such as HNF-1α, Nkx-2.5, Pit-1, and TTF-1 (14–17); Ets family PU.1 (18); MADS box protein SRF (19); RUNT domain protein AML1 (20); AP1 (21); NF-ATc (6); and Smad3 (22). During cell differentiation, GATA factors may be associated with cofactors, such as basal transcription factors, FOG-1, FOG-2, CBP/p300, and histone deacetylase-3 (23–29). Taken together, these data suggest that GATA transcription factors are involved in highly complex regulatory pathways and that the dissection of these networks may provide valuable insight into the transcriptional control of endothelial phenotypes.

To that end, the goal of the present study was to identify partner proteins that interact with GATA factors in endothelial cells. Using a yeast two-hybrid system, we describe a physical interaction between GATA family of transcription factors and the homeobox protein Hex. The interaction between Hex-GATA is associated with reduced GATA binding activity and transcriptional activity, decreased expression of the GATA-2 target gene, *flk-1/KDR*, and secondary attenuation of VEGF-mediated[1] signaling. Finally, we present data supporting the notion that TGF-β exerts its anti-angiogenesis effect by a Hex-GATA-*flk-1/KDR*-dependent mechanism.

## EXPERIMENTAL PROCEDURES

*Cell Culture*—Human umbilical vein endothelial cells (HUVEC) (Clonetics, La Jolla, CA) were cultured in EGM-2 MV medium (Clonetics). Human embryonic kidney (HEK)-293 cells (ATCC CRL-1573) and COS-7 cells (ATCC CRL-1651) were cultured in Dulbecco's modified Eagle's medium supplemented with 10% heat-inactivated fetal bovine serum. HUVEC were used within the first eight passages.

*Plasmids*—The construction of the *flk-1/KDR*-luc, *flk-1/KDR* (GATA mut), and *flk-1/KDR* (SP1 mut) plasmids were previously described (30). Human GATA-2 expression plasmid (pMT$_2$-GATA2) was a kind

* This work was supported by funds from the Uehara Memorial Foundation in Japan (to T. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ These authors contributed equally to this work.
¶ To whom correspondence should be addressed: Research Center for Advanced Science and Technology, the University of Tokyo, 4-6-1 Komaba, Meguro Tokyo, 153-8904. Tel.: 81-3-5452-5403; Fax: 81-3-5452-5232; E-mail: minami@med.rcast.u-tokyo.ac.jp.

§§ Supported in part by National Institutes of Health Grants HL 60585-03, HL 63609-02, and HL 65216-02.

[1] The abbreviations used are: VEGF, vascular endothelial growth factor; TGF, transforming growth factor; HUVEC, human umbilical vein endothelial cell(s); HEK, human embryonic kidney; GBD, Gal4 DNA-binding domain; GAD, Gal4 activation domain; UTR, untranslated region; EGFP, enhanced green fluorescence protein; IRES, internal ribosome entry sites; GAPDH, glyceraldehyde-3-phosphate dehydrogenase.


This paper is available on line at http://www.jbc.org

gift from Dr. Stuart H. Orkin (Harvard Medical School, Boston, MA). Human GATA-3 and -6 and Hex cDNA fragments were amplified using PCR from reverse-transcribed HUVEC total RNA. To generate the Hex expression vector, Hex cDNA was subcloned into the pcDNA3 vector (Invitrogen). To generate the plasmids expressing Gal4 DNA-binding domain (GBD) fused with GATA-2, -3, and -6, each of the three GATA cDNA fragments was inserted into pGBKT7 (Clontech, Palo Alto, CA). To construct the plasmid expressing Gal4 activation domain (GAD) fused to Hex, the Hex cDNA fragment was subcloned into pGAD424 (Clontech). To generate a FLAG-tagged Hex (pFLAG-Hex) and FLAG-tagged GATA-2 (pFLAG-GATA2), Hex and GATA-2 cDNA fragments were subcloned into pFLAG (Sigma), respectively. For construction of pGEM-hflk, a 266-bp human *flk-1/KDR* cDNA fragment was amplified from reverse-transcribed HUVEC total RNA and subcloned into pGEM-T-easy (Promega). Similarly, pGEM-hHex and pGEM-hGAPDH were derived by ligating a 296-bp human Hex cDNA fragment and a 283-bp human GAPDH fragment into pGEM-T-easy. Orientation was confirmed by automated DNA sequencing.

*Yeast Two-hybrid Screening and β-Galactosidase Assays*—A HUVEC cDNA library was constructed using a two-hybrid cDNA library construction kit (Clontech). AH109 yeast were transfected sequentially with the GATA-6 bait vector and cDNA library and then spread on synthesized dropout medium plate in the absence of tryptophan, leucine, and histidine. After 5 days of incubation at 30 °C, colonies complemented by histidine autotrophy were isolated and confirmed to be positive by β-galactosidase assay according to the manufacturer's instruction (Clontech; yeast protocols handbook). The plasmids from the positive colonies were purified, and the inserts were subsequently sequenced. To analyze the protein-protein interaction, Y190 yeast were co-transfected with pGAD-Hex and a plasmid in which GBD was fused either with GATA-2, -3, and -6. Similarly, Y190 yeast were co-transfected with pGBD-GATA2 and a plasmid containing GAD fused with Hex. The cells were spread and incubated on the synthesized dropout medium plate without tryptophan and leucine.

*Transfection of COS-7 Cells and Immunoprecipitation Assays*—COS-7 cells were co-transfected with either pFLAG-Hex and $pMT_2$-GATA2, or pFLAG-GATA2 and pcDNA3-Hex expression plasmids using the FuGENE 6 reagent (Roche Applied Science) as instructed by the manufacturer. Two days later, the transfected cells were freeze-thawed three times and incubated for 30 min on ice with cell lysis buffer (0.1% IGEPAL CA-630, 50 mM Tris-HCl, 150 mM NaCl, 5 mM EDTA, 0.5 μg/ml pepstatin A, 10 μg/ml leupeptin, 2 μg/ml aprotinin, 200 μM phenylmethylsulfonyl fluoride, pH 7.5), followed by centrifugation at 20,000 × *g* for 5 min. The supernatant was incubated with anti-FLAG polyclonal antibody (Sigma) overnight at 4 °C. The resulting mixture was then mixed and incubated with protein G-Sepharose (Amersham Biosciences) for 1 h at 4 °C. The immobilized beads were washed five times with 1 ml of cell lysis buffer containing 1.5% IGEPAL CA-630. Each sample was separated on 12% SDS-polyacrylamide gel and transferred to a nitrocellulose membrane (Amersham Biosciences). The membrane was incubated either with anti-FLAG M5 monoclonal antibody (Sigma), anti-Myc monoclonal antibody (Invitrogen) or anti-GATA-2 antibody (Santa Cruz, CA).

Alternatively, subconfluent HUVEC ($2 \times 10^7$ cells) were harvested for nuclear extracts according to the mild nitrogen cavitation method (31) to keep intact the protein-protein associations. 1 μg of nuclear extracts was precleaned by centrifugation with 4 μg of control IgG (Santa Cruz). The resulting supernatant was incubated with 30 μg of agarose-conjugated anti-GATA-2 monoclonal antibody (Santa Cruz, sc-267 AC) or an identical amount of isotype-matched mouse control IgG (Santa Cruz, sc-2343) overnight at 4 °C. The immobilized samples were separated and transferred to a polyvinylidene difluoride membrane. The membrane was then probed for Hex and GATA-2 by Western blot analysis, using anti-Hex antibody (generated by Dr. Tamio Noguchi, Japan) and anti-GATA-2 antibody (Santa Cruz, sc-16044, and generated by Dr. Stuart H. Orkin, Harvard Medical School, Boston), respectively. The complexes were visualized with an ECL advance Western blotting detection kit (Amersham Biosciences).

*RNA Isolation and RNase Protection Assays*—HUVEC were serum-starved in EBM-2 medium containing 0.5% fetal bovine serum. 18 h later, HUVEC were treated with 2.5–10 ng/ml TGF-β1 (Peprotec, Rocky Hill, NJ) for 18 h (for Hex) or 24h (for *flk-1/KDR*). Alternatively, HUVEC were infected with adenoviruses encoding the IRES-mediated green fluorescence protein (EGFP) (Adeno-Blank) or IRES coupled to Hex and EGFP (Adeno-Hex). Adeno-Blank and Adeno-Hex were generated with a ligation into adenovirus cosmid vector and a co-transfection into HEK-293 (32). Infections were carried out at a multiplicity of infection of 20 for 24 h. HUVEC were harvested for total RNA at the

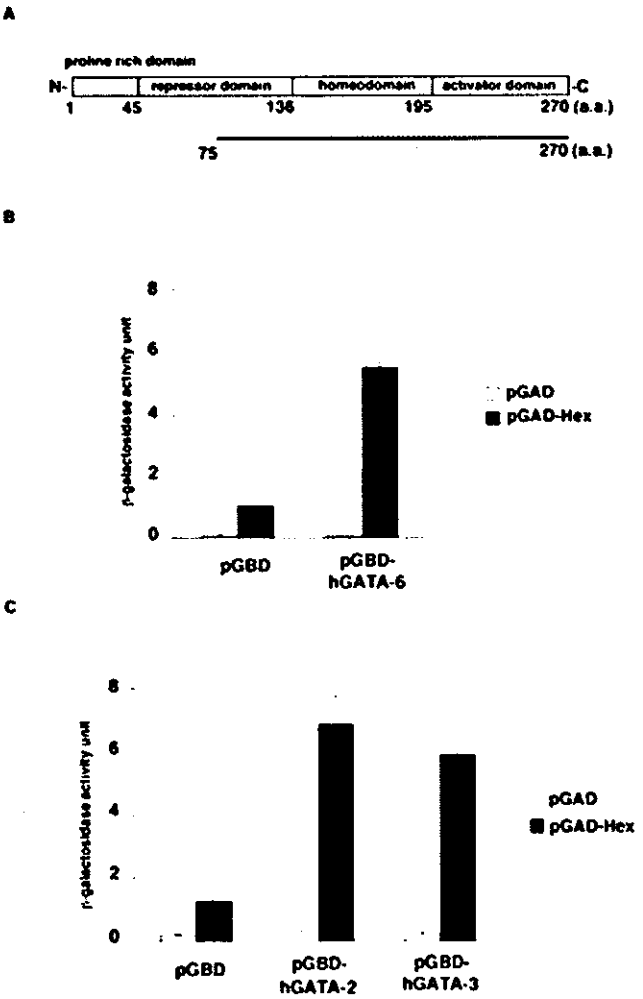



FIG. 1. **Hex interacts the GATA family proteins.** *A*, schematic representation of Hex. The *black bar* indicates the isolated fragment by two-hybrid screening. *B* and *C*, the yeast strain Y190 was co-transformed with HEX-pGAD and with human GATA-2, GATA-3, or GATA-6 in pGBD. The cells were plated in synthesized dropout medium lacking tryptophan, leucine, and histidine. β-Galactosidase assays and calculation of the activity units were performed as described under "Experimental Procedures." The values are the means ± S.E. of three separate experiments. *a.a.*, amino acids

times indicated, using the Isogen reagent (Nippon Gene). For *in vitro* transcription, *flk-1/KDR*-, Hex-, and GAPDH-specific $^{32}P$-labeled riboprobes were synthesized from pGEM-hflk, pGEM-hHex, and pGEM-hGAPDH, respectively. Riboprobes were synthesized using SP6 (for *flk-1/KDR* and GAPDH) or T7 (for Hex) RNA polymerase (Ambion, Austin, TX) and purified with a G-50 spun column (Amersham Biosciences). RNase protection assays were performed with a RPA III kit (Ambion) according to the manufacturer's instructions.

*Transfection of HEK-293 Cells or HUVEC and Analysis of Luciferase Activity*—HEK-293 cells and HUVEC were transfected as described previously (30). Briefly, either $2 \times 10^5$ cells/well of HEK-293 cells or $1 \times 10^5$ cells/well of HUVEC were seeded onto 12-well plates 18–24 h before transfection. 0.05 pmol of the luciferase reporter plasmid (either *KDR*-, *KDR* (GATA mut)-, or *KDR* (SP1 mut)-luc), 50 ng of pRL-CMV (Promega), 0.075 pmol of the GATA expression vector and either 0.0375 or 0.075 pmol of Hex expression vector were incubated with 2 μl of FuGENE 6. As a negative control, empty vector ($pMT_2$ and pcDNA3) was transfected instead of GATA-2 and Hex expression vector, respectively. 24 h later for HEK-293 cells and 48 h later for HUVEC, the cells were washed with phosphate-buffered saline, lysed, and assayed for luciferase activity using the dual luciferase reporter assay system (Promega) and a Lumat LB 9507 luminometer (Berthold, Gaithersburg, MD).

*Sandwich Tube Formation Assays*—400-μl aliquots of type-I collagen gel (Koken) containing EGM-2 MV medium (Clonetics) without basic fibroblast growth factor were added to 24-well plates and allowed to gel

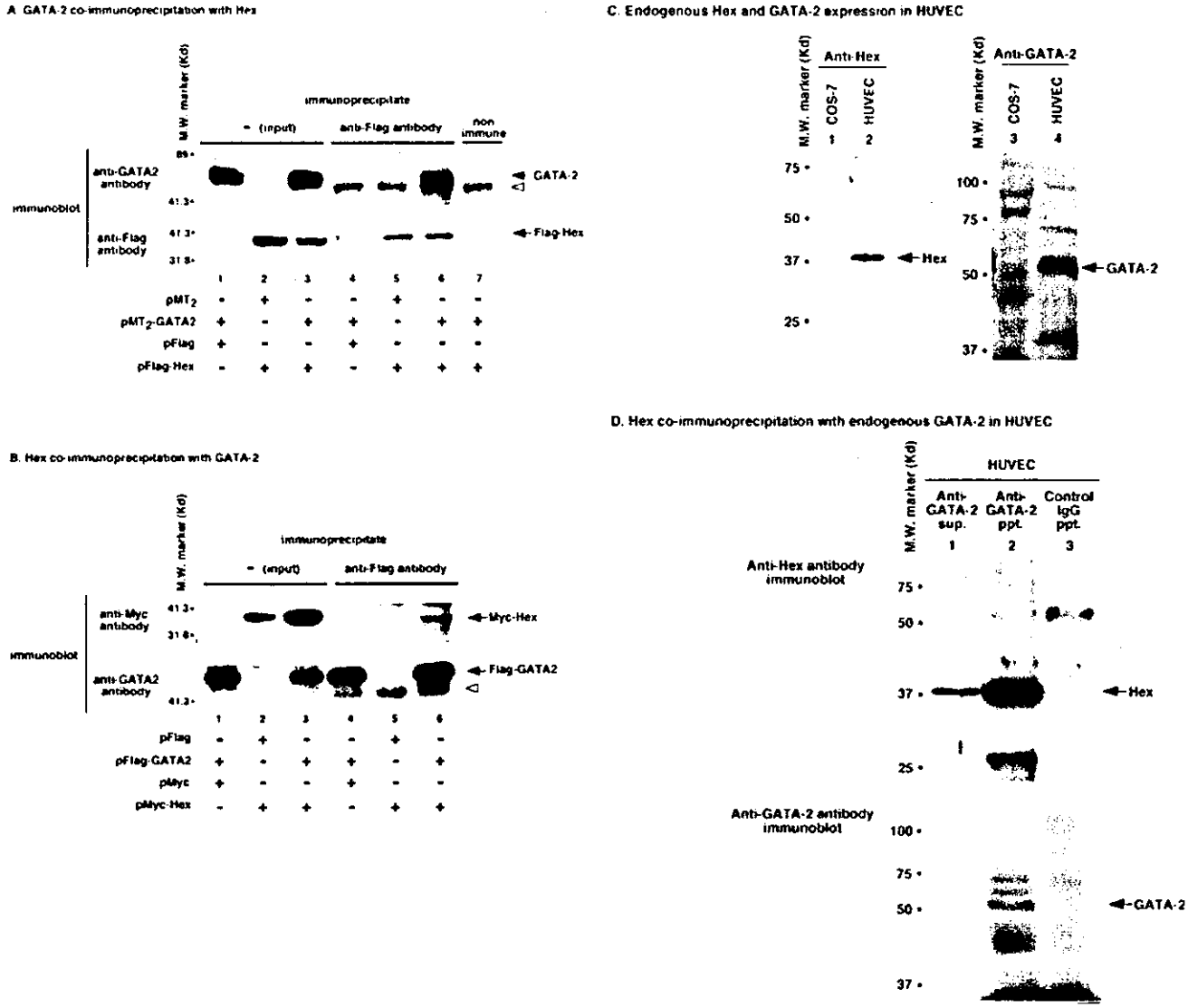



FIG. 2. **Physical interaction between GATA-2 and Hex by immunoprecipitation analysis.** *A*, expression vector in combinations either $pMT_2$ (*lanes 2* and *5*), $pMT_2$-GATA2 (*lanes 1, 3, 4, 6,* and *7*), pFLAG (*lanes 1* and *4*), or pFLAG-HEX (*lanes 2, 3, 5, 6,* and *7*) were co-transfected into COS-7 cells. Extracted proteins from the cells were untreated (*lanes 1–3*) or precipitated by either anti-FLAG antibody (*lanes 4–6*) or nonimmune antibody (*lane 7*) and then separated by 8% SDS-PAGE. The transferred membrane was immunoblotted with anti-GATA2 (*upper panel*) or FLAG (*lower panel*) antibody. The *closed arrowhead* indicates the specific GATA-2 and FLAG-tagged Hex complex. *B*, expression vectors either pFLAG (*lanes 2* and *5*), pFLAG-GATA2 (*lanes 1, 3, 4,* and *6*), pMyc (*lanes 1* and *4*), or pMyc-HEX (*lanes 2, 3, 5,* and *6*) were co-transfected into COS-7 cells. Extracted proteins from the cells were untreated (*lanes 1–3*) or precipitated by anti-FLAG antibody (*lanes 4–6*) and then separated by 8% SDS-PAGE. The transferred membrane was immunoblotted with anti-Myc (*upper panel*) or GATA-2 (*lower panel*) antibody. The *closed arrowhead* indicates the specific FLAG-tagged GATA-2 and Myc-tagged Hex complex. The faster migration complex (*open arrowhead*) represents nonspecific immunoreactants. The results are representative of three independent experiments. *C*, 80 μg of COS-7 (*lanes 1* and *3*) and HUVEC (*lanes 2* and *4*) were separated by 10%SDS-PAGE. The transferred membrane was immunoblotted with anti-Hex antibody (*left panel*) and anti-GATA-2 antibody (*right panel*). *D*, nuclear extracts from HUVEC were immunoprecipitated by either anti-GATA2 antibody or control mouse IgG, then separated according to supernatant (*sup., lane 1*) and precipitate (*ppt., lanes 2* and *3*). The samples were electrophoresed and immunoblotted with anti-Hex and anti-GATA-2 antibodies. The *arrow* indicates the specific Hex and GATA protein.

for 1 h at 37 °C. HUVEC infected with Adeno-Blank or Adeno-Hex were seeded at $1 \times 10^5$ cells/well and incubated for 24 h in 5% $CO_2$. The medium was removed, and the HUVEC were covered with 400 μl of the gel. The plate was incubated for 30 min at 37 °C. The cells were incubated with 1 ml of EGM-2 MV media in the absence of basic fibroblast growth factor, in the presence or absence of SU1498 (Calbiochem, San Diego, CA). Two days later, a branched capillary network was visualized under a microscope. Images from at least three different areas in each well were captured by a digital camera under a microscope.

*Nuclear Extracts and Electrophoretic Mobility Shift Assays*—Nuclear extracts were prepared as previously described (30, 33). Double-stranded oligonucleotides were labeled with [α-$^{32}$P]dCTP and Klenow fragment and purified by spun column (Amersham Biosciences). 10 μg of HUVEC nuclear extracts were incubated with 10 fmol of $^{32}$P-labeled probe, 1 μg of poly(dI-dC), and 3 μl of 10× binding buffer (100 mM Tris HCl, pH 7.5, 50% glycerol, 10 mM dithiothreitol, 10 mM EDTA) for 20 min at the room temperature, followed by 30 min at 4 °C. The following oligonucleotides sequences were used for probes: 5′-UTR GATA motifs, 5′-GGCAGCCTGGATATCCTCTCCTA-3′; GATA-mut motifs, 5′-GGCAGCCTGTTTAAGCTCTCCTA-3′; *flk-1/KDR* SP1 motifs, 5′-GGTGAGGGGCGGGGCTGGCCGC-3′; and *flk-1/KDR* SP1-mut motifs, 5′-GGTGAGGTTCGGTTCTGGTTGC-3′. To test the effect of antibodies on DNA-protein binding, the nuclear extracts were preincubated with monoclonal antibody to GATA-2 (we prepared from the antigen GATA-2 (amino acids 9–258)) or polyclonal antibody to p65 (Santa Cruz) for 30 min at the room temperature. DNA-protein complexes were resolved on a 5% nondenaturing polyacrylamide gel containing 5% glycerol in 0.5× TBE (50 mM Tris, 50 mM boric acid, and 1 mM EDTA). The loaded gel was fixed with 10% methanol and 10% acetic acid and then imaged by

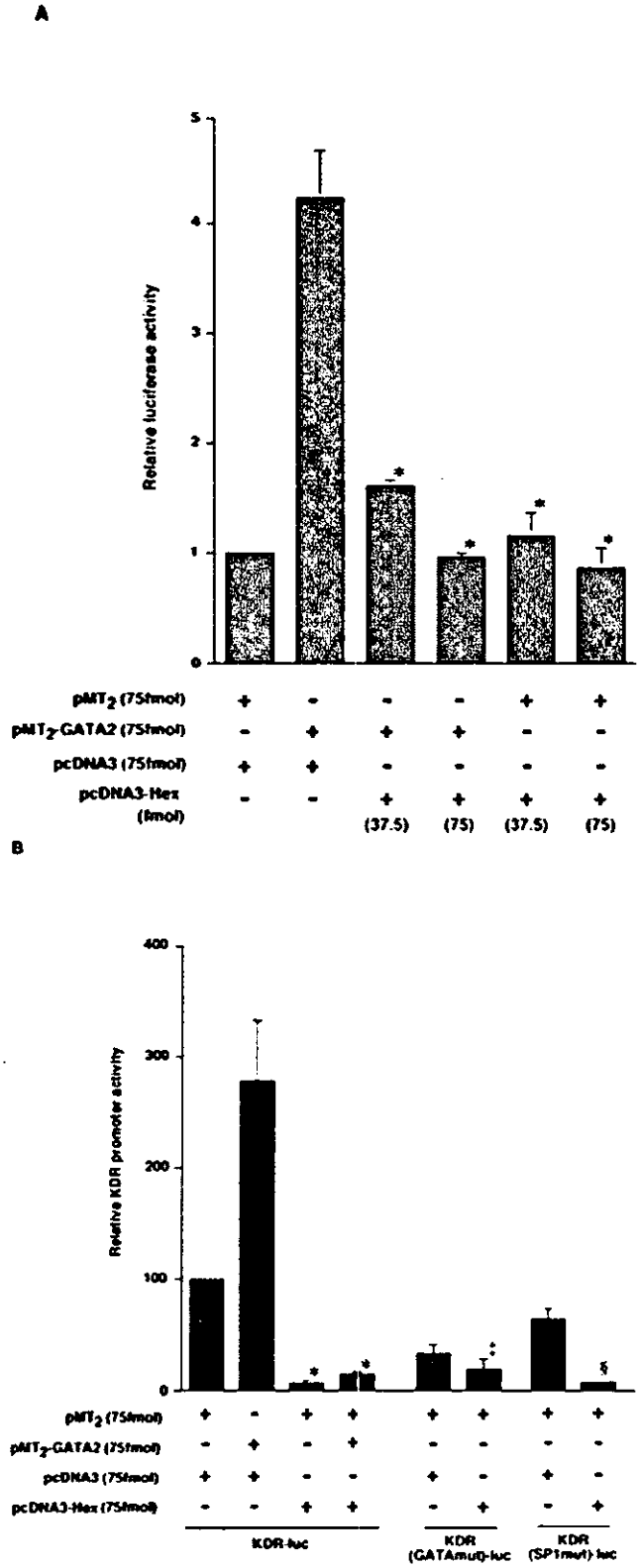



FIG. 3. **Hex suppresses GATA-2-mediated *flk-1/KDR* promoter activity.** *A*, HEK-293 cells were transiently transfected with luciferase reporter plasmid containing the *flk-1/KDR* promoter fragment (−115 to +296) (*KDR*-luc), either human GATA2 expression plasmid ($pMT_2$-GATA2) or vector alone ($pMT_2$), and either human Hex expression plasmid (pcDNA3-HEX) or vector alone (pcDNA3). After 24 h of incubation, the cells were lysed and assayed for luciferase activity. *, $p < 0.01$ compared with the activity from co-transfection with *KDR* promoter, $pMT_2$-GATA2, and pcDNA3. *B*, HUVEC were transiently transfected as above. After 48 h of incubation, the cells were lysed and assayed for luciferase activity. *, $p < 0.002$ compared with the activity from co-transfection with *KDR*-luc, $pMT_2$-GATA2, and pcDNA3; §, $p = 0.002$ compared from co-transfection with *KDR* (SP1 mut)-luc, $pMT_2$, and pcDNA3; ‡, $p = 0.08$ (not significant) compared from co-transfection with *KDR* (GATA mut)-luc, $pMT_2$, and pcDNA3. The expression levels were normalized to pRL-SV40 activity and expressed as fold induction relative to co-transfection with vector alone. The means and standard deviations were derived from at least three separate experiments performed in triplicate.

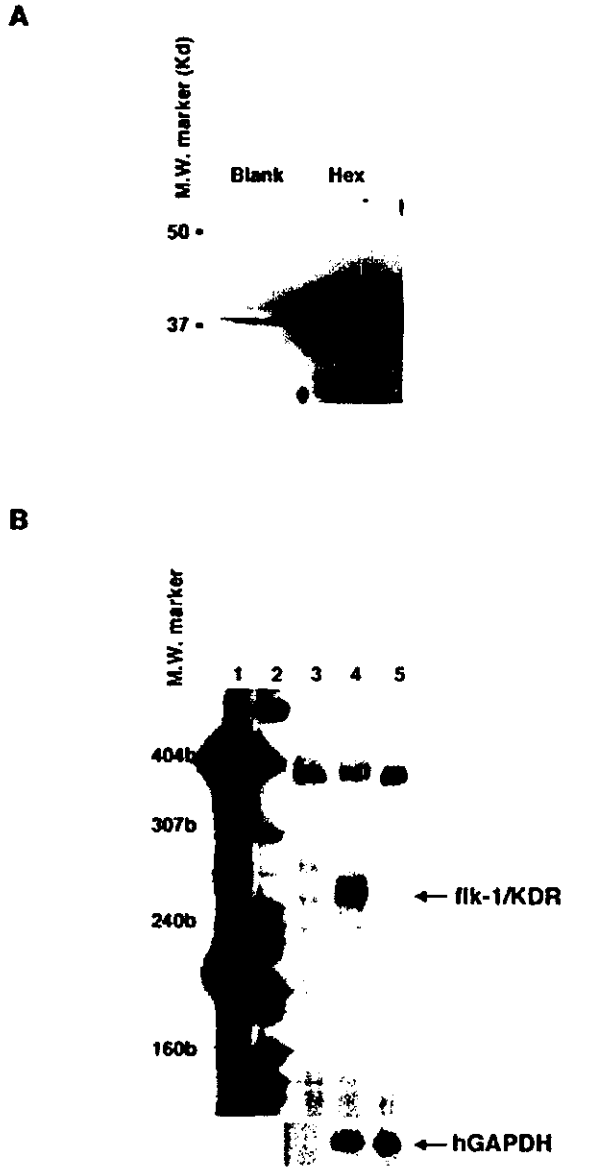



FIG. 4. **Hex suppresses endogenous *flk-1/KDR* mRNA in human endothelial cells.** *A*, HUVEC were infected with adenovirus expressing IRES-mediated EGFP (*Blank*) or IRES-mediated rat Hex and EGFP (HEX) at a multiplicity of infection of 20 plaque-forming units/cells and then incubated for 48 h. The cell extracts were separated by 10% SDS-PAGE and transferred to nitrocellulose membrane. The membrane was immunoblotted with anti-HEX antibody. *M.W.*, molecular mass. *B*, HUVEC were infected with adenovirus expressing IRES-mediated EGFP (*Blank*) or expressing IRES-mediated rat Hex and EGFP (*HEX*) at a multiplicity of infection of 20 plaque-forming units/cells for 48 h, at which time the total RNA was isolated. In RNase protection assays, an [$\alpha$-$^{32}$P]UTP-labeled 384-bp human *flk-1/KDR* riboprobe was incubated with no RNA (*lane 1*), 10 $\mu$g of yeast tRNA (*lane 3*), 10 $\mu$g of total RNA from control (*Blank*) HUVEC (*lane 4*), or Hex-overexpressed (*Hex*) HUVEC (*lane 5*). The protected fragment (266 bp) represents the human *flk-1/KDR* transcript. An [$\alpha$-$^{32}$P]UTP-labeled GAPDH riboprobe was hybridized with total RNA as an internal control.

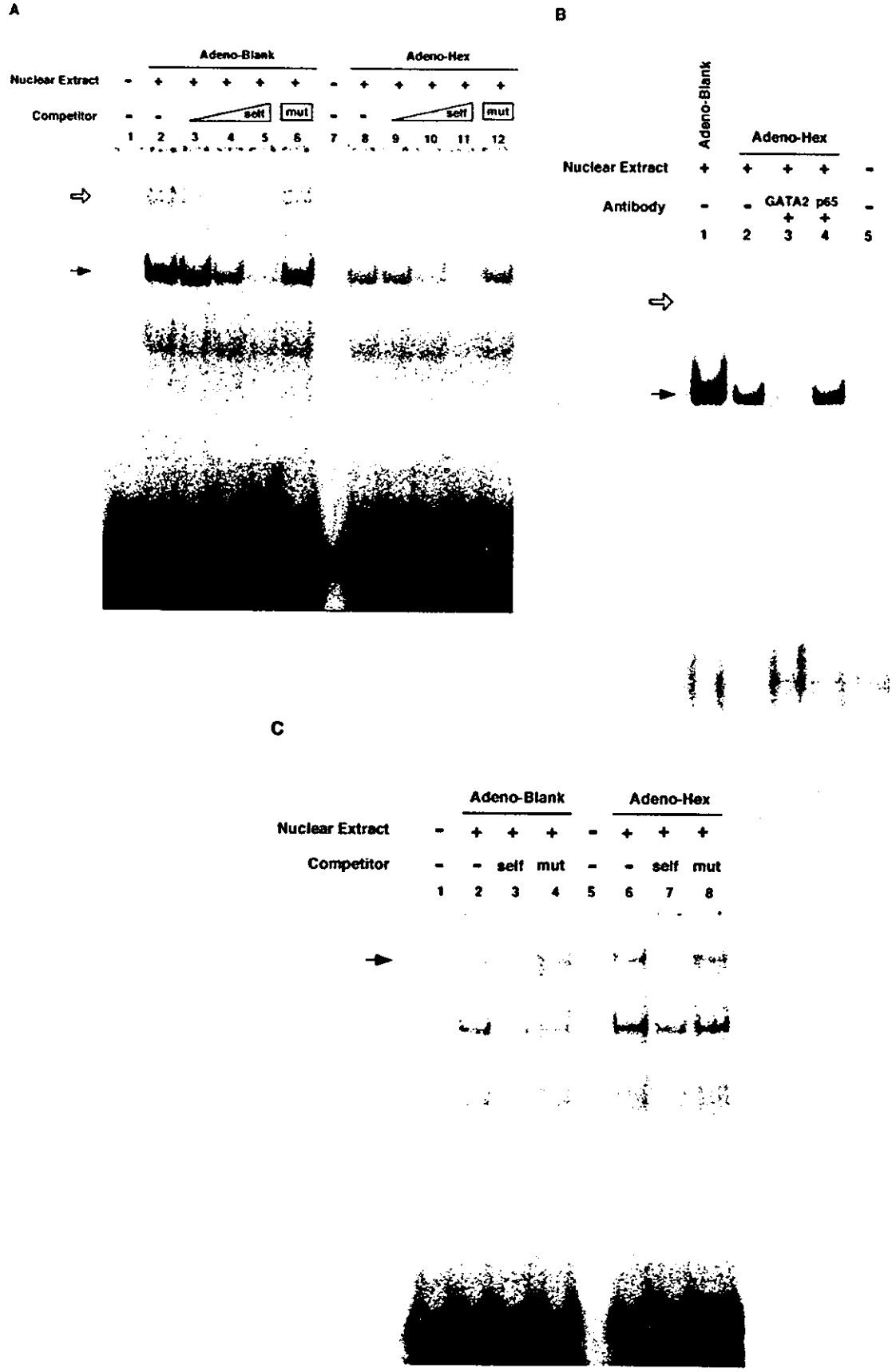



FIG. 5. **Hex reduces binding of GATA motif on the *flk-1/KDR* and GATA-binding proteins.** *A*, electrophoretic mobility shift assays were performed with $^{32}$P-labeled 5′-UTR GATA probe and 10 μg of nuclear extract from HUVEC infected with adenovirus expressing IRES-mediated EGFP (*Adeno-Blank*) or adenovirus expressing IRES-mediated Hex and EGFP (*Adeno-Hex*). The *open* and *closed arrows* indicate specific DNA-protein complexes. In competition assays, 10-, 100-, and 500-fold molar excesses of unlabeled self-probe and mutant 5′-UTR GATA probe

BAS-1800 (Fuji Film, Japan). The signals were quantified with NIH Image.

## RESULTS

*Identification of Hematopoietically Expressed Hex as a GATA-interacting Protein by Yeast Two-hybrid Screening*—The GATA family of transcription factors has been implicated not only in the early differentiation of endothelial cells but also in the transduction of extracellular signals in the adult endothelium. Several members of the GATA family have been identified in endothelial cells, including GATA-2, -3, and -6. Previous studies have shown that GATA-1 and -2 interact with partner proteins in erythroid and megakaryocyte cells (34). Our goal was to identify the proteins that interact with GATA transcription factors in endothelial cells. To that end, we employed a yeast two-hybrid system in which full-length human GATA-6 served as bait. From a screen of $8.6 \times 10^6$ clones, a total of 58 clones demonstrated histidine auxotroph. After elimination of false positives, 16 clones were selected, one of which encoded a fragment (amino acids 75–270) of Hex (Fig. 1*A*). To confirm the specificity of interaction between Hex and GATA-6, constructs containing the Gal4-activating domain fused with full-length Hex (pGAD-Hex) and either the Gal4-binding domain fused with GATA-6 (pGBD-GATA6) or the Gal4-binding domain alone (pGBD) were co-expressed in yeast AH109. As shown in Fig. 1*B*, co-expression of pGAD-Hex and pGBD-GATA6 resulted in a 4.2-fold increase in β-galactosidase activity compared with co-expression of pGBD and pGAD-Hex.

We next wished to determine whether Hex interacts with human GATA-2 and -3. To that end, pGAD-Hex was co-expressed in yeast AH109 with constructs containing the Gal4-binding domain fused either with GATA-2 or -3 (pGBD-GATA2 or GATA3, respectively) As shown in Fig. 1*C*, co-expression with human GATA-2 and -3 resulted in 5.8- and 4.7-fold induction of the β-galactosidase activity, respectively, compared with co-expression with Gal4-binding domain alone (pGBD). Taken together, these results suggest that Hex interacts with GATA-2, -3, and -6.

*Physical Interaction between Hex and GATA-2 in Mammalian Cells*—Having identified the interaction between GATA and Hex in the yeast two-hybrid system, we wished to determine whether this interaction occurs in mammalian cells. Of the various members of the GATA family of transcription factors, GATA-2 is expressed most abundantly in cultured endothelial cells (data not shown) and is believed to play a predominant role in endothelial cell biology. Therefore, we chose to focus on the interaction between Hex and GATA-2. To that end, COS-7 cells were transiently transfected with expression plasmids for human GATA-2 ($pMT_2$-GATA2), FLAG-tagged human GATA-2 (pFLAG-GATA2), FLAG- or Myc-tagged human Hex (pFLAG-Hex or pMyc-Hex), or vector alone ($pMT_2$, pMyc or pFLAG) and then processed for immunoprecipitation. The transfected cells expressed high levels of GATA-2 and Hex (Fig. 2, *A* and *B*, *lanes 1–3*). In co-transfections ($pMT_2$-GATA2 and pFLAG-Hex), GATA-2 co-precipitated with the anti-FLAG antibody (Fig. 2*A*, *lane 6*, *closed arrowhead*) but not with the nonimmune control (Fig. 2*A*, *lane 7*). In contrast, in experiments in which either ($pMT_2$-GATA2 and pFLAG) or ($pMT_2$ and pFLAG-tagged Hex) were transfected, GATA-2 did not co-precipitate (Fig. 2*A*, *lanes 4* and *5*). Furthermore, in co-transfections (pFLAG-GATA2 and pMyc-Hex), Myc-tagged Hex co-precipitated with the anti-FLAG antibody (Fig. 2*B*, *lane 6*, *closed arrowhead*). Although in experiments in which either (pFLAG and pMyc-Hex) or (pFLAG-GATA2 and pMyc) were transfected, Hex did not co-precipitate (Fig. 2*B*, *lanes 4* and *5*).

Next, we wished to identify the physical interaction between endogenous GATA-2 and Hex in endothelial cells. To that end, nuclear extracts were prepared and processed for immunoblotting and co-immunoprecipitations. As shown in Fig. 2*C*, endogenous GATA-2 and Hex were detected in nuclear extracts from HUVEC but not COS-7. Endogenous Hex was co-precipitated with anti-GATA2 antibody but not with isotype-matched control IgG (Fig. 2*D*, *lanes 2* and *3*). In semi-quantitative calculations from two independent experiments, anti-GATA-2 antibody immunoprecipitated 75.2% of total GATA-2 in HUVEC. Moreover, 8.29% of total cellular Hex was physically associated with GATA-2. Collectively, these findings suggest that GATA-2 and Hex specifically interact with one another in cultured endothelial cells.

*Hex Inhibits GATA-2-mediated flk-1/KDR Promoter Activity*—We next wished to study the functional relevance of the interaction between GATA-2 and Hex. We have previously shown that GATA-2 binds to a GATA motif in the 5′-UTR region of the *flk-1/KDR* promoter and that this effect is necessary for full expression (35). To determine the effect of Hex on GATA-2-mediated activation of *flk-1/KDR*, transactivation assays were carried out in HEK-293 cells (nonendothelial cells) and HUVEC (endothelial cells) co-transfected with *KDR*-luc (−115 and +296 *flk-1/KDR* coupled to luciferase) and an expression plasmid containing either human GATA-2 ($pMT_2$-GATA2) or Hex (pcDNA3-HEX). As a negative control, the cells were co-transfected with vector alone ($pMT_2$ or pcDNA3). Consistent with our previous findings, the basal level of *flk-1/KDR* promoter activity in HEK-293 cells was significantly transactivated by 4.2-fold with overexpression of GATA-2 (Fig. 3*A*). GATA-2-mediated stimulation of promoter activity was completely abrogated by co-expression of Hex in a dose-dependent manner. Moreover, overexpression of Hex did not change the basal level of the promoter activity (Fig. 3*A*). In HUVEC, a high level of *flk-1/KDR* promoter activity (8.7-fold higher than SV40 promoter plus enhancer construct (pGL2-control), not shown) occurred, and overexpression of GATA-2 resulted in 2.8-fold transactivation of the promoter (Fig. 3*B*). More importantly, overexpression of Hex resulted in huge reduction of the *flk-1/KDR* promoter activity (15.9-fold compared without expression of GATA-2 and Hex). Co-expression of GATA-2 failed to recover the Hex-mediated attenuation of the *flk-1/KDR* promoter activity (2.3-fold induction compared Hex expression alone) (Fig. 3*B*). To confirm that the Hex-mediated down-regulation of the promoter activity was mediated by the 5′-UTR GATA motif on the *flk-1/KDR* gene, either the GATA element or the SP1 element (as a control) point-mutated plasmid was transfected into HUVEC. *KDR* promoter activity from SP1 point-mutated plasmid was down to 63.1% and markedly reduced to 3.4% by the co-expression with Hex. In contrast, the promoter activity from GATA point-mutated plasmid was down to 33.1%, whereas there was no significant reduction in the presence of Hex (Fig. 3*B*). Previous studies have shown that Hex directly binds to a

(*lanes 3* and *9*, *lanes 4* and *10*, *lanes 5* and *11*, and *lanes 6* and *12*, respectively) were added to the reaction mixture. *B*, electrophoretic mobility shift assays were performed with 10 μg of nuclear extract from HUVEC infected with adenovirus expressing IRES-mediated EGFP (*Adeno-Blank*) (*lane 1*) or adenovirus-expressing IRES-mediated Hex and EGFP (*Adeno-Hex*) (*lanes 2–4*). To test the effect of antibodies on the DNA-protein complexes, antibody against GATA-2 (*lane 3*) or against p65 (*lane 4*) was added to the reaction mixture. *C*, electrophoretic mobility shift assays were performed with $^{32}$P-labeled *flk-1/KDR* SP1 probe and 10 μg of nuclear extract from HUVEC infected as above. The arrows indicate specific DNA-protein complexes. In competition assays, a 100-fold molar excess of unlabeled self-probe or mutant SP1 probe (*lanes 2* and *5* and *lanes 3* and *6*, respectively) was added to the reaction mixture. The results are representative of two independent experiments.

consensus element (5′-CAATTAAA-3′) in the promoter region of its target genes, resulting in transcriptional activation (36). A search of the 4-kb 5′-flanking region and 3-kb first intron of the *flk-1/KDR* failed to reveal such a motif. Taken together, these results suggest that Hex represses the GATA-mediated *flk-1/KDR* promoter activation.

*Hex Suppresses flk-1/KDR mRNA Expression in Primary Human Endothelial Cells*—To determine whether Hex modulates the expression of the endogenous *flk-1/KDR* gene, HUVEC were infected with adenovirus expressing either IRES-mediated-EGFP (Blank) or IRES-mediated-rat Hex and EGFP (Hex) at a multiplicity of infection of 20. Using this approach, over 80% of the cells were infected as determined by EGFP expression. Western blot assays of infected cells demonstrated high levels of Hex protein (Fig. 4*A*). More importantly, overexpression of Hex in HUVEC resulted in a 85% reduction (mean of three independent experiments) in *flk-1/KDR* mRNA by RNase protection assay (Fig. 4*B*, compare *lanes 4* and *5*). These findings indicate that Hex suppresses *flk-1/KDR* activity not only at the level of the promoter but also at the level of the endogenous gene.

*Hex Inhibits Binding of GATA-2 to the flk-1/KDR 5′-UTR GATA Motif*—Based on the above findings, we hypothesized that Hex inhibits *flk-1/KDR* expression by interfering with GATA binding to the 5′-UTR. To test this hypothesis, we performed electrophoretic mobility shift assays in which nuclear extracts derived from HUVEC either expressing IRES-mediated EGFP (Adeno-Blank) or IRES-mediated Hex and EGFP (Adeno-Hex) were incubated with a radiolabeled probe spanning the 5′-UTR GATA motif (+98 to +122). As shown in Fig. 5 (*A* and *B*), incubation of nuclear extract from Adeno-Blank infected HUVEC with the $^{32}$P-labeled probe resulted in the appearance of two specific DNA-protein complexes (*closed* and *open arrows*). These DNA-protein complexes were inhibited by the addition of 10-, 100-, and 500-fold molar excess unlabeled self-competitor (Fig. 5*A*, *lanes 3–5*) but not by 500-fold molar excess GATA mutant competitor (Fig. 5*A*, *lane 6*). The mobility shift pattern was identical in Hex-overexpressing cells (the more slowly migrated complexes designated with the *open arrows* appeared with longer exposure time). However, Hex overexpression resulted in a significant reduction (54.2% reduction by densitometry) in the intensity of the GATA-binding complexes (Fig. 5, *A*, compare *lanes 3–5* and *9–11*, and *B*, compare *lanes 1* and *2*). As previously reported, the DNA-protein complexes were inhibited by preincubation with the anti-GATA-2 antibody (34). In contrast, the complex was not inhibited by preincubation with anti-p65 antibody (Fig. 5*B*, *lanes 3* and *4*). Compared with control HUVEC, Hex overexpression did not significantly alter SP1 DNA binding activity (Fig. 5*C*). Together, these results suggest that Hex-mediated down-regulation of the *flk-1/KDR* is associated with an inhibition of GATA-2 binding.

*Hex Suppresses VEGF-mediated Tube Formation in Primary Human Endothelial Cell Cultures*—VEGF interaction with Flk-1/KDR is a critical determinant of endothelial cell proliferation and angiogenesis (37). In the next set of experiments, we wished to determine whether Hex inhibits the VEGF-Flk-1/KDR signaling axis. To that end, we carried out sandwich tube formation assays on collagen gel. In the absence of basic fibroblast growth factor, VEGF stimulated tube formation of HUVEC infected with adenovirus expressing IRES-mediated-EGFP (Adeno-Blank), an effect that was 50.8% inhibited by the preincubation with the Flk-1/KDR inhibitor, SU1498 (Fig. 6). Interestingly, VEGF-mediated tube formation was significantly (58.5%) inhibited by overexpression of Hex and profoundly (77.0%) abrogated when combined with the presence of

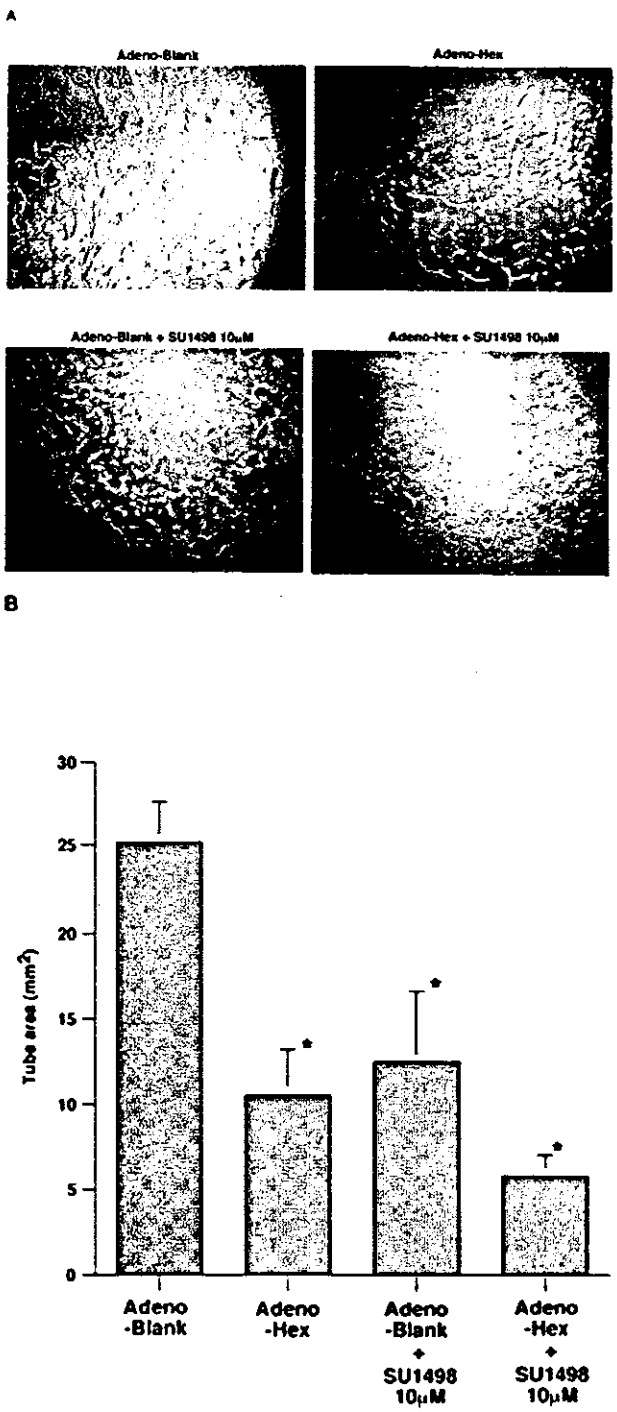



FIG. 6. **Hex inhibits the Flk-1/KDR-dependent tube formation activities in human endothelial cells.** *A*, sandwich tube formation model was performed as described under "Experimental Procedures." IRES-mediated EGFP expressing adenovirus (*Adeno-Blank*) or IRES-mediated Hex and EGFP expressing adenovirus (*Adeno-Hex*) were infected to HUVEC, after 24 h HUVEC were seeded on the collagen gel in the presence or absence of *flk-1* tyrosine kinase inhibitor SU1498 (10 μM). Tube formations were observed by phase contrast microscopy at 40-fold optical magnification. The results were representative of four independent experiments. *B*, shown in the quantification of the tube area. *, $p < 0.03$ compared with the activity from HUVEC infected with Adeno-Blank. The means and standard deviations were calculated with NIH Image from four independent experiments.

SU1498 (Adeno-Hex + SU1498) (Fig. 6). These findings suggest that Hex-mediated inhibition of *flk-1/KDR* gene expression results in a down-regulation of VEGF signaling.

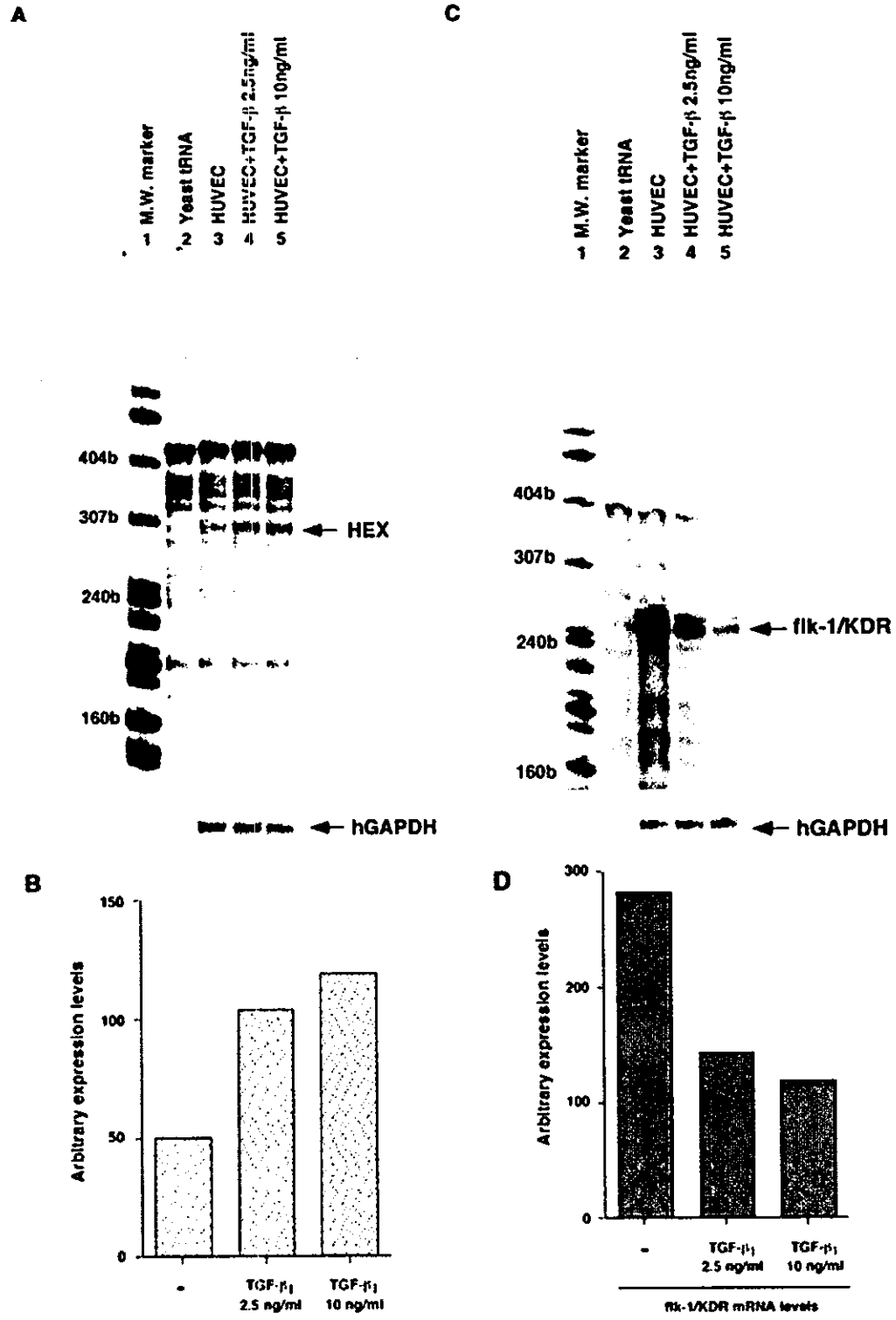



FIG. 7. **TGF-$\beta_1$ down-regulates endogenous *flk-1/KDR* and up-regulates endogenous Hex expression in human endothelial cells.** *A*, a confluent HUVEC were serum-starved and then incubated in the absence or presence of 2.5 or 10 ng/ml TGF-$\beta_1$ for 18 h, at which time total RNA was isolated. In RNase protection assays, an [$\alpha$-$^{32}$P]UTP-labeled 414-bp human Hex riboprobe was incubated with 10 $\mu$g of yeast tRNA (*lane 2*), 10 $\mu$g of total RNA from control HUVEC (*lane 3*), 2.5 ng/ml TGF-$\beta_1$-treated HUVEC (*lane 4*), or 10 ng/ml TGF-$\beta_1$-treated HUVEC (*lane 5*). The protected fragment (296 bp) represents the human Hex transcript. An [$\alpha$-$^{32}$P]UTP-labeled GAPDH riboprobe was hybridized with total RNA as an internal control. *B*, shown is the quantification of RNase protection assays. Densitometry was used to calculate the ratio of Hex and GAPDH signals (arbitrary expression level). The data represent the averages from two independent experiments. *C*, a confluent HUVEC were serum-starved and then incubated in the absence or presence of 2.5 or 10 ng/ml TGF-$\beta_1$ for 24 h, at which time total RNA was isolated. In RNase protection assays, an [$\alpha$-$^{32}$P]UTP-labeled 393-bp human *flk-1/KDR* riboprobe was incubated with 10 $\mu$g of yeast tRNA (*lane 2*), 10 $\mu$g of total RNA from control HUVEC (*lane 3*), 2.5 ng/ml TGF-$\beta_1$-treated HUVEC (*lane 4*), or 10 ng/ml TGF-$\beta_1$-treated HUVEC (*lane 5*). The protected fragment (266 bp) represents the human *flk-1/KDR* transcript. An [$\alpha$-$^{32}$P]UTP-labeled GAPDH riboprobe was hybridized with total RNA as an internal control. *M.W.*, molecular mass. *D*, shown is the quantification of RNase protection assays. Densitometry was used to calculate the ratio of *flk-1/KDR* and GAPDH signals (arbitrary expression level). The data represent the averages from two independent experiments.

*Hex Induces TGF-β Expression in Primary Human Endothelial Cells*—Previous studies have shown that whereas tumor necrosis factor α, VEGF, and thrombin induce *flk-1/KDR* expression in endothelial cells (38–40), TGF-β1 has the opposite effect (41). In keeping with these findings, the addition of 2.5 and 10 ng/ml TGF-β1 to HUVEC for 24 h resulted in 49 and 58% reduction of *flk-1/KDR* mRNA, respectively (Fig. 7, *A*, *lanes 3–5*, and *B*). Interestingly, incubation of HUVEC with 2.5 and 10 ng/ml TGF-β1 for 18 h resulted in 2.1- and 2.4-fold stimulation of Hex mRNA, respectively. Taken together, these results suggest that TGF-β1 may exert its inhibitory effects through a Hex-GATA-2-dependent pathway.

## DISCUSSION

The GATA family of transcription factors has been implicated not only in the early differentiation of the endothelial cells but also in the transduction of extracellular signals. For example, insulin-like growth factor 1, tumor necrosis factor α, and thrombin have each been shown to induce GATA-2 activity, whereas estrogens and TGF-β may actually inhibit GATA binding (6, 30, 35, 42). Together, these studies suggest that GATA transcription factors may behave like immediate early genes, serving to couple short changes in the extracellular environment to long term changes in downstream gene expression.

An important clue to understanding the multifaceted role of a transcription factor is found in its repertoire of protein-protein interactions. Indeed, previous studies have uncovered an array of partner proteins that interact with GATA transcription factors to regulate gene transcription (8, 14–18, 20, 22, 23, 25, 27). In the present study, we have extended the list of protein partners by demonstrating a novel interaction between GATA proteins and the homeobox protein Hex in primary human endothelial cells. Hex is a member of the homeobox transcription factors and has been shown to bind to a consensus motif (5′-CAATTAAA-3′) in the promoter region of downstream target genes, resulting in repression of gene transcription (43). In addition, Hex may indirectly modulate gene expression through protein-protein interactions. For example, Hex has been shown to associate with cAMP-responsive element-binding protein to induce the SMemb/nonmuscle myosin heavy chain B gene in vascular smooth muscle cells (44). Moreover, Hex-Jun protein interactions have been reported to suppress c-Jun, JunB, and JunD-mediated gene activation (45).

In the present study, we have demonstrated GATA-2, -3, and -6 that have been identified in endothelial cells were physically interacted with Hex in cultured cells by immunoprecipitation analysis. GATA-2, -3, and -6 have highly conserved zinc finger domain and the region involved the usual protein-protein interactions (2, 10, 13, 15, 18, 19, 20, 29, 37, 39). Further studies will be required to determine whether these motifs are responsible for mediating the binding of Hex to GATA-2, GATA-3, and GATA-6.

In a previous study, we demonstrated that TGF-β inhibits *flk-1/KDR* expression through a mechanism that involves reduced binding of GATA-2 to a palindromic GATA site in the 5′-UTR (35). In keeping with the role for GATA-2 in mediating *flk-1/KDR* expression, Hex overexpression resulted in reduced binding of GATA-2 to 5′-UTR GATA motif and a dose-dependent inhibition of *flk-1/KDR* promoter activity. It is interesting to speculate that the specific association between Hex and GATA-2 interferes with zinc finger domain-mediated DNA binding activity. Adenovirally mediated overexpression of Hex did not completely abrogate (54.2% inhibition) GATA-2 binding activity but did result in a more profound reduction of *flk-1/KDR* mRNA and the promoter activity (Figs. 3 and 4). It has been shown that Hex and Jun interaction inhibits Jun-mediated transactivation without interfering with the Jun-DNA binding activity (45). In addition, Hex contains the transrepression domain at the N terminus. These observations suggest that Hex may exert its inhibitory effect through two mechanisms, namely through competitive inhibition of GATA-2-DNA to the *flk-1/KDR* promoter and direct repression of transcription.

The results of the present study provide new insight into how TGF-β mediates its inhibitory effect on the GATA-2-*flk-1/KDR* signaling axis. The observation that TGF-β treatment of endothelial cells resulted in concomitant up-regulation of Hex mRNA and down-regulation of *flk-1/KDR* in endothelial cells suggests that TGF-β signaling promotes transcriptionally inactive or inhibitory complexes between GATA-2 and Hex. The functional consequence of this interaction was borne out in *in vitro* studies of angiogenesis, in which the overexpression of Hex abrogated VEGF-*flk-1/KDR*-dependent endothelial cell tube formation.

The mechanism by which TGF-β1 induces Hex expression remains to be elucidated. TGF-β1 is known to induce Smad-2 and -5 activity in endothelial cells (46). Moreover, Smad-mediated signaling has been implicated in the control of Hex expression (47). Mice that are null for Smad-2 deficiency are embryonic lethal at E7.5 and lack detectable levels of Hex (48). Taken together, these observations raise the possibility that TGF-β1 signaling is coupled to Hex-GATA-mediated inhibition of *flk-1/KDR* through a Smad-2-dependent pathway.

Most recently, Nakagawa *et al.* (49) reported that Hex acts as a negative regulator of angiogenesis. In the latter study, Hex was shown to completely abrogate the VEGF-mediated proliferation, migration, and invasion of HUVEC (49). Although our data are consistent with those of Nakagawa *et al.*, they are novel in that they: 1) provide a link between a natural inhibitor of angiogenesis (TGF-β) and Hex and 2) reveal a mechanistic connection between Hex, GATA-2, and *flk-1/KDR* expression. Indeed, based on these findings, we propose that Hex, as well as GATA-2, represent new therapeutic targets for anti-angiogenesis therapy.

*Acknowledgments*—We are grateful to Stuart. H. Orkin (Harvard Medical School, Boston, MA) for kindly providing the human GATA-2 expression plasmid (pMT$_2$-GATA2) and anti-GATA-2 antibody. We also thank Drs. Ryuichiro Sato and Juro Sakai (The University of Tokyo, Japan) for technical help with co-immunoprecipitation assays.

## REFERENCES

1. Tjian, R., and Maniatis, T. (1994) *Cell* **77,** 5–8
2. Umetani, M., Mataki, C., Minegishi, N., Yamamoto, M., Hamakubo, T., and Kodama, T. (2001) *Arterioscler. Thromb. Vasc. Biol.* **21,** 917–922
3. Charron, F., and Nemer, M. (1999) *Semin. Cell Dev. Biol.* **10,** 85–91
4. Orkin, S. H. (1995) *Curr. Opin. Cell Biol.* **7,** 870–877
5. Weiss, M. J., and Orkin, S. H. (1995) *Exp. Hematol.* **23,** 99–107
6. Musaro, A., McCullagh, K. J., Naya, F. J., Olson, E. N., and Rosenthal, N. (1999) *Nature* **400,** 581–585
7. Minami, T., Abid, M. R., Zhang, J., King, G., Kodama, T., and Aird, W. C. (2003) *J. Biol. Chem.* **278,** 6976–6984
8. Gregory, R. C., Taxman, D. J., Seshasayee, D., Kensinger, M. H., Bieker, J. J., and Wojchowski, D. M. (1996) *Blood* **87,** 1793–1801
9. Osada, H., Grutz, G. G., Axelson, H., Forster, A., and Rabbitts, T. H. (1997) *Leukemia* **11,** (Suppl. 3) 307–312
10. Wadman, I. A., Osada, H., Grutz, G. G., Agulnick, A. D., Westphal, H., Forster, A., and Rabbitts, T. H. (1997) *EMBO J.* **16,** 3145–3157
11. Blobel, G. A., Sieff, C. A., and Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 3147–3153
12. Crossley, M., Merika, M., and Orkin, S. H. (1995) *Mol. Cell. Biol.* **15,** 2448–2456
13. Mackay, J. P., Kowalski, K., Fox, A. H., Czolij, R., King, G. F., and Crossley, M. (1998) *J. Biol. Chem.* **273,** 30560–30567
14. van Wering, H. M., Huibregtse, I. L., van der Zwan, S. M., de Bie, M. S., Dowling, L. N., Boudreau, F., Rings, E. H., Grand, R. J., and Krasinski, S. D. (2002) *J. Biol. Chem.* **277,** 27659–27667
15. Lee, Y., Shioi, T., Kasahara, H., Jobe, S. M., Wiese, R. J., Markham, B. E., and Izumo, S. (1998) *Mol. Cell. Biol.* **18,** 3120–3129
16. Gordon, D. F., Lewis, S. R., Haugen, B. R., James, R. A., McDermott, M. T., Wood, W. M., and Ridgway, E. C. (1997) *J. Biol. Chem.* **272,** 24339–24347
17. Liu, C., Glasser, S. W., Wan, H., and Whitsett, J. A. (2002) *J. Biol. Chem.* **277,** 4519–4525
18. Rekhtman, N., Radparvar, F., Evans, T., and Skoultchi, A. I. (1999) *Genes Dev.*

**13,** 1398–1411
19. Morin, S., Paradis, P., Aries, A., and Nemer, M. (2001) *Mol. Cell. Biol.* **21,** 1036–1044
20. Elagib, K. E., Racke, F. K., Mogass, M., Khetawat, R., Delehanty, L. L., and Goldfarb, A. N. (2003) *Blood* **101,** 4313–4341
21. Kawana, M., Lee, M. E., Quertermous, E. E., and Quertermous, T. (1995) *Mol. Cell. Biol.* **15,** 4225–4231
22. Blokzijl, A., ten Dijke, P., and Ibanez, C. F. (2002) *Curr. Biol.* **12,** 35–45
23. Tsang, A. P., Visvader, J. E., Turner, C. A., Fujiwara, Y., Yu, C., Weiss, M. J., Crossley, M., and Orkin, S. H. (1997) *Cell* **90,** 109–119
24. Tevosian, S. G., Deconinck, A. E., Cantor, A. B., Rieff, H. I., Fujiwara, Y., Corfas, G., and Orkin, S. H. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 950–955
25. Svensson, E. C., Tufts, R. L., Polk, C. E., and Leiden, J. M. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 956–961
26. Chang, A. N., Cantor, A. B., Fujiwara, Y., Lodish, M. B., Droho, S., Crispino, J. D., and Orkin, S. H. (2002) *Proc. Natl. Acad. Sci. U. S. A.* **99,** 9237–9242
27. Blobel, G. A., Nakajima, T., Eckner, R., Montminy, M., and Orkin, S. H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95,** 2061–2066
28. Ozawa, Y., Towatari, M., Tsuzuki, S., Hayakawa, F., Maeda, T., Miyata, Y., Tanimoto, M., and Saito, H. (2001) *Blood* **98,** 2116–2123
29. Aird, W. C., Parvin, J. D., Sharp, P. A., and Rosenberg, R. D. (1994) *J. Biol. Chem.* **269,** 883–889
30. Minami, T., and Aird, W. C. (2001) *J. Biol. Chem.* **276,** 47632–47641
31. Brock, T. G., Paine, R., 3rd, and Peters-Golden, M. (1994) *J. Biol. Chem.* **269,** 22059–22066
32. Chiba, T., Kogishi, K., Wang, J., Xia, C., Matsushita, T., Miyazaki, J., Saito, I., Hosokawa, M., and Higuchi, K. (1999) *Am. J. Pathol.* **155,** 1319–1326
33. Dignam, J. D., Lebovitz, R. M., and Roeder, R. G. (1983) *Nucleic Acids Res.* **11,** 1475–1489
34. Cantor, A. B., and Orkin, S. H. (2002) *Oncogene* **21,** 3368–3376
35. Minami, T., Rosenberg, R. D., and Aird, W. C. (2001) *J. Biol. Chem.* **276,** 5395–5402
36. Crompton, M. R., Bartlett, T. J., MacGregor, A. D., Manfioletti, G., Buratti, E., Giancotti, V., and Goodwin, G. H. (1992) *Nucleic Acids Res.* **20,** 5661–5667
37. Shalaby, F., Rossant, J., Yamaguchi, T. P., Gertsenstein, M., Wu, X. F., Breitman, M. L., and Schuh, A. C. (1995) *Nature* **376,** 62–66
38. Giraudo, E., Primo, L., Audero, E., Gerber, H. P., Koolwijk, P., Soker, S., Klagsbrun, M., Ferrara, N., and Bussolino, F. (1998) *J. Biol. Chem.* **273,** 22128–22135
39. Wang, D., Donner, D. B., and Warren, R. S. (2000) *J. Biol. Chem.* **275,** 15905–15911
40. Wang, J., Morita, I., Onodera, M., and Murota, S. I. (2002) *J. Cell. Physiol.* **190,** 238–250
41. Pepper, M. S., Vassalli, J. D., Orci, L., and Montesano, R. (1993) *Exp. Cell Res.* **204,** 356–363
42. Simoncini, T., Maffei, S., Basta, G., Barsacchi, G., Genazzani, A. R., Liao, J. K., and De Caterina, R. (2000) *Circ. Res.* **87,** 19–25
43. Pellizzari, L., D'Elia, A., Rustighi, A., Manfioletti, G., Tell, G., and Damante, G. (2000) *Nucleic Acids Res.* **28,** 2503–2511
44. Sekiguchi, K., Kurabayashi, M., Oyama, Y., Aihara, Y., Tanaka, T., Sakamoto, H., Hoshino, Y., Kanda, T., Yokoyama, T., Shimomura, Y., Iijima, H., Ohyama, Y., and Nagai, R. (2001) *Circ. Res.* **88,** 52–58
45. Schaefer, L. K., Wang, S., and Schaefer, T. S. (2001) *J. Biol. Chem.* **276,** 43074–43082
46. Goumans, M. J., Valdimarsdottir, G., Itoh, S., Rosendahl, A., Sideras, P., and ten Dijke, P. (2002) *EMBO J.* **21,** 1743–1753
47. Zhang, W., Yatskievych, T. A., Cao, X., and Antin, P. B. (2002) *J. Biol. Chem.* **277,** 45435–45441
48. Heyer, J., Escalante-Alcalde, D., Lia, M., Boettinger, E., Edelmann, W., Stewart, C. L., and Kucherlapati, R. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 12595–12600
49. Nakagawa, T., Abe, M., Yamazaki, T., Miyashita, H., Niwa, H., Kokubun, S., and Sato, Y. (2003) *Arterioscler. Thromb. Vasc. Biol.* **23,** 231–237

*Physiol Genomics* 13: 31–46, 2003.
First published January 7, 2003; 10.1152/physiolgenomics.00116.2002.

# Expression imbalance map: a new visualization method for detection of mRNA expression imbalance regions

**Makoto Kano,[1] Kunihiro Nishimura,[2] Shumpei Ishikawa,[3] Shuichi Tsutsumi,[3] Koichi Hirota,[4] Michitaka Hirose,[4] and Hiroyuki Aburatani[3]**

*[1]School of Engineering and [2]School of Information Science and Technology, University of Tokyo, Tokyo 113-8655; and [3]Genome Science Division, and [4]Intelligent Cooperative System, Department of Information Systems, Research Center for Advanced Science and Technology, University of Tokyo, 153-8904, Japan*

Submitted 4 September 2002; accepted in final form 20 December 2002

**Kano, Makoto, Kunihiro Nishimura, Shumpei Ishikawa, Shuichi Tsutsumi, Koichi Hirota, Michitaka Hirose, and Hiroyuki Aburatani.** Expression imbalance map: a new visualization method for detection of mRNA expression imbalance regions. *Physiol Genomics* 13: 31–46, 2003. First published January 7, 2003; 10.1152/physiolgenomics.00116.2002.—We describe the development of a new visualization method, called the expression imbalance map (EIM), for detecting mRNA expression imbalance regions, reflecting genomic losses and gains at a much higher resolution than conventional technologies such as comparative genomic hybridization (CGH). Simple spatial mapping of the microarray expression profiles on chromosomal location provides little information about genomic structure, because mRNA expression levels do not completely reflect genomic copy number and some microarray probes would be of low quality. The EIM, which does not employ arbitrary selection of thresholds in conjunction with hypergeometric distribution-based algorithm, has a high tolerance of these complex factors. The EIM could detect regionally underexpressed or overexpressed genes (called, here, an expression imbalance region) in lung cancer specimens from their gene expression data of oligonucleotide microarray. Many known as well as potential loci with frequent genomic losses or gains were detected as expression imbalance regions by the EIM. Therefore, the EIM should provide the user with further insight into genomic structure through mRNA expression.

gene expression profiling; allelic imbalance; chromosome mapping; hypergeometric distribution; computing methodologies

THE RECENT DEVELOPMENT of microarray technology has enabled simultaneous measurement of genome-wide expression profiles. Many research studies have revealed strong correlations between the expression profiles and cancer classifications. The next era of gene expression analysis would involve systematic integration of expression profiles and other types of gene information, such as locus, gene function, and sequence information. In particular, integration between expression profiles and locus information should be effective in detecting gene structural abnormalities such as genomic gains and losses.

In general, cancer progression is not a single but a multistep process and includes many genomic structural abnormalities. Among them, genomic gains and losses, particularly deletion of tumor suppressor genes and amplification of oncogenes, are associated with cancer progression and its malignant phenotype, although the affected lesion varies among different types of cancers. Comparative genomic hybridization (CGH) for detecting genome-wide abnormalities such as copy number changes, has been applied to various types of cancers (5), but its low resolution (~20 Mb, corresponding to about 200 genes) makes it difficult to identify the causal genes, the structural alternation of which is critical for cancer biological behavior.

Integration of gene expression profiles and gene locus information might allow detection of copy number changes at a much higher resolution. Several studies using oligonucleotide probe arrays suggested a strong relationship between genomic structural abnormalities and expression imbalances (underexpression or overexpression). Mukasa et al. (7) reported that the expression levels of a significant number of genes in the 1p region were reduced to about 50%, in oligodendrogliomas with 1pLOH. Furthermore, Virtaneva et al. (12) reported that acute myeloid leukemia with trisomy 8 was associated with overexpression of genes on chromosome 8. Recently, a genome-wide transcriptome map of non-small cell lung carcinomas based on gene expression profiles generated by serial analysis of gene expression (SAGE) was conducted (3). However, the simple spatial mapping of the expression profiles on chromosomal location sometimes hardly provides information about genomic structure for the following reasons: *1*) since some microarray probes are of low quality, the microarray signal intensities do not always reflect their target mRNA expression levels; and 2) mRNA expression level does not completely reflect genomic copy number. The aim of the present study was to develop a new method with high tolerance of such complex factors, designed to detect regionally underexpressed or overexpressed genes in cancer specimens compared with the corresponding normal tissues. The expression imbalance region, constituted by

Article published online before print. See web site for date of publication (http://physiolgenomics.physiology.org).

Address for reprint requests and other correspondence: M. Kano, Tokyo Research Laboratory, IBM Japan, 1623-14 Shimotsuruma, Yamato-shi, Kanawaga 242-8502, Japan (E-mail: mkano@jp.ibm.com).

1094-8341/03 $5.00 Copyright © 2003 the American Physiological Society





Physiological Genomics

Downloaded from physiolgenomics.physiology.org on May 11, 2005

these genes, likely reflects genomic structural changes such as chromosomal gain and loss.

When developing the methodology that integrates the expression profiles and locus information, two significant problems have to be dealt with. First, a definition of what constitutes an expression imbalance region is not yet clarified. How many base pairs on chromosome should be considered as a genomic region (referred to below as chromosomal proximity)? To consider that a certain gene is differentially expressed in cancer and normal tissue, how much difference in the gene expression level is needed between the two (referred to below as cancer specificity)? It is generally very difficult to determine adequate thresholds for chromosomal proximity and cancer specificity. Arbitrary selection of thresholds would involve a risk of overlooking significant genes (that is, "threshold problem"). In addition, to detect expression imbalance regions, it is necessary to search for genes with both cancer specificity and chromosomal proximity. Because determining these two thresholds synergistically increases the risk of overlooking significant genes, the "threshold problem" is more critical in this case.

When selecting thresholds, several statistical theories such as hypothesis testing are helpful. However, commonly used statistical criteria are also arbitrarily determined. If thresholds are automatically determined based on statistical theory, the user cannot search more genes with potential significance, because the information of genes overlooked is almost unknown. Therefore, to detect as many significant genes as possible, a comprehensive presentation of the distribution of the "false balance" (that is, the balance of false negative and false positive) is quite significant rather than an attempt to seek potentially optimal statistical criterion.

Second, there are many candidate expression imbalance regions. Some of them may be a family of genes that are tandemly repeated and are under similar transcriptional regulations. To confirm that a candidate locus is biologically significant, human curation is necessary, using a variety of biological information. Therefore, it is important to present large genome-wide data in a comprehensive manner, indicating which genes are to be further examined. That is, a broadband interface between humans and computers is essential.

We focused on visualization technology as the key technology to solve these two problems. Visualization is effective in providing, genome-wide, the false-balance distribution and indication of the genes that are worth examining. The visualization used in our report would make it possible to present the images of all genes that have both cancer specificity and chromosomal proximity.

In this study, we developed a novel visualization method for detecting expression imbalance regions at much higher resolution than conventional technologies such as CGH, called the expression imbalance map (EIM). The EIM was applied to gene expression data of lung squamous cell carcinoma measured by oligonucleotide microarray and detected many known as well as potential loci with frequent genomic losses or gains as regional signal images on chromosomes (expression imbalance regions). In addition, the EIM could detect not only the expression imbalance common to all cancer specimens, but also individual differences among cancer specimens.

## MATERIAL AND METHODS

### *Data Sets*

In this article, the EIM is illustrated using the gene expression data of lung cancer from the study of Bhattacharjee et al. (1). In this experiment, total mRNA was extracted from histologically defined specimens of squamous cell lung carcinomas (abbreviated here as "SQ"; $n = 21$) and normal lung tissues (abbreviated here as "NL"; $n = 17$). The expression profiles were obtained using human U95A oligonucleotide probe arrays (GeneChip; Affymetrix, Santa Clara, CA). The SQ-NL gene expression data set (SQ, $n = 21$; NL, $n = 17$) was then analyzed using the EIM.

### *Feature Selection and Logarithmic Transformation*

To compensate for distortion in the expression level, changes in the expression level were limited from 1 to 8,000. In addition, 4,083 probes with a mean expression above 50 and CV (CV = mean/standard deviation) above 0.2 were selected to eliminate potential low-quality probes. The common logarithm of the gene expression data was used for the following analysis.

### *Translation from Probe to UniGene*

To associate gene locus information with gene expression profiles, each "probeID" on the U95A array was translated to UniGene, using information on the UniGene web site of the National Center for Biotechnology Information (NCBI), by referring to the corresponding original GenBank accession number of each probe set. Then, 11,334 of 12,533 probes on the U95A array were translated into 8,851 UniGenes.

### *Gene Locus Information*

Gene locus information was obtained from the web sites for Genes On Sequence Map (*Homo sapiens* build 27) of NCBI and is defined as "LocusID." Among the LocusIDs on chromosome 1 to 22 of Genes On Sequence Map, the 12,063 LocusIDs, which had the corresponding UniGenes, were utilized to identify the chromosome locations of genes. Since the gene expression data utilized in this study were obtained from both sexes, the X and Y chromosomes were excluded. However, by using the data obtained from only males or females, the EIM can be applied to the analysis of chromosome X and Y. Since the 12,063 LocusIDs had one-to-one correspondence with UniGenes, they were translated into 12,063 UniGenes. However, only 6,652 of the 12,063 UniGenes were in common with the 8,851 UniGenes translated from the probes on the U95A array (Fig. 1). In this article, these 6,652 UniGenes are called "Key-UniGenes." The distributions of the UniGenes and Key-UniGenes on each arm of the chromosome are shown in Table 1. The number of total Key-UniGenes was defined as $U$ (=6,652).

### *Quantization of Each Chromosome Arm Region*

For easier handling of the gene locus information, each chromosome arm region was quantized by unit region called



Physiological Genomics

Downloaded from physiolgenomics.physiology.org on May 11, 2005