Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7305–7309, October 1987
Medical Sciences

Genomic sequence of *hst*, a transforming gene encoding a protein homologous to fibroblast growth factors and the *int-2*-encoded protein

(oncogene/transformation/growth factor)

TERUHIKO YOSHIDA*, KIYOSHI MIYAGAWA*, HIROKI ODAGIRI*, HIROMI SAKAMOTO*, PETER F. R. LITTLE†, MASAAKI TERADA*, AND TAKASHI SUGIMURA*

*Genetics Division, National Cancer Center Research Institute, Tsukiji 5-chome, Chuo-ku, Tokyo 104, Japan; and †Institute of Cancer Research, Chester Beatty Laboratories, Fulham Road, London SW3 6JB, United Kingdom

Contributed by Takashi Sugimura, July 6, 1987

ABSTRACT **_hst_ is a transforming gene first identified from transformed NIH 3T3 cells that were transfected with DNA of a human stomach cancer. A genomic fragment of _hst_ obtained directly from a human genomic library also has transforming activity. This fragment has a coding sequence identical to that of the _hst_ cDNA prepared from an NIH 3T3 transformant induced by DNA from a stomach cancer. The deduced amino acid sequence of the _hst_ protein is 43%, 38%, and 40% homologous, respectively, to human basic fibroblast growth factor, human acidic fibroblast growth factor, and mouse _int-2_ protein in selected regions. This suggests that _hst_ encodes a protein related to fibroblast growth factors, which are wide-spectrum mitogens, and to the _int-2_ protein, a potential oncogene product implicated in murine mammary carcinogenesis.**

The *hst* transforming gene was first identified by transfection of murine NIH 3T3 cells with three different human DNA samples obtained from a gastric cancer, metastatic lymph node of a gastric cancer, and a noncancerous gastric mucosa (1). We cloned a cDNA copy of *hst* mRNA derived from T361-2nd-1, a transformed cell line isolated after transfection with DNA from a gastric cancer (2). An open reading frame required for transforming activity was identified and designated as ORF1. It was predicted that ORF1 encoded a 206 amino acid transforming protein with a signal peptide-like sequence at the amino terminus. We further reported successful isolation of cosmid clones containing *hst* directly from a library prepared from DNA extracted from peripheral leukocytes of a patient who was subsequently shown to be suffering from chronic myelogenous leukemia (3). All these cosmid clones, one of which was termed LpH-A, were shown to have transforming activity on NIH 3T3 cells. Subsequently, *hst* was also identified in human DNA specimens derived from three gastric cancers (4), three hepatomas (ref. 5 and Y. Yuasa, personal communication), one colon cancer, and two noncancerous colon mucosae (M. Nagao, personal communication). Thus *hst* is at present the most frequently encountered transforming gene other than the *ras* gene family.

Here we report the complete nucleotide sequence of a genomic fragment of *hst* that has transforming activity on NIH 3T3 cells.‡ The coding sequence is identical to that of T361-2nd-1 *hst* cDNA, strongly suggesting that the deduced *hst* protein is not a fusion protein, as is produced occasionally as a result of an artificial recombination during a gene-transfer experiment (6–9). This fact is the basis for the significance of further studies on the *hst* protein, including a homology search against a protein data base. The deduced amino acid sequence of the *hst* protein was found to be homologous to those of fibroblast growth factors (FGFs) and the protein encoded by the *int-2* gene, suggesting that they may constitute a gene family that is involved in cell growth.

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

MATERIALS AND METHODS

Culture and Cells. NIH 3T3 cells and the transformants were cultured as described (1). T361-2nd-1 is a secondary transformant induced by a DNA sample from a stomach cancer (no. 361). DNA-mediated gene transfer was carried out as described (1), using salmon sperm DNA as a carrier.

Plasmids. Cloning of a cosmid clone of *hst*, LpH-A, from leukemic leukocyte DNA has been described (3). A 6.2-kilobase-pair (kb) *Bam*HI–*Sal* I fragment of LpH-A was designated BS6.2 and subcloned into pBR322 to generate plasmid pLBS6.2. The BS6.2 fragment was then purified from pLBS6.2 by digestion with *Bam*HI and *Sal* I, fractionation in an agarose gel, and electroelution essentially as described (10). Linearized pLBS6.2 was prepared by digestion at the single *Bam*HI site, electrophoresed, and electroeluted from an agarose gel. pKOc5 is a eukaryotic expression vector in which the coding sequence of *hst* cDNA is driven by the simian virus 40 early promoter (2).

DNA Sequence Analysis. The BS6.2 fragment was further separated into five restriction fragments (using *Eco*RI, *Sst* I, and *Sal* I) cloned into appropriate sites of M13mp18 and M13mp19 phages, and sequenced by the dideoxy chain-termination method (11). Series of overlapping subclones were generated by the stepwise deletion method (12) for clones with large inserts. Nucleotide and amino acid sequences were analyzed using the GENETYX programs (Software Development Co. Ltd., Tokyo, Japan) for a microcomputer and the IDEAS programs (13) for the VAX/VMS computer. The National Biomedical Research Foundation protein data base§ was used for the homology search.

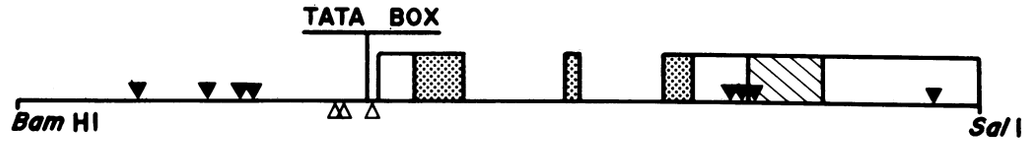
RESULTS

Transforming Activity of the *hst* Gene. pLBS6.2 transformed NIH 3T3 cells with the same efficiency (≈1 focus per nmol) as LpH-A. The resulting transformants were injected subcutaneously into athymic nude mice (5×10^6 cells per mouse), and tumors developed in all the mice within 2 weeks. The BS6.2 fragment and pLBS6.2 linearized at the *Bam*HI

Abbreviation: FGF, fibroblast growth factor.

‡This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J02986).

§Protein Identification Resource (1986) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 11.0.




```
ggatccagaaggcatccccgagtggctactccaatggagtggcttctccattcaggcaaacctgaatgggataagtcattggcaggaagatctggggccgggggtcatccagtgggaagg  120
ggagagatgacgcggtcagcatggcgggaacacaggagcagaaaggaagcaggtgggaagccaggtcaagggccaggggcagcggaaaggggtcagatgcagataagtgagtgcttcctgg  240
tgcatccttcatccgcaattcatccttacctgtgcttttgttgcctccattgcacagctgaggaggccagggcctgcggaggttgagagtgtgctcagggagcccccggagcaaagtgga  360
agccagattccagatcagttctgctgggaattcccagctcccaaaagccctgctggctgtcagtccccagtcaccacaagcacctatcctgtgtgggtgggcctgcagttctgggagata  480
tatcagctgcctgcagcgtcctttgctgaactcacagcaaataggagagacagggagggggtccttgggaagccctaaattgagcttgctgtgggagtcctgggaagaaaggagcctcatc  600
ctatcaaaagccgggggggaagacatcagagtccctctgctcaggtcagctggcacaggtgggtctccaggcctgggtctcacttccccagagggtgtgttcgggtggccccaggctgagg  720
gaggaaagcccacctcccatgtcattttgcaaatggggagtcagggacctagagatggaaagacaacacagcaagtgagggatgggttctaggtcccctgcaccctgcaccctgcaccct  840
ggccaacgatgtctatttggcaccagatctgcaggctcatctggggaccccaggacccagaggcagccgggttgcatctcgaagctgtgagctgcagcccaggaaggtccaggtctggg  960
tggcgctgcccaagcaggctgcaggcccaaggaggaacaaagatcctctcaaggggtgcggagctgaggttccggtcctgccaaagccacttgatgaccccaagtgccccctttctgc 1080
acctcagagaagagccctcaagcctcccaggtcccctccaggggcacgaataagccccagcagggttctgaaggggtcccaggaatctccctgtggggatgcggtggaggtggaggaggc 1200
tgcggtggcctggggacatctctggtcacaggtgctggtggtatgagagatggggtaggcaccaagccccctgcagctgtggctaggcgggcctgcaggaagggccaggcaggctcctca 1320
gggaccacaaagaacagggggttttcacacctaggtgggcctgcatctagctaggccagtccccatcaggccataatgggcacagtgggagtagaaccatgagtgagagaggggaggctt 1440
ccagaggcctggcctgggtccctgctagattgagggctctggctatggtacatggatatttctgctgtggaatcaaaggagcaggggatgctgaatatcccctctggccctatgccctgc 1560
tacctgtcctttcacggaagggtgtgtgtgtaggggtgcaggaccaggcctccctgggtgcatctctgccaccttgccctttggctcaggtggacctccaccaggtattcagaactcca 1680
gcccagaaacgcgccaagcctgtggggccaagacctaggggggtgggggtggcctccctcccgcctgtagccaaagggtcctcccttgcccagccaggccccggtgtcgcttactgctctt 1800
atccaccctcctcccaggccggtcctcaaggccccagcaaaggaaccaagttcccgtgagcctccgaaaggcgaagggcaggcagcagccgctggcttctgcgcccactaggagcttc 1920
ggatgcccgagttagggctgcgccaaggcggccggagcagagagggagacggggacggggacaggcagggacaaagtgcaagaggcaaaactggctgaaaagcagaagtgtaggagccgc 2040
caaggggcgggacgaacaggtccgtgggccgggcggagccaaggggtgggggccggggtccctccaggtggcactcgcggcgctagtccccagcctcctccctcccccggccctgattgg 2160
caggcggcctgcgaccagccgcgaacgccacagcgccccgggcgcccaggagaacgcgaacggccccccgcgggagcgggcgagtaggaggggggcgccgggctatatatatagcggctcg 2280
gcctcgggcgggcctggcgctcagggaggcgcGCACTGCTCCTCAGAGTCCCAGCTCCAGCCGCGCGCTTTCCGCCCGGCTCGCCGCTCCATGCAGCCGGGGTAGAGCCCGGCGCCCGGG 2400
GGCCCCGTCGCTTGCCTCCCGCACCTCCTCGGTTGCGCACTCCCGCCCGAGGTCGGCCGTGCGCTCCCGCGGGCCGCCACAGGCGCAGCTCTGCCCCCCAGCTTCCCGGGCGCACTGACC 2520

GCCTGACCGACGCACGGCCCTCGGGCCGGGATGTCGGGGCCCGGGACGGCCGCGGTAGCGCTGCTCCCGGCGGTCCTGCTGGCCTTGCTGGCGCCCTGGGCGGGCCGAGGGGGCGCCGCC 2640
                                MetSerGlyProGlyThrAlaAlaValAlaLeuLeuProAlaValLeuLeuAlaLeuLeuAlaProTrpAlaGlyArgGlyGlyAlaAla

GCACCCACTGCACCCAACGGCACGCTGGAGGCCGAGCTGGAGCGCCGCTGGGAGAGCCTGGTGGCGCTCTCGTTGGCGCGCCTGCCGGTGGCAGCGCAGCCCAAGGAGGCGGCCGTCCAG 2760
AlaProThrAlaProAsnGlyThrLeuGluAlaGluLeuGluArgArgTrpGluSerLeuValAlaLeuSerLeuAlaArgLeuProValAlaAlaGlnProLysGluAlaAlaValGln

AGCGGCGCCGGCGACTACCTGCTGGGCATCAAGCGGCTGCGGCGGCTCTACTGCAACGTGGGCATCGGCTTCCACCTCCAGGCGCTCCCCGACGGCCGCATCGGCGGCGCGCACGCGGAC 2880
SerGlyAlaGlyAspTyrLeuLeuGlyIleLysArgLeuArgArgLeuTyrCysAsnValGlyIleGlyPheHisLeuGlnAlaLeuProAspGlyArgIleGlyGlyAlaHisAlaAsp

ACCCGCGACAgtgagtggcgcggccaggcgcgaaggggcgggggcgggggcaacggccgccgggccaacccgctcagtcacactctgagaccctcggcgggcacctgctcgggggcccc 3000
ThrArgAspS

gggaaccggggcggactcgggctccggtcccttctgacgcggggctggggacgcagacactcttggctccggcagcccagcgcaaccccctgaggtcgggcgccgcctcccgccttcagaa 3120
actcgggctccgagcgccgaattccagcgccttcgcccgtgggcacagggcgcgcggtgcagccacaggggcccgagacacgcgccccggcctggcccaggctggggaaccgctggggt 3240
cgggctcgcgtctgaaggtccgggactgggtgcggccgccgggggtcccctacacaggcaagctaatctgagctagcgcaggcttgggctccggaggccctagagggcagcttgggctct 3360
ggaggccctgggggcggctgcgccgggaaccctggccctttatccccaaccccaccccagaaatagggtccccggaggcgaacaagccgaggggcggagtgggccagggatcacctgcc 3480

ccgcaatgacctgcgccccgcccccagGCCTGCTGGAGCTCTCGCCCGTGGAGCGGGGCGTGGTGAGCATCTTCGGCGTGGCCAGCCGGTTCTTCGTGGCCATGAGCAGCAAGGGCAAGC 3600
                            erLeuLeuGluLeuSerProValGluArgGlyValValSerIlePheGlyValAlaSerArgPhePheValAlaMetSerSerLysGlyLysL

TCTATGGCTCGgtgagtaccgcaggggtctggctaggcacctagttgggaacagcggacatggctagcaggctcgtggcttctccagccccacctgtgcctgggtcttggaggggtggca 3720
euTyrGlySer

gggtcaccaggtcacgggaccggcaggcctccccagacaaaggaagcagccccaaggcaggaacaatgaggttcctgccatccctgagtgggcccctcccagaccgaggaaaggggcgcta 3840
ttgagagcccttcccttctctagtccagaggggtaggtctcagtgttggaactgcgggcttgaggctggacacgcagggaatgaattctctggctgctaggtgcagggcaggtggtgaga 3960
gcaccagctgttgtgggctggccatgtccccttctcaccctgtgtgggtcttgacaccttaactgctcagcagagacatctcagcccagggtgggggtgggacagaaggggggttctgac 4080

ccctggcttcaggctgggtaccttgcccaagaggtgccccagccctgacactgccctgctttgctgcagCCCTTCTTCACCGATGAGTGCACGTTCAAGGAGATTCTCCTTCCCAACAAC 4200
                                                                        ProPhePheThrAspGluCysThrPheLysGluIleLeuLeuProAsnAsn

TACAACGCCTACGAGTCCTACAAGTACCCCGGCATGTTCATCGCCCTGAGCAAGAATGGGAAGACCAAGAAGGGGAACCGAGTGTCGCCCACCATGAAGGTCACCCACTTCCTCCCCAGG 4320
TyrAsnAlaTyrGluSerTyrLysTyrProGlyMetPheIleAlaLeuSerLysAsnGlyLysThrLysLysGlyAsnArgValSerProThrMetLysValThrHisPheLeuProArg

CTGTGACCCTCCAGAGGACCCTTGCCTCAGCCTCGGGAAGCCCCTGGGAGGGCAGTGCCGAGGGTCACCTTGGTGCACTTTCTTCGGATGAAGAGTTTAATGCAAGAGTAGGTGTAAGAT 4440
LeuEnd

ATTTAAATTAATTATTTAAATGTGTATATATTGCCACCAAATTATTTATAGTTCTGCGGGTGTGTTTTTTTAATTTTCTGGGGGGAAAAAAAAGACAAAACAAAAAACCAACTCTGACTTTT 4560
CTGGTGCAACAGTGGAGAATCTTACCATTGGATTTCTTTAACTTGTCAAAAGTTGTCACGAGTGTGCTGCTATTCTGTGTTTTAAAAAAAGGTGACATTGGATTCCGATGTCATCCCCTG 4680
TAGTATGGCGTGGAGCATCTCTGTCTGGAAAGGCCCGCCTGAGGCTTGGGCAGCCAGTTCAGGGAGCTCCCAGGCTTGGCTCTCGGCTAGCATCCTCAGAGGCCCACTCCCTTTGTGCCC 4800
TGTTGCTATTAATCGGGACATATCGGTTTACTTCGGGTACAGAAAGTGCGGTGTTGAAGTCCTCGCTGCCACTCTGTTTTTAGATCTGCCAAGACTGACCTTTGAACTTTCCTGTAGTCA 4920
ATCTTCCTCGATCTACCAGATGGGAGAGACCCTTGGACAACTTTATAAACTCCTGTTTGCCTTTTTTGGATCAGCGACAGCCCCCATCGCTGTGACTATTGGGGAAAAGACGAAGCTCTT 5040
TCATAAATTCCATGGAGAGGAATCAATATCCCACTGGAAGGCTAGAAATGGACAAGATAGTGTATTTGCAATCACAAACAAAACCCTAGTGATGAAAAATAATTTGTGATGGCAGATGCT 5160
TCTGATGGTGTGATAGAATATGTTTTTGAAAACAAACCATCGAACCCCCCGCCCCACCCCCAAAACGGGCTTCCCTGTGTTTAGGGAGCTTTGGGCTAGAACTAGCTACGATTTTTAGGT 5280
GAAATGTCCTTGTAATTGTACAAAGCACTTGGTGCAGTGTTTGCGTGGAGCAGCCTGCTGCTTTCTGATGCATTCCCTGTTTAAGTGCGTTTAACATCTACCTCACAAGCCCTGAAACCC 5400
CAGGCAAAACCCACAGAAAGCTCATACCCGGTGCAGGAGTTTGCCATCCCAAGTGGCTTTTTTTCCATATGTAGCCAAAAAGGATTGCAGATAGCGTCGGTGCGTCCCATTCGAACCTTG 5520
TCACGTTTGAGCTATCTTTACCCTGTGATTTACTTTTAGTAAGGGTGATCATGGTGAAAATATTTGCAGACAGCTGTTACAGTACACTATATGGTCACCAAGTAACCTTATATTTTTCTT 5640
TATATATTTTACAAATGTAACCCCTGTCATTGAAGCAACCGTGGAAGAGGCAGGGTCGGTGATGTTTAAAAAAAGTTCCGAGGTGATGGCAAACATTTAATTTTAATGAATGACTTTTTA 5760
GAGTTTATACAAAATGACCTTAGCTTGCTACCAGAAATGCTCCGAATGTTTCGTCAAGACTTTAATACTCTCCTAGGATGTTTCTGAACTGTCTCCCGAATTAACTTTATGGGAGTCTAC 5880
AGACAGCAAGACTGGAAAATCTGATTGGAGTTTTTGTCTTTCACATTCCTTTTGAAAACTCTTTGTTCGAATGCAAATCATCGACTTAAAATACTATTCTTAACCAAGGCCTGGAAGAAA 6000
GAAGACACTTGCAAAGCCGCTAAGACAGGACCACACATCTTAAACTGCTGTTCCTACCATGCACTAAACTGTTTTTAAGTTTTAAACCACACCCTAGGCTCCAGGAGTGTTCAGGAAAGA 6120
TGGTGTTTGTAGGTCTCCATGCTGTTTGGCGTTGGGGGGTGTGGAGGGATCATCCGTCGAC 6181
```

FIG. 1. Nucleotide sequence of BS6.2, the 6.2-kb *Bam*HI–*Sal* I insert of pLBS6.2. The sequence is schematically shown at the top, where regions existing in the cDNA of T361-2nd-1 are boxed and considered to represent exons of *hst*. Coding region corresponding to ORF1 is stippled. The other open reading frame in the cDNA (ORF2), which is unnecessary for transformation (2), is hatched. Black triangles denote enhancer core sequences, and white triangles denote "GC boxes" where transcription factor Sp1 can bind. Below the scheme, the complete nucleotide sequence of BS6.2 is presented. Sequences that do not appear in the T361-2nd-1 *hst* cDNA are in lowercase characters, while those present in cDNA are in uppercase characters with coding sequences translated into amino acids. The TATA box preceding the coding region is boxed, enhancer core sequences are underlined, and GC boxes are underlined with wavy lines. The three nucleotide sequence differences between BS6.2 and the cDNA of T361-2nd-1 are marked by triangles, and the nucleotide present in the cDNA is shown below them.

site also transformed NIH 3T3 cells with similar efficiency. pKOc5, a simian virus 40-based eukaryotic expression vector harboring the coding sequence of *hst* cDNA, transformed NIH 3T3 cells at the efficiency of 15 foci per nmol. The morphological characteristics of the transformants induced by transfection with pLBS6.2 or BS6.2 were indistinguishable from those induced by pKOc5; the cells were refractile and showed a criss-cross and piled-up arrangement.

Genomic Sequence of *hst*. The sequence of BS6.2 was determined (Fig. 1) and compared with the sequence of *hst* cDNA prepared from T361-2nd-1 cells. Only three base changes in the noncoding portion of the cDNA were detected in these two sequences. Examination of the genomic and cDNA sequences of *hst*, each of which was derived from different cells, reveals four exon–intron boundaries. The sequences at these boundaries agree with the reported consensus sequences for splice acceptor and donor sites (14). There is a "TATA box" located 42–50 base pairs (bp) upstream of the first nucleotide of the T361-2nd-1 cDNA. No "CAAT box" is present, but there are three putative Sp1-binding sites, or "GC" boxes," characterized by the sequence GGGCGG (15), upstream of the cDNA sequence. There are eight copies of a sequence that is homologous to the classical enhancer core, (G)TGG$\overset{AAA}{TTT}$(G), four of which reside 800–1537 bp upstream of the first nucleotide in the cDNA. The region spanning the 5′ noncoding region and the first exon of *hst* is high in G+C content (75%) and enriched for CpG pairs (Fig. 2). The 3′ end of BS6.2 maps to the *Sal* I site at position 2709 of the cDNA. It lacks a classical polyadenylylation signal (AATAAA) or any of its known variants (ATTAAA, AGTAAA, TATAAG, AAGAAC, AATACA). There are two possible open reading frames in the sequence of *hst* cDNA, designated ORF1 and ORF2. ORF2 spans positions 4685–5146 of the genomic sequence, a region devoid of introns (Fig. 1). A TATA box is present 220 bp upstream of the first nucleotide of ORF2, and four enhancer core sequences are noted in and around this region. The G+C content of this 3′ end of BS6.2 containing ORF2 is about 40% (Fig. 2).

Homologies to FGFs and the *int-2* Protein. As shown in Fig. 3, residues 72–204 of the *hst* protein have 43% homology to human basic FGF (16), residues 79–204 have 38% homology to human acidic FGF (17), and residues 72–174 have 40% homology to the mouse *int-2* protein (18). Human basic FGF and the mouse *int-2* protein share 44% homology in selected regions. In basic FGF two functional domains have been postulated (19), a cell-attachment site and a heparin-binding site. Human basic FGF has the consensus sequence for the cell-attachment site (Arg-Gly-Asp-Xaa) in an inverted orientation at two locations (Fig. 3); one is also present in the *hst* protein, but neither human acidic FGF nor the mouse *int-2* protein has such a sequence. The hallmark of a heparin-binding site is clusters of basic residues or pairs of basic and aromatic residues. Two such sites are found in human acidic and basic FGFs (Fig. 3). The *hst* protein also has a potential heparin-binding sequence at the corresponding positions. The location of two cysteine residues, Cys-88 and Cys-155, is highly conserved among these four proteins and they are present in a homologous region. Finally, in contrast to the *hst* protein, neither FGF has a classical signal-peptide sequence or internal hydrophobic domains.

DISCUSSION

In this paper we report the sequence of a genomic fragment of *hst*, BS6.2, from a leukemic leukocyte DNA. This fragment can transform NIH 3T3 cells upon transfection. Mammalian cells have two major classes of promoters, TATA promoters and non-TATA, G+C-rich promoters (20). The latter class is characterized by multiple GC boxes and the absence of any TATA box; promoters of this class are found in many "housekeeping" genes. The *hst* gene has a TATA box, whereas the basic FGF gene and the *int-2* gene do not. All three of these genes have several GC boxes in the 5′ flanking regions. Analysis of the genomic *hst* sequence revealed that the 5′ noncoding region and the first exon were high in G+C content and enriched for CpG pairs. This is a characteristic of many housekeeping genes, although several exceptions are known, and is designated as the *Hpa* II tiny

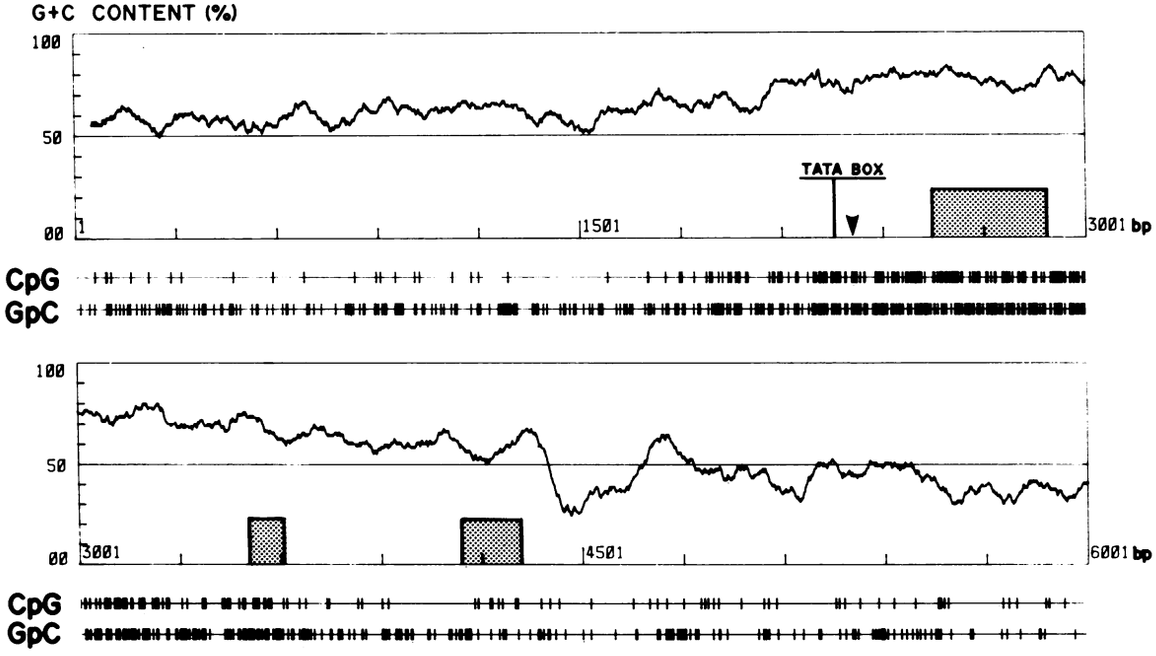


FIG. 2. The G+C content of the genomic *hst* clone is presented; below the graphs, the positions of CpG and GpC doublets are indicated by vertical bars. Stippled boxes indicate coding regions of the gene (ORF1), and the *hst* cDNA from T361-2nd-1 starts at the position marked by an arrowhead.

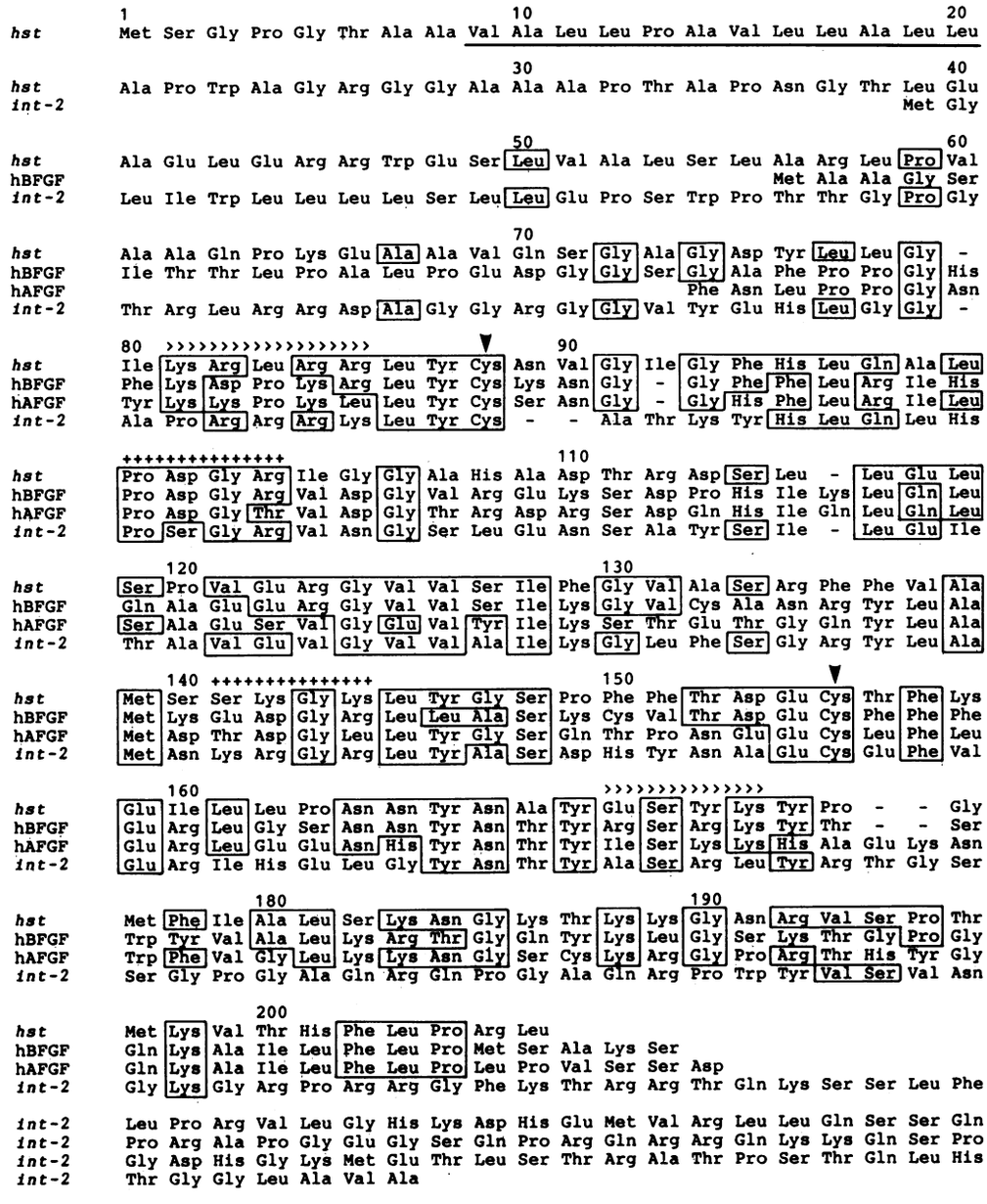


FIG. 3. Homology among the *hst* protein, FGFs, and the *int-2* protein. Entire amino acid sequences of the *hst* protein, human acidic FGF (hAFGF), human basic FGF (hBFGF), and the mouse *int-2* protein were aligned and compared. Dashes indicate gaps inserted for optimal alignment. Residues identical to the *hst* sequence are boxed. Numbers above the sequence lines refer to the *hst* residues. The cell-attachment sites and heparin-binding sites of human basic FGF are marked by ++ and >>, respectively. Two cysteine residues embedded in regions that are highly conserved among these four genes are indicated by arrowheads. Core sequence of a signal peptide is underlined in the *hst* sequence.

fragment (HTF)-like feature (21). The CpG pairs are usually hypomethylated in this region. The *hst* gene, the gene for human basic FGF (16), and the mouse *int-2* gene (18) all have HTF-like features in their 5′ portions (data not shown for the FGF and *int-2* genes). Although pLBS6.2 transforms NIH 3T3 cells efficiently, its cloned *hst* fragment lacks the classical polyadenylylation signal or any of its known variants. Since the isolated fragment, BS6.2, transforms cells with the same efficiency as linearized DNA that has the pBR322 sequence at its 3′ end, a cryptic polyadenylylation sequence in pBR322 (22) is not required for efficient transformation by this gene. One likely explanation of the sequence data and observed biological activity of BS6.2 is that the transcription of *hst* in BS6.2 terminates before the 3′ *Sal* I site, and the resulting mRNA is polyadenylylated using an unidentified sequence within the fragment as a polyadenylylation signal.

Conservation of sequences around the splice sites between the BS6.2 genomic clone and the T361-2nd-1 cDNA clone of *hst* permitted unambiguous alignment of the two sequences. It should be noted that *hst* isolated by biologic selection (the cDNA) and by molecular hybridization (the genomic clone) are identical with the exception of three base changes in the noncoding region. Both transform NIH 3T3 cells, and these transformants are tumorigenic in nude mice. Furthermore, a genomic fragment of *hst* was cloned from a normal healthy individual, again as a 6.2-kb *Bam*HI–*Sal* I fragment, and this fragment also was able to transform NIH 3T3 cells (unpublished data). Thus *hst*—like *sis*, which encodes the β chain of platelet-derived growth factor (23)—appears not to require activation by point mutation. The three base changes in the noncoding sequences of BS6.2 and T361-2nd-1 cDNA may represent genetic polymorphisms between the two individuals or point mutations that occurred in T361-2nd-1 DNA during the transfection process (24, 25). These facts suggest that the *hst* protein deduced from the cDNA sequence obtained from the NIH 3T3 transformant T361-2nd-1 did not

represent a fusion protein, which is produced occasionally by a gene rearrangement as a result of a DNA-mediated gene transfer.

Two putative open reading frames, ORF1 and ORF2, were deduced from the sequence of the T361-2nd-1 *hst* cDNA, and only ORF1 was found to be required for the transforming activity (2). The probabilities of these two sequences to encode proteins were evaluated by the method of Fickett (26), which is based on nonrandom compositional patterns of nucleotide sequences in coding regions. This test revealed that ORF1, but not ORF2, was likely to encode a protein (data not shown). ORF2 is embedded in the region that corresponds to the 3′ third of BS6.2. This region contains a TATA box in front of ORF2 and four enhancer core sequences but seems to have no introns. The G+C content of this region is about 40%, the average percentage for the human genome as a whole. The deduced amino acid sequence of ORF2 revealed no significant homology to known protein sequences stored in the National Biomedical Research Foundation protein data base (release 11.0). Although we are inclined to suppose that ORF2 is merely a fortuitous open reading frame, its significance remains unknown.

Significant homologies exist among the *hst* protein, human FGFs, and the mouse *int-2* protein, which suggests that this group may constitute a family of genes involved in cell growth. FGFs are potent mitogens for a variety of cell lineages, including those of mesodermal, neuroectodermal, and epithelial origins (27). They may play important roles in tumor development (28) and in normal angiogenic processes such as tissue repair, and they may also be involved in organogenesis (29). Although there are many FGF analogues, they fall into one of two classes—acidic or basic FGF, which share 55% homology in amino acid sequences. Both classes have high affinity for heparin, a complex proteoglycan, and they may bind to the same receptor on the cell surface (30). Both FGFs have homology with the *int-2* protein, a potential oncogene product implicated in murine breast cancer induced by mouse mammary tumor virus (18). Two possible functional domains of basic FGF (19) are also conserved in the *hst* protein. One of these is a cell-attachment site found in proteins such as fibronectin, fibrinogen, collagen, and thrombin. The other is a heparin-binding site found in fibronectin, antithrombin III, and platelet factor 4. Conservation of such domains may signify not only structural, but also some functional, homologies among these proteins. However, there is a notable difference between the *hst* protein and FGFs. The *hst* protein has a classical signal-peptide sequence, whereas FGFs have neither this sequence nor internal hydrophobic domains. A signal-peptide sequence and/or internal hydrophobic domains are features present in many other well-characterized extracellular proteins. It is plausible that the FGFs are usually intracellular proteins, segregated from the cell surface receptors, and that they are liberated by lysis of the FGF-producing cells when required. In view of the fact that every growth-factor gene might be a potential oncogene, it is interesting that the *hst* protein has, but FGFs do not have, a signal-peptide sequence that may facilitate secretion and easy access of the protein to potential target molecules on the cell surface.

We thank S. J. Silverstein for critical reading of the manuscript. This work was supported in part by a grant-in-aid from the Ministry of Health and Welfare for a Comprehensive 10-Year Strategy for Cancer Control, Japan.

1. Sakamoto, H., Mori, M., Taira, M., Yoshida, T., Matsukawa, S., Shimizu, K., Sekiguchi, M., Terada, M. & Sugimura, T. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 3997–4001.
2. Taira, M., Yoshida, T., Miyagawa, K., Sakamoto, H., Terada, M. & Sugimura, T. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2980–2984.
3. Yoshida, T., Sakamoto, H., Miyagawa, K., Little, P. F. R., Terada, M. & Sugimura, T. (1987) *Biochem. Biophys. Res. Commun.* **142,** 1019–1024.
4. Koda, T., Sasaki, A., Matsushima, S. & Kakinuma, M. (1987) *Gann* **78,** 325–328.
5. Nakagama, H., Ohnishi, S., Imawari, M., Hirai, H., Takaku, F., Sakamoto, H., Terada, M., Nagao, M. & Sugimura, T. (1987) *Gann* **78,** 651–654.
6. Martin-Zanca, D., Hughes, S. H. & Barbacid, M. (1986) *Nature (London)* **319,** 743–748.
7. Birchmeier, C., Birnbaum, D., Waitches, G., Fasano, O. & Wigler, M. (1986) *Mol. Cell. Biol.* **6,** 3109–3116.
8. Ishikawa, F., Takaku, F., Nagao, M. & Sugimura, T. (1987) *Mol. Cell. Biol.* **7,** 1226–1232.
9. Takahashi, M. & Cooper, G. M. (1987) *Mol. Cell. Biol.* **7,** 1378–1385.
10. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 97–172.
11. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
12. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene* **33,** 103–119.
13. Kanehisa, M. (1982) *Nucleic Acids Res.* **10,** 183–196.
14. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S. & Sharp, P. A. (1986) *Annu. Rev. Biochem.* **55,** 1119–1150.
15. Gidoni, D., Dynan, W. S. & Tjian, R. (1984) *Nature (London)* **312,** 409–413.
16. Abraham, J. A., Whang, J. L., Tumolo, A., Mergia, A., Friedman, J., Gospodarowicz, D. & Fiddes, J. C. (1986) *EMBO J.* **5,** 2523–2528.
17. Jaye, M., Howk, R., Burgess, W., Ricca, G. A., Chiu, I.-M., Ravera, M. W., O'Brien, S. J., Modi, W. S., Maciag, T. & Drohan, W. N. (1986) *Science* **233,** 541–545.
18. Moore, R., Casey, G., Brookes, S., Dixon, M., Peters, G. & Dickson, C. (1986) *EMBO J.* **5,** 919–924.
19. Esch, F., Baird, A., Ling, N., Ueno, N., Hill, F., Denoroy, L., Klepper, R., Gospodarowicz, D., Böhlen, P. & Guillemin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6507–6511.
20. Dynan, W. S. (1986) *Trends Genet.* **2,** 196–197.
21. Bird, A. P. (1986) *Nature (London)* **321,** 209–213.
22. Kessler, M. M., Westhafer, M. A., Carson, D. D. & Nordstrom, J. L. (1987) *Nucleic Acids Res.* **15,** 631–642.
23. Josephs, S. F., Ratner, L., Clarke, M. F., Westin, E. H., Reitz, M. S. & Wong-Staal, F. (1984) *Science* **225,** 636–639.
24. Hauser, J., Levine, A. S. & Dixon, K. (1987) *EMBO J.* **6,** 63–67.
25. Calos, M. P., Lebkowski, J. S. & Botchan, M. R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3015–3019.
26. Fickett, J. W. (1982) *Nucleic Acids Res.* **10,** 5303–5318.
27. Crabb, J. W., Armes, L. G., Carr, S. A., Johnson, C. M., Roberts, G. D., Bordoli, R. S. & McKeehan, W. L. (1986) *Biochemistry* **25,** 4988–4993.
28. Klagsbrun, M., Sasse, J., Sullivan, R. & Smith, J. A. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 2448–2452.
29. Slack, J. M. W., Darlington, B. G., Heath, J. K. & Godsave, S. F. (1987) *Nature (London)* **326,** 197–200.
30. Neufeld, G. & Gospodarowicz, D. (1986) *J. Biol. Chem.* **261,** 5631–5637.