vav, a novel human oncogene derived from a locus ubiquitously expressed in hematopoietic cells

Shulamit Katzav, Dionisio Martin-Zanca and Mariano Barbacid¹

Departmental Oncology Section, BRI-Basic Research Program, Frederick Cancer Research Facility, PO Box B, Frederick, MD 21701, USA

¹Present address: Department of Molecular Biology, Squibb Institute for Medical Research, PO Box 4000, Princeton, NJ 08543, USA

Communicated by J.Schlessinger

A novel human oncogene, designated *vav*, was generated by a genetic rearrangement during gene transfer assays. The *vav* oncogene directs the synthesis of a 3.0 kb mRNA from which we isolated a 2.8 kb-long complementary DNA copy. Nucleotide sequence analysis of this *vav* oncogene cDNA clone revealed that its 5' 167 bp were derived from pSV2neo DNA cotransfected as a selectable marker during gene transfer. The remaining 2597 bp were unrelated to genes included in current data banks, indicating that the *vav* oncogene is likely to be derived from a novel human locus. The *vav* oncogene cDNA clone encompasses a 2391 bp long open reading frame (ORF) capable of directing the synthesis of a 797 amino acid long polypeptide. The predicted *vav* oncogene protein sequence exhibits several motifs reminiscent of transcriptional factors. They include a highly acidic amino-terminal region separated from two putative nuclear localization signals by a proline-rich sequence, presumably a hinge region. In addition, we identified two zinc-finger-like domains, one of which conforms to the canonical pattern Cys-X₂-Cys-X₁₃-Cys-X₂-Cys previously found to confer *trans*-activating activity to the adenovirus E1A protein. Transcription of its normal allele, the *vav* proto-oncogene, has been exclusively observed in cells of hematopoietic origin, including those of erythroid, lymphoid and myeloid lineages. These findings raise the possibility that this novel locus might play an important role in hematopoiesis.

Key words: *vav*/oncogene/hematopoietic cells/zinc finger

Introduction

Malignant transformation has been used as a genetic marker to identify vertebrate loci that play active roles in the regulation of cellular proliferation and/or differentiation. These loci harbor the normal alleles of transforming genes present in acute transforming retroviruses, retroviral-induced tumors and spontaneous and carcinogen-induced malignancies (Bishop, 1987). In addition to those oncogenes present in naturally occurring tumours, transforming genes can be generated by rearrangements that occur during gene transfer assays. Some of these *in vitro* activated oncogenes are homologous to those previously found in human tumors such as *trk* (Martin-Zanca *et al.*, 1986; Kozma *et al.*, 1988) or

in certain retroviruses such as *raf* (Rapp *et al.*, 1983; Muller and Muller, 1984; Fukui *et al.*, 1985; Shimizu *et al.*, 1985). Other oncogenes including *mas* (Young *et al.*, 1986), *db1/mcf-2* (Fasano *et al.*, 1984; Eva and Aaronson, 1985), *ret* (Takahashi *et al.*, 1985), *hst/K-fgf* (Delli Bovi *et al.*, 1987; Taira *et al.*, 1987), *fgf-5* (Zhan *et al.*, 1988), *B-raf* (Ikawa *et al.*, 1988) and *tre* (Nakamura *et al.*, 1988) represent previously unidentified loci. Characterization of these transforming genes and their normal alleles will expand the number of loci available for the study of molecular pathways involved in the control of cell proliferation and differentiation.

We describe here the isolation and molecular characterization of *vav*, a new human oncogene that became activated during the course of gene transfer assays aimed at determining the presence of transforming genes in human esophageal carcinomas suspected of having a chemical etiology. Molecular characterization of this oncogene revealed a novel locus that encodes a zinc finger-containing protein whose expression is restricted to cells of hematopoietic origin.

Results

Identification of the *vav* oncogene

DNAs isolated from several esophageal carcinomas kindly provided by Dr R.Montesano (WHO, Lyon) were tested for the ability to transform NIH3T3 mouse fibroblasts. Although none of the DNAs was able to induce foci of morphologically transformed cells, one DNA sample was positive in the *in vivo* nude mouse tumorigenicity assay (Blair *et al.*, 1982; Fasano *et al.*, 1984). NIH3T3 cells transfected with this tumor DNA in the presence of the selectable marker pSV2neo (Southern and Berg, 1982) were expanded *in vitro*, pooled and inoculated into mice. These cells elicited the appearance of tumors in 9 out of the 10 mice injected in <4 weeks. Southern transfer analysis of DNA isolated from these nude mouse tumors revealed the presence of multiple human *Alu*⁺ repetitive sequences.

When the DNA isolated from the nude mouse tumors was used in additional cycles of transfection, tumors were induced in 100% of the mice inoculated. DNA isolated from second- and third-cycle nude mouse tumors exhibited a consistent pattern of human *Alu*⁺ sequences (*Eco*RI DNA fragments of 30, 9 and 5 kb), a result indicative of the presence of a human oncogene (Figure 1A). The human sequences present in these human *Alu*⁺ DNA fragments did not hybridize, even under low-stringency conditions, to any of 20 oncogene-specific probes available at the time these experiments were conducted (data not shown). These results indicated that these nude tumors might contain a novel human oncogene. Since this transforming gene represented the sixth distinct oncogene identified in our laboratory, it was designated by the acronym *vav*, the sixth letter of the Hebrew alphabet, the native language of one of us (S.K.).

In view of the complexity of the *vav* locus (≥ 35 kb) we wished to isolate *vav* cDNA clones. To identify such clones, we generated a genomic probe containing *vav* exonic sequences. DNA isolated from a third-cycle nude mouse tumor was partially digested with *Sau3AI* and used to create a genomic library in λ EMBL-4. One million recombinant phages were screened with a probe consisting of ^{32}P -labeled nick-translated human genomic DNA. One hundred phages were found to be positive. Five plaque-purified phages carrying inserts of 15–20 kb were grown up, digested with restriction endonucleases and several DNA fragments free of human repetitive sequences were isolated by preparative gel electrophoresis. These DNA fragments were screened for the possible presence of exon sequences by hybridizing them to genomic DNAs isolated from several mammalian species. This test was based on the assumption that exonic sequences are more likely to be evolutionarily conserved than intron sequences. Among those DNA fragments tested, a 800 bp *Bam*HI–*Eco*RI probe efficiently hybridized with single-copy DNA sequences present in the hamster, mouse, mink, macaque and human genomes (data not shown). This probe was subsequently used in Northern transfer analysis to identify a single 3.0 kb mRNA species specifically expressed in *vav*-induced nude mouse tumors (Figure 1B). This transcript is likely to represent the transcriptional product of the *vav* oncogene.

Isolation of cDNA clones

Poly(A)-selected RNA isolated from a *vav*-induced third-cycle nude mouse tumor was used to prepare a 10^6 -member *Eco*RI cDNA library in the λ gt10 cloning vector. When this library was screened with the 800 bp *Bam*HI–*Eco*RI *Alu*⁺ probe, $\sim 10^3$ phages were positive. Twenty recombinant phages were picked at random from those exhibiting the strongest signal. The longest insert found in these recombinant phages was a 2.3 kb *Eco*RI fragment, likely to represent an incomplete cDNA clone. To isolate a more representative *vav* cDNA clone, a 300 bp *Pst*I DNA fragment corresponding to the 5' end of this insert was purified and used to rescreen the cDNA library. Eleven phages carrying inserts of 2.5–2.8 kb were isolated. The one exhibiting the longest insert was selected for further studies.

Transforming activity of the *vav* oncogene

To test whether the *vav* cDNA clone (2.8 kb) was biologically active, we subcloned it into the unique *Eco*RI site of pMEX, a mammalian expression vector that carries a multiple cloning site flanked by a Moloney murine sarcoma virus (MSV) long terminal repeat (LTR) and polyadenylation signal from SV40. The resulting expression plasmid, pSK27, induced 10^3 – 10^4 foci per μg of DNA when transfected into NIH3T3 cells (Graham and van der Eb, 1973). A similar construction in which the cDNA clone was inserted in the opposite orientation did not have detectable biological activity (data not shown). *vav*-induced foci consist of dense, non-refractile cells that closely resemble those that spontaneously appear in dense cultures of NIH3T3 cells (Figure 2A). In addition, the *vav*-induced foci often exhibit formation of syncytia resulting in giant multinucleated cells (Figure 2B). Cells derived from either NIH3T3 foci or from nude mouse tumors retain a rather normal morphology in

subconfluent cultures. However, the *vav*-transformed NIH3T3 cells grow in semisolid media with relatively good efficiency (0.2–1%) and readily induce tumors when injected into nude mice.

Nucleotide sequence analysis

The nucleotide sequence of the biologically active 2.8 kb cDNA clone of the *vav* oncogene was determined by an automated chain-termination method using a difluoresceinated primer and laser excitation as described (Brumbaugh *et al.*, 1988). Single-stranded DNA purified from 5' and 3' deletion mutants generated by unidirectional deletions using the Exo III/Mung bean nuclease method were used as templates for sequencing. Figure 3 depicts the entire nucleotide sequence of the 2765-bp *vav* oncogene cDNA clone and the deduced amino acid sequence of its putative gene product.

Several biologically relevant features could be deduced from the nucleotide sequence: (i) nucleotides 1–167 were identical to the pSV2neo sequences corresponding to the end of the SV40 early promoter region (nucleotides 1–17) and the beginning of the bacterial Tn5 sequences harboring the *neo* gene (nucleotides 18–167); (ii) no significant homology could be detected among sequences located between nucleotides 168 and 2765 of the *vav* cDNA clone and those stored in gene data banks; (iii) the cDNA clone contains a 2391 bp-long open reading frame (nucleotides 111–2501) capable of coding for a polypeptide of 797 amino acid residues with a calculated relative molecular mass of 91 641 daltons; (iv) this open reading frame (ORF) is followed by

Fig. 1. (A) Detection of human *Alu*⁺ sequences in nude mouse tumors induced by the *vav* oncogene. DNAs (20 μg) isolated from (a) NIH3T3 cells, (b) second and (c) third-cycle *vav*-induced nude mouse tumors were digested with *Eco*RI and submitted to Southern transfer analysis (Southern, 1975). Hybridization was conducted for 48 h under stringent conditions (42°C in 5 $\times$ SSC, 50% formamide, 1 $\times$ Denhardt's solution) using 5 $\times 10^7$ c.p.m. of ^{32}P -labeled nick-translated high mol. wt human DNA. Filters were exposed to Kodak XAR-5 film at -70°C for 20 h in the presence of intensifier screens. λ *Hind*III DNA fragments were used as mol. wt markers. The migration of the three major *Alu*⁺ DNA fragments is indicated by arrows. (B) Identification of *vav* oncogene transcripts. Poly(A)-selected RNA (3 μg) was isolated from (a) NIH3T3 cells and (c) a third-cycle *vav*-induced nude mouse tumor and submitted to Northern transfer analysis (Lehrach *et al.*, 1977). Nitrocellulose filters were hybridized for 48 h under stringent conditions (50% v/v formamide, 42°C) to 5 $\times 10^7$ c.p.m. of a ^{32}P -labeled nick-translated 800 bp *Bam*HI–*Eco*RI DNA derived from an *Alu*⁺ genomic fragment of the *vav* oncogene. The filter was exposed to Kodak XAR-5 film for 30 min. 28S and 18S RNA of *Saccharomyces cerevisiae* and 23S and 16S RNA of *Escherichia coli* R-13 were used as mol. wt markers. The migration of the 3.0 kb *vav* oncogene transcript is indicated by an arrow.

Fig. 2. Morphology of NIH3T3 cells transfected with the *vav* oncogene cDNA clone under the transcriptional control of retroviral regulatory elements. (A) A representative focus. (B) occasional syncytia resulting in the formation of giant multinucleated cells.

264 bp of 3' non-coding sequences that contain a consensus polyadenylation signal (AATAAA, nucleotides 2738–2743) followed 14 residues downstream by a short (8 bp) poly(A) tail.

Predicted amino acid sequences

The nucleotide sequence of the cDNA clone of the *vav* oncogene predicts the synthesis of a 797 amino acid long polypeptide with the following relevant features (Figure 3):

A

B

SV40 Tn5 Transposase

CTAGGCTTTTGC AAAAGCTTCACGCTGCCGCAAGCACTCAGGGCGCAAGGGCTGCTAAAGGAAGCGGAACACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGG 110

ATGAATGTCAGTACTGGCTATCTGGACAAGGAAAACGCAAGCGCAAGAGAAAGCAGTTCTCTGCTTAAAGAACATTAGAACCTTCTGTCACCTGCTGAGAGTTCGGCCTC 230

MetAsnValSerTyrTrpAlaIleThrTrpArgGluAsnAlaSerAlaLysArgLysGlnPheLeuLysLeuLysAsnIleArgThrPheLeuSerThrCysGluLysPheGlyLeu 40

AAAGCGAGCGAGCTCTTGAAGCCTTTGACCTCTCGATGTGCAGGATTTGGCAAGGTCATCATACCCCTGTCTGCTGTCTGCGACCCCGATCGCCAGCAACAGGGGATCATGCC 350

LysArgSerGluLeuPheGluAlaPheAspLeuPheAspValGlnAspPheGlyLysValIleTyrThrLeuSerAlaLeuSerTrpThrProIleAlaGlnAsnArgGlyIleMetPro 80

TTCCCCCAGGAGGAGAGTGTAGGTGATGAAGACATCTACAGTGGCTGTCCAGCAGATCGACGACCGGTGGAGAGGATGAGACCTGTATGACTGCTGGGAATGAGGAGCGG 470

PheProThrGluGluSerValGlyAspGluAspIleTyrSerGlyLeuSerAspGlnIleAspAspThrValGluGluAspGluAspLeuTyrAspCysValGluAsnGluAla 120

GAAGCGACGAGATCTATGAGGACCTATCGCTCGGAGCCCGTCCATGCCGCCAAGATGACAGAGTATGACAAGCGCTGCTGCTGCGGGAGATCCAGCAGACGGAGGAGAAG 590

GluGlyAspGluIleTyrGluAspLeuMetArgSerGluProValSerMetProProLysMetThrGluTyrAspLysArgCysCysLeuArgGluIleGlnGlnThrGluGlyLys 160

TACACTGACACGCTGGGCTCCATCCAGCAGCATTTCTGAAGCCCTGCAACGGTCTCTGAAACCTCAAGACATTGAGATCATCTTTATCAACATTGAGGACCTGCTCGTGTCACT 710

TyrThrAspThrLeuGlySerIleGlnGlnHisPheLeuLysProLeuGlnArgPheLeuLysProGlnAspIleGluIleIlePheIleAsnIleGluAspLeuLeuArgValHisThr 200

CACCTCTCAAGGAGATGAAGGAAGCCCTGGGCAACCGAATCTCTACCGGCTCTCATCAAATACAAGGAGAGGTTCTCGTCTATGGCCGCTACTGACCCAGGTGGAG 830

HisPheLeuLysGluMetLysGluAlaLeuGlyThrProGlyAlaProAsnLeuTyrGlnValPheIleLysTyrLysGluArgPheLeuValTyrGlyArgTyrCysSerGlnValGlu 240

TCAGCGACAAACACCTGGACCGTGTGGCCGAGCCCGGAGGACGTCAGATGAAGCTGAGGAATGCTCTCAGAGAGCAACACGGGAGTTCACTGCGGACCTGCTGATGGTCC 950

SerAlaSerLysHisLeuAspArgValAlaAlaAlaArgGluAspValGlnMetLysLeuGluGlyCysSerGlnArgAlaAsnAsnGlyArgPheThrAlaArgProAlaAspGlyAla 280

TATGACGCGAGTTCTCAAATATCACCTCTCTCCAGGAGCTGGTGAACACACGAGGAGCGATGGAGCAAGAACTGCGGCTGGCCCTGGATGCCATGAGGACCTGGCTCAGTGG 1070

TyrAlaAlaSerSerGlnIleSerProProSerProGlyAlaGlyGluThrHisAlaGlyIleTyrAspGlyAlaArgLysLeuLeuAlaLeuAspAlaMetArgAspLeuAlaGlnCys 320

GTGAACGAGGTCAAGCGAGACAACGAGACATGCGACAGATCCCAATTTCCAGCTGTCCATTGAGAACCTGGACAGTCTCTGGCTCACTATGGCCGCCCAAGATCGACGGGGAAC 1190

ValAsnGluValLysArgAspAsnGluLeuArgGlnIleThrAsnPheGlnLeuSerIleGluAsnLeuAspGlnSerLeuAlaHisTyrGlyArgProLysIleAspGlyIle 360

AAGATCACCTCGGTGGAACGGCGCTCAAGATGGACAGGTATGCCCTCTGCTCGACAAGCTCTACTCATCTGTAAGCGCAGGGGAGACTCTATGACCTCAAGGACTTTGTAACCTG 1310

LysIleThrSerValGluArgArgSerLysMetAspArgTyrAlaPheLeuLeuAspLysAlaLeuLeuIleCysLysArgArgGlyAspSerTyrAspLeuLysAspPheValAsnLeu 400

CACAGCTTCCAGGTTCGGGATGACTCTTCAGGAGACCCGAGACAACAAGATGGAGGCACATGTTCTCTCTGAGGAGCAAGGTGCCAGGGCTATGAGCTGTTCTTCAAGACAAGA 1430

HisSerPheGlnValArgAspAspSerSerGlyAspArgAspAsnLysLysTrpSerHisMetPheLeuLeuIleGluAspGlnGlyAlaGlnGlyTyrGluLeuPhePheLysThrArg 440

GAATTGAAGAAGAAGTGGATGGAGCAGTTTGATAGTCCATCTCCAACATCTATCCGGGAATGCCACCGCAACGGGCATGACTTCCAGATGTTCTCTTTGAGGAGACCATCTG 1550

GluLeuLysLysLysTrpMetGluGlnPheGluMetAlaIleSerAsnIleTyrProGluAsnAlaThrAlaAsnGlyHisAspPheGlnMetPheSerPheGluGluThrThrSerCys 480

AAAGCGTTCAGATGCTGCTAGAGTACCTTCTATCAGGGCTACCGCTGCTCGGTCGGGCATCTGCACACAAGGAGTGTCTGGGAGGGTCCCTCCATGTGGCCGACATGGGCAA 1670

LysAlaCysGlnMetLeuLeuArgGlyThrPheTyrGlnGlyTyrArgCysHisArgCysArgAlaSerAlaHisLysGluCysLeuGlyArgValProProCysGlyArgHisGlyGln 520

GATTTCACGAGAACTATGAAGAAGGACAAACTACATCGCAGGGCTCAGGACAAAAGGAATGAGCTGGGCTGCGCAAGATGGAGGTGTTTACGGAATACTACGGCTCTCTCCACCC 1790

AspPheProGlyThrMetLysLysAspLysLysArgAlaGlnAspLysLysArgAsnGluLeuGlyLeuProLysMetGluValPheGlnGluTyrTyrGlyLeuProPro 560

CCTGGAGCATTGGACCTTTCTACGGCTCAACCTGGAGACATTTGTGAGCTCACGAAGGCTGAGGCTGAACAGAAGTGGTGGGAGGGCAGAAATACATCTACTAATGAAATGGCTGG 1910

ProGlyAlaIleGlyProPheLeuArgLeuAsnProGlyAspIleValGluLeuThrLysAlaGluAlaGluGlnAsnTrpTrpGluGlyArgAsnThrSerThrAsnGluIleGlyTrp 600

TTTCCTTGAACAGGGTGAAGCCCTATGTCATGGCCCTCTCAGGACCTGTCTGTTCTATCTGCTGACGAGGCCCATGGAGCGGGCAGGGCAGAGAGCATCTGCCCAACCGCTCG 2030

PheProCysAsnArgValLysProTyrValHisGlyProProGlnAspLeuSerValHisLeuTrpTyrAlaGlyProMetGluArgAlaGlyAlaGluSerIleLeuAlaAsnArgSer 620

GACGGGACTTCTTGGTGGCGCAGAGGGTGAAGGATGCAGCAGAAATTTGCCATCAGCATTAATATAACGTGAGGTCAAGCACCGGTTAAATCATGACAGCAGAGGAGCTGTACCGG 2150

AspGlyThrPheLeuValArgGlnArgValLysAspAlaAlaGluPheAlaIleSerIleLysTyrAsnValGluValLysHisThrValLysIleMetThrAlaGluGlyLeuTyrArg 660

ATCACAGAGAAAAGGCTTTCGGGGGCTTACGGAGCTGGTGGAGTTTACCAGCAGAACTCTTAAAGGATGCTTCAAGTCTCTGGACACCCACTTGCAGTTCCTCTCAAGGAGCCT 2270

IleThrGluLysLysAlaPheArgGlyLeuThrGluLeuValGluPheTyrGlnGlnAsnSerLeuLysAspCysPheLysSerLeuAspThrThrLeuGlnPheProPheLysGluPro 720

GAAAAGAGAACCATCAGCAGGCCAGCAGTGGGAAGCACAAGATTTTGGCAGCGCAAGCCCGCTATGACTTTCGGCCCGTGACCGTTACAGAGCTGCGCTCAAGGAGGGTGACATC 2390

GluLysArgThrIleSerArgProAlaValGlySerThrLysTyrPheGlyThrAlaLysAlaArgTyrAspPheCysAlaArgAspArgSerGluLeuSerLeuLysGluGlyAspIle 760

ATCAAGATCCTTAAACAAGGAGCAGCAAGGCTGGTGGCGAGGGGAGATCTATGGCCGGGTTGGCTGGTTCCTGCCAACTACGTGGAGGAAGATTTATCTGAATATGCTGAGCCCTG 2510

IleLysIleLeuAsnLysLysGlyGlnGlnGlyTrpTrpArgGlyGluIleTyrGlyArgValGlyTrpPheProAlaAsnTyrValGluGluAspTyrSerGluTyrCys*** 780

GTGCTTGGCAGAGACGAGAACTCCAGGCTCTGAGCCCGGCTGGCGAGGACGCGGACGAGGGCTGTGACAGCTCCGCGGGTGGAGACTTTGGATGGACTGGAGGAGGCCAGCG 2630

TCCAGCTGGCGGTGCTCCCGGATGTGCCCTGACATGGTTAATTTATAACCCCGATTTTCTCTTGGGTCCCTCAAGCAGACGGGGCTCAAGGGGTACATTTAATAAAGGATG 2750

AAGATGAAAAA 2765

(i) the 19 amino-terminal residues are coded by sequences derived from the bacterial Tn5 sequences harboring the *neo* gene, albeit in a different reading frame than the bacterial enzyme; (ii) there are six potential N-glycosylation sites (Asn-X-Ser/Thr), two of which are located within the rearranged sequences contributed by the bacterial Tn5 gene; (iii) a 45 amino acid long domain (residues 84–128) composed of mostly acidic (Glu/Asp) amino acid residues (23 residues or 51%); (iv) two proline-rich sequences including Pro-Pro-Ser-Pro (residues 289–292) and Pro-Pro-Pro-Pro (residues 558–561) that may represent hinge regions needed for the appropriate folding of the *vav* oncogene product; (v) an Arg-Arg-Gly-Asp-Ser-Tyr motif (residues 387–392) that may be a phosphorylation site for protein kinase A (Creighton, 1984); (vi) two putative nuclear localization signals, Lys-Thr-Arg-Glu-Leu-Lys-Lys-Lys (residues 438–445) and Lys-Lys-Asp-Lys-Leu-His-Arg-Arg (residues 527–534) (Dingwall and Kaskey, 1986); and (vii) two putative zinc-finger-like motifs (Sunderman and Barber, 1988). The first of these motifs, Cys-X₂-Cys-X₁₃-Cys-X₂-Cys (residues 480–500), conforms to the canonical pattern Cys-X₂₋₄-Cys-X₂₋₁₅-Cys-X₂₋₄-Cys (Berg, 1986). Moreover, this type of motif has been found to confer *trans*-activating activity to the larger (289 amino acids) of the two proteins coded for by the adenovirus E1A oncogene (Culp *et al.*, 1988). The second motif, His-X₂-Cys-X₆-Cys-X₂-His (residues 505–518), exhibits two histidine residues in the flanking positions. Although histidine residues often replace cysteines in the Cys-X₂₋₄-Cys domain, this particular structure has not been previously described. These features, taken together, suggest that the *vav* oncogene might encode a DNA-binding phosphoprotein, perhaps a transcriptional factor (Ptashne, 1988).

Mechanism of activation

The above nucleotide sequence suggests that the *vav* oncogene became activated by a rearrangement that placed SV40 regulatory sequences present in the cotransfecting pSV2neo plasmid DNA in front of a novel cellular proto-oncogene. Since nucleotide 1 of our *vav* cDNA clone is located 46 bp downstream from the major site used *in vivo* for the initiation of SV40 early mRNA synthesis (SV40 nucleotide 5237), it is likely that transcription of *vav* oncogene sequences is directed by SV40 regulatory elements.

Transcriptional activation of the *vav* oncogene by the SV40 early promoter raised the possibility that this transforming gene might have been generated during the course of gene transfer assays. To prove this hypothesis, DNA isolated from the original esophageal carcinoma was tested by Southern transfer analysis along with DNA prepared from white blood

cells of healthy donors and from *vav*-induced nude mouse tumors. When the DNAs were probed with a 900 bp *Pst*I DNA fragment derived from the 5' end of the *vav* oncogene cDNA clone, only the mouse tumor DNAs induced by this oncogene contained rearranged sequences (data not shown). These results indicate that the *vav* oncogene became activated during *in vitro* manipulations and is unlikely to play a role in the genesis of human esophageal tumors.

To determine whether additional mutations were required to activate the *vav* oncogene, we isolated its normal allele from a cDNA library of K562 cells (Shtivelman *et al.*, 1985). K562 is a hematopoietic cell line derived from a chronic myelogenous leukemia which expresses significant levels of a 2.9 kb-long *vav*-related transcript. Two million recombinant λ gt10 phages were screened with the 5' 900 bp *Pst*I DNA fragment of the *vav* oncogene. Five phages were found to be positive, and the one containing the longest insert (2.8 kb) was subcloned into Bluescript KS and used in subsequent studies.

Comparison of the restriction endonuclease map of this normal *vav* cDNA clone with that of its transforming allele revealed complete identity except at their respective 5' termini (data not shown). Nucleotide sequence analysis of the 5' sequences of the normal *vav* gene transcript confirmed that they were replaced by those derived from pSV2neo during the generation of the *vav* oncogene (Figure 4). Next, we prepared a chimeric *vav* gene construct in which the 2.6 kb *Sac*I–*Eco*RI DNA fragment of the *vav* oncogene was replaced by the corresponding sequences derived from the normal *vav* proto-oncogene (Figure 5). The resulting plasmid, pSK77, exhibited a transforming activity (10^3 f.f.u./ μ g) comparable to that of the *vav* oncogene (Figure 5). These findings demonstrate that insertion of the pSV2neo sequences in the 5' domain of the normal *vav* proto-oncogene was sufficient for its malignant activation.

vav proto-oncogene expression

In order to gain information regarding the physiological role of the normal human *vav* locus, we screened a battery of human cell lines from distinct lineages for expression of proto-*vav* gene sequences. Representative results are depicted in Figure 6. None of the epithelial, mesenchymal or neuroectodermal cell lines tested expressed detectable levels of the normal 2.9 kb proto-*vav* gene transcript (Table I). In contrast, this transcript could be readily observed in mRNAs prepared from cell lines of hematopoietic origin (Figure 6). Cell lines representative of each of the three major hematopoietic differentiation lineages—lymphoid, myeloid and erythroid—were found to express approximately equivalent levels of the 2.9 kb proto-*vav* gene transcript

Fig. 3. Nucleotide sequence of a *vav* oncogene cDNA clone. (A) Schematic diagram. Untranslated sequences are depicted by a thin bar. Coding sequences (flanked by the initiator ATG and terminator TGA codons) are represented by the thicker box. Highlighted domains include sequences derived from pSV2neo (▨), a region rich in acidic amino acid residues (■), two proline-rich stretches (▤), a potential target for protein kinase A phosphorylation (▧), two putative nuclear localization signals (▩) and two zinc-finger-like domains (□). Cysteine residues (●) and consensus glycosylation sites (▽) are also shown. (B) Nucleotide and predicted amino acid sequence of the 2765 bp insert of the *vav* oncogene cDNA clone. The sequences of the flanking *Eco*RI linkers have been omitted. Nucleotide numbers are indicated in the right column. Amino acid numbers are indicated underneath the corresponding residue. Nucleotides 1–17 (SV40 early promoter) and 18–167 (Tn5 bacterial gene) are derived from pSV2neo. Nucleotide 168–2765 are derived from the *vav* proto-oncogene. The vertical arrow (↓) indicates the separation between pSV2neo and *vav* proto-oncogene-derived sequences. The presumed initiator ATG codon is boxed and the terminator TGA codon is marked with asterisks. Other in-frame termination codons are underlined. A consensus polyadenylation signal (AATAAA) is underlined by a wavy line. Sequences encompassing each of the domains highlighted in the above schematic diagram are indicated as follows: cysteine residues are shaded; consensus glycosylation sites (Asn-X-Ser/Thr) are underlined by broken lines; the highly acidic region is indicated by solid black underlining; proline rich sequences are underlined by open bars; the putative nuclear localization signals are indicated by a hatched underlining; the potential site for protein kinase A phosphorylation is underlined by a crossed bar; and the two zinc-finger-like sequences are boxed.

Fig. 4. Mechanism of activation of the *vav* oncogene. Comparative nucleotide sequence analysis of the 5' domains of the *vav* proto-oncogene and oncogene cDNA clones. Vertical lines indicate identical nucleotides. The putative initiator codon (ATG) of the *vav* oncogene is boxed. In-frame terminator codons are underlined. The black vertical arrow indicates the breakpoint between pSV2neo and *vav* gene sequences. Diagnostic restriction endonuclease cleavage sites are also indicated.

Fig. 5. Transforming activity of *vav* cDNA clones. Expression plasmids including pSK27, which contains the entire *vav* oncogene cDNA sequences (hatched box), and pSK77, which encompasses a chimeric *vav* gene containing 5' *vav* oncogene (hatched box) and 3' *vav* proto-oncogene (open box) sequences, were tested for their ability to transform NIH3T3 cells in gene transfer assays (Graham and van der Eb, 1975). Sequences derived from pSV2neo are indicated by a solid black box. Coding cDNA sequences are represented by the thick boxes. Non-coding sequences are indicated by thin boxes. Transcriptional regulatory elements (MSV-LTR) and polyadenylation signal (PA) present in the pMEX expression plasmid are also indicated.

Fig. 6. *vav* proto-oncogene expression in human cells. Poly(A)-selected RNAs were isolated from (a) K562; (b) MOLT-4; (c) RPMI-6666; (d) CMS; (e) U937; (f) KG-1; (g) IM-9; (h) CCRF-CEM; (i) A431; (j) HOS; (k) T98G; (l) IMR32; (m) Y79; (n) A172; (o) M413; and (p) SK-N-SH human cell lines. The developmental lineages of each of these cell lines are described in Table I. RNAs (3 µg) were submitted to Northern transfer analysis (Lehrach *et al.* 1977) and hybridized for 48 h under stringent conditions (42°C in 5× SSC, 50% formamide, 1× Denhardt's solution) to ³²P-labeled nick-translated probes specific for the *vav* proto-oncogene cDNA clone (2.8 kb *EcoRI* DNA fragment) and chicken β -actin (2 kb *HindIII* DNA fragment of plasmid B2000, Cleveland *et al.*, 1980). In panel A, the *vav* proto-oncogene probe was stripped from the filter prior to hybridization of the β -actin probe. In panel B, both probes (7 × 10⁷ c.p.m. of each) were mixed in the hybridization reaction. Hybridized filters were exposed to Kodak XAR-5 film at -70°C for 4 h in the presence of intensifier screens. Molecular weight markers are those described in the legend to Figure 1. The migration of the *vav* proto-oncogene and β -actin transcripts is indicated by arrows.

(Table I). Similar results were obtained when normal human B-cells, T-cells (either untreated or PHA-stimulated), monocytes or platelets were used instead of cell lines (Table I). These results indicate that the *vav* proto-oncogene is specifically expressed in cells of the hematopoietic system.

Discussion

The *vav* oncogene was generated by a genetic rearrangement that replaced its 5' domain by sequences derived from the bacterial Tn5 gene present in the cotransfecting pSV2neo DNA used as a selectable marker in gene transfer assays. This rearrangement also resulted in the transcriptional activation of *vav* gene sequences presumably by the neighboring SV40 early promoter contained within pSV2neo. At the present time, we do not know whether the transforming properties of the *vav* oncogene are due to its ectopic expression in NIH3T3 cells, the modification of its amino-terminal domain or a combination of both. Isolation of full-length cDNA clones of the *vav* proto-oncogene should provide conclusive information regarding the mechanism of activation of this human gene.

Analysis of the predicted amino acid sequence of the *vav* gene product reveals a particularly hydrophilic molecule rich in both acidic and basic residues. Glutamic acid is the most abundant amino acid residue (64 residues or 8.4%). Other highly charged amino acid residues are lysine (53 residues), arginine (52 residues) and aspartic acid (46 residues). In total, charged amino acids represent 31.1% of all residues. Acidic residues are concentrated between positions 84 and 128 (23 out of these 45 residues are either glutamic acid or aspartic acid with a net charge of -23) with total exclusion of basic residues. Acidic domains are believed to be involved in protein-protein interactions (Sigler, 1988), thus suggesting that the *vav* gene product might be part of a functional protein complex. In addition, the predicted sequence of the *vav* protein contains two clusters of basic residues that could represent nuclear localization signals (Dingwall and Kaskay, 1986) and a putative site for protein kinase A phosphorylation (Creighton, 1984). These features suggest that the *vav* gene might encode a nuclear phosphoprotein.

The predicted sequence of the *vav* gene product exhibits two zinc-finger-like domains. Putative zinc finger domains have been identified in transcriptional factors such as the yeast GAL4 and the *Xenopus* TFIIIA proteins; several morphogenic gene products of *Drosophila*; the mammalian transcriptional activator SP1; and the steroid and retinoic acid receptor families (Berg, 1986; Evans, 1988; Sunderman and Barber, 1988). However, it has been recently pointed out that zinc fingers might not be exclusive to DNA-binding proteins (Frankel and Pabo, 1988). In the *vav* protein the two zinc-finger-like motifs and the N-terminal acidic region are separated by at least a cluster of prolines (residues 289-292) that are likely to serve as a hinge region separating two domains presumably involved in DNA binding and transcriptional activation. This structural feature has been found in well-characterized transcriptional activators such as the yeast GCN4 and GAL-4 proteins, the *jun* oncogene and the steroid receptor family, independently of whether they mediate their DNA interactions with zinc finger motifs or not (Ptashne, 1988).

Analysis of *vav* proto-oncogene transcripts in cells of various developmental lineages has revealed a truly

Table 1. *vav* proto-oncogene expression in human cells

Hematopoietic cells	<i>vav</i> proto-oncogene expression	Non-hematopoietic cells	<i>vav</i> proto-oncogene expression
Lymphoid		Epithelial	
Raji	+	HeLA	—
IM-9	+	A431	—
Normal B-cells	+	MCF-7	—
CCRF-CEM	+	PA-1	—
MOLT-4	+	TERA2	—
RPMI-6666	+	M413	—
Normal T-cells	+		
Monocytic		Mesenchymal	
CMS	+	HOS	—
KG-1	+	HT-1080	—
U-937	+		
Normal monocytes	+		
Multipotential		Neuroectodermal	
K562	+	U-373 MG	—
HEL	+	HS 683	—
HEL + hemin ^a	+	SK-N-SH	—
HEL + TPA ^b	+	A172	—
		Y79	—
		IMR 32	—
		T98G	—
Other		Other	
Normal platelets	+	HuAC11	—

^aErythroid lineage.^bMonocytic lineage.

remarkable pattern of expression. Whereas no *vav* mRNA could be detected in cell lines of epithelial, mesenchymal or neuroectodermal origin, high levels of *vav* proto-oncogene expression were observed in each of the hematopoietic cells tested, including those of lymphoid, myeloid and erythroid lineages. The ubiquitous expression of the proto-*vav* gene in hematopoietic cells, along with the resemblance of its gene product to transcriptional factors, raise the possibility that this locus might be involved in the transcriptional machinery required for the proliferative maintenance of the hematopoietic system.

Materials and methods

Cell lines and gene transfer assays

Human cell lines were obtained from the American Type Culture Collection (ATCC), except for CMS (Monaco *et al.*, 1982) and HuAC11 (Peebles *et al.*, 1973). Platelets, and B and T lymphocytes were isolated as described (Timmen and Saksela, 1980; Abe *et al.*, 1986). Isolation of monocytes was performed by using counterflow centrifugal elutriation (Wahl *et al.*, 1983). HEL cells were treated with the phorbol ester TPA and with hemin according to published procedures (Papayannopoulou *et al.*, 1983; Larson and Papayannopoulou, 1985). NIH3T3 mouse cells were transfected with 20 µg of high mol. wt cellular DNA or with 1 µg of plasmid DNA by the calcium phosphate precipitation technique (Graham and van der Eb, 1973). Foci of transformed cells were scored after 10–14 days. Tumorigenicity assays were performed as described (Blair *et al.*, 1982; Fasano *et al.*, 1984). Tumor appearance was followed up to 2 months after inoculation.

Isolation of genomic and cDNA clones

A *vav* genomic library was made in λEMBL-4. DNA isolated from a third-cycle *vav*-induced nude mouse tumor was partially digested with *Sau3A*I and fractionated on sucrose gradients. DNA of 15–20 kb was ligated to *Bam*HI-digested λEMBL-4 DNA. 10⁶ phage clones were screened by

hybridization with ³²P-labeled nick-translated high mol. wt human DNA under stringent conditions (42°C in 5 × SSC, 50% formamide, 1 × Denhardt's solution). Filters were washed three times at room temperature with 2 × SSC, 0.1% SDS and twice at 55°C with 0.1 × SSC, 0.1% SDS. The *vav* cDNA library in λgt10 was prepared from poly(A)-selected RNA of a third-cycle *vav*-induced nude mouse tumor using a cDNA cloning kit (Amersham). One million phages were screened under stringent conditions using a ³²P-labeled nick-translated 800 bp *Bam*HI–*Eco*RI *Ala*⁺ *vav* genomic DNA fragment as a probe. Among those recombinant λgt10 phages isolated, that containing the longest (2.8 kb) insert was selected for further studies. Its insert was subcloned into the *Eco*RI site of Bluescript KS in both orientations to generate pSK33 and pSK47. A λgt10 cDNA library (2 × 10⁶ clones) prepared from human K562 cells (Shtivelman *et al.*, 1985) was hybridized under stringent conditions to a ³²P-labeled nick-translated 2.8 kb *vav* cDNA probe. A recombinant phage carrying a 2.9 kb *vav* proto-oncogene insert was isolated and subcloned in Bluescript KS in both orientations to generate pSK65 and pSK66.

Southern and Northern transfer analysis

High mol. wt DNA was digested to completion with appropriate restriction endonucleases, electrophoresed in 0.8% agarose gel and submitted to Southern transfer analysis as described (Southern, 1975). Total cellular RNA was extracted by the guanidium thiocyanate method (Chirgwin *et al.*, 1979) and purified by centrifugation through cesium chloride. Poly(A)-containing RNA was isolated by retention on oligo(dT) columns (Collaborative Research). Total RNA (10 µg) or poly(A)-selected RNA (3 µg) were submitted to Northern transfer analysis (Lehrach *et al.*, 1977). The nitrocellulose filters were hybridized to various ³²P-labeled nick-translated probes for 48 h under stringent conditions (42°C in 5 × SSC, 50% formamide, 1 × Denhardt's solution).

Nucleotide sequencing

A series of nested deletions were generated from pSK33 and pSK47 by the combined use of Exonuclease III and Mung bean nuclease (Stratagene). *Escherichia coli* MV1193 cells were transformed with the corresponding deletion mutants and single-strand phages rescued by subsequent infection with the helper M13 K07 phage. Single-stranded DNAs were prepared from

these phages and submitted to nucleotide sequence analysis by an automated chain termination method using primers with multiple fluorophores (Brumbaugh *et al.*, 1988). To determine the sequence of the 5' domain of the normal *vav* gene, a 270 bp *SacI* fragment from pSK65 was subcloned in both orientations in Bluescript KS. Single-stranded DNA was obtained from these plasmids and used as a template for sequencing by the dideoxynucleotide chain termination technique (Sanger *et al.*, 1977) using the Sequenase kit (USB).

Expression plasmids

pSK27 was generated by subcloning the entire 2.8 kb *EcoRI* *vav* oncogene cDNA insert of pSK33 into the unique *EcoRI* site of pMEX. pMEX is a mammalian expression vector in which the polylinker sequences of pUC18 (minus the *HindIII* recognition site) are flanked by a Moloney MSV-LTR and the polyadenylation signal of SV40. pSK77 was generated by subcloning the 5' 270 bp *EcoRI*–*SacI* DNA fragment of the *vav* oncogene and the 3' 2.6 kb *SacI*–*EcoRI* DNA fragment of the *vav* proto-oncogene into the *EcoRI* site of pMEX.

Acknowledgements

We are indebted to J.Brumbaugh and D.Conway for their help in determining the nucleotide sequence of the *vav* oncogene. We would also like to thank R.Montesano for providing tumor samples, E.Canaani for the K562 cDNA library, M.Ernst for help with gene transfer assays, and F.Coulier, S.Hughes and B.Stanton for helpful discussions. Research sponsored by the National Cancer Institute, DHHS, under contract no. N01-CO-74101 with Bionetics Research, Inc. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

- Abe, H., Rossio, J.L., Ruscehi, F.W., Matsushima, K. and Oppenheim, J.J. (1986) *J. Immunol. Methods*, **90**, 111–123.
- Berg, J.M. (1986) *Science*, **232**, 485–487.
- Bishop, M. (1987) *Science*, **235**, 305–310.
- Blair, D.G., Cooper, C.S., Oskarsson, M.K., Eader, L.A. and Vande Woude, G.F. (1982) *Science*, **281**, 1122–1125.
- Brumbaugh, J.A., Mitterdorf, L.R., Grone, D.L. and Ruth, J.L. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 5610–5614.
- Chirgwin, J., Przybyla, A., MacDonald, R. and Rutter, W. (1979) *Biochemistry*, **18**, 5294–5299.
- Cleveland, D.W., Lopata, M.A., MacDonald, R.J., Cowan, N.J., Rutter, W.J. and Kirschner, M.W. (1980) *Cell*, **20**, 95–105.
- Creighton, T.F. (1984) *Proteins: Structural and Molecular Principles*. Freeman, New York, pp. 83–84.
- Culp, J.S., Webster, L.C., Friedman, D.J., Smith, C.L., Huang, W.-J., Wu, F.Y.-H., Rosenberg, M. and Ricciardi, R.P. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6450–6454.
- Delli Bovi, P., Curatola, A.M., Kern, F.G., Greco, A., Ihmarm, M. and Basilico, C. (1987) *Cell*, **50**, 729–737.
- Dingwall, C. and Kaskey, R.A. (1986) *Annu. Rev. Cell. Biol.*, **2**, 367–390.
- Eva, A. and Aaronson, S.A. (1985) *Nature*, **316**, 273–275.
- Evans, R. (1988) *Science*, **240**, 889–895.
- Fasano, O., Birnbaum, D., Edlund, L., Fogh, J. and Wigler, M. (1984) *Mol. Cell. Biol.*, **4**, 1695–1705.
- Frankel, A.D. and Pabo, C.O. (1988) *Cell*, **53**, 675.
- Fukui, M., Yamamoto, T., Kawai, S., Marvo, K. and Toyoshima, K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5954–5958.
- Graham, F.L. and van der Eb, A.J. (1973) *Virology*, **52**, 456–467.
- Ikawa, S., Fukui, M., Veyama, Y., Tamaok, N., Yamamoto, T. and Toyoshima, K. (1988) *Mol. Cell. Biol.*, **8**, 2651–2654.
- Kozma, S.C., Redmond, S.M.S., Xiao-Chung, F., Saurer, S.M., Groner, B. and Hynes, N.E. (1988) *EMBO J.*, **7**, 147–154.
- Larson, A.E. and Papayannopoulou, T. (1985) *Exp. Hematol.*, **13**, 129–135.
- Lehrach, H., Diamond, D., Wozney, J.M. and Boedtker, H. (1977) *Biochemistry*, **16**, 4743–4751.
- Martin-Zanca, D., Hughes, S. and Barbacid, M. (1986) *Nature*, **319**, 743–748.
- Monaco, G., Vigneti, E., Lancieri, M., Cornaglia-Ferraris, P., Lambertenghi-Delilieri, G. and Revoltella, R. (1982) *Cancer Res.*, **42**, 4182–4189.
- Muller, R. and Muller, D. (1984) *EMBO J.*, **3**, 1121–1127.
- Nakamura, T., Hillova, J., Mariage-Samson, R. and Hill, M. (1988) *Oncogene Res.*, **2**, 357–370.
- Papayannopoulou, T., Nakamoto, B., Yokochi, T., Chait, A. and Kannagi, R. (1983) *Blood*, **62**, 832–845.
- Peebles, P.T., Fischinger, P.J., Bassin, R.H. and Papageorge, A.G. (1973) *Nature New Biol.*, **242**, 98–101.
- Ptashne, P. (1988) *Nature*, **335**, 683–689.
- Rapp, U.R., Reynolds, F.H. and Stephenson, J.R. (1983) *J. Virol.*, **45**, 914–924.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Shimizu, K., Nakatsu, Y., Sekiguchi, M., Hokamura, K., Tanaka, K., Terada, M. and Sugimura, T. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 5641–5645.
- Shtivelman, E., Lifshitz, B., Gale, R.P. and Canaani, E. (1985) *Nature*, **315**, 550–554.
- Sigler, P.B. (1988) *Nature*, **333**, 210–211.
- Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
- Southern, P.J. and Berg, P. (1982) *J. Mol. Appl. Gen.*, **1**, 327–341.
- Sunderman, F.W. and Barber, A.M. (1988) *Ann. Clin. Lab. Sci.*, **18**, 267–288.
- Taira, M., Yoshida, T., Miyagawa, K., Sakamoto, H., Terada, M. and Sugimura, T. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 2980–2984.
- Takahashi, M., Ritz, J. and Cooper, G.M. (1985) *Cell*, **42**, 581–588.
- Timmen, T. and Saksela, E. (1980) *J. Immunol. Methods*, **36**, 285–291.
- Wahl, L.M., Katona, I.M., Wilder, R.L., Winter, C.C., Haroui, B., Scher, I. and Wahl, S.M. (1984) *Cell. Immunol.*, **85**, 373–383.
- Young, D., Waitches, G., Birchmeier, C., Fasano, O. and Wigler, M. (1986) *Cell*, **45**, 711–719.
- Zhan, X., Bates, B., Hu, X. and Goldfarb, M. (1988) *Mol. Cell. Biol.*, **8**, 3487–3495.

Received February 20, 1989; revised March 29, 1989