# A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences

## Dionisio Martin-Zanca*, Stephen H. Hughes† & Mariano Barbacid*

* Developmental Oncology Section and † Gene Expression Section, Basic Research Program, Frederick Cancer Research Facility, Frederick, Maryland 21701, USA

*A biologically active complementary DNA clone of a transforming gene present in a human colon carcinoma contains gene sequences of both tropomyosin and a previously unknown protein tyrosine kinase. The predicted protein (641 amino acids) encoded by this oncogene seems to have been formed by a somatic rearrangement that replaced the extracellular domain of a putative transmembrane receptor by the first 221 amino acids of a non-muscle tropomyosin molecule.*

TRANSFECTION of NIH 3T3 cells with genomic DNAs has allowed the identification of transforming genes in human neoplasms and in carcinogen-induced animal tumours (for recent reviews, see refs 1-3). Most of these oncogenes are members of the *ras* gene family. However, other transforming genes without apparent sequence homology to *ras* genes have been identified: those of human origin include *onc*D (carcinoma of the ascending colon[4]), *met* (MNNG-HOS cells[5]), *mel* (NK14 melanoma cells[6]), *dbl* (diffuse B-cell lymphoma[7]), and a transforming allele of c-*raf*-1 (carcinoma of the stomach[8]). Two additional oncogenes, designated *mcf*-3 (ref. 9) and *ret*[10], became activated during transfection of NIH 3T3 cells with human DNA. Finally, a series of non-*ras* transforming genes which exhibit differentiation-stage specificity have been consistently identified by Cooper *et al.*[11] in a variety of human tumours.

During gene transfer assays aimed at detecting transforming genes in human tumours, our laboratory identified four different transforming genes (designated *onc*A to *onc*D) based on their distinct pattern of *Alu*-positive restriction endonuclease fragments[4]. Three of these oncogenes (*onc*A, B and C) were identified as transforming alleles of the cellular Ha-*ras*-1, Ki-*ras*-2 and N-*ras* loci; however, *onc*D did not show detectable sequence homology with members of the *ras* gene family even under relaxed hybridization conditions[4,12]. *onc*D exhibits a genetic complexity of at least 30 kilobases (kb), as all tertiary NIH 3T3 transformants analysed contain two *Alu*-positive *Eco*RI DNA fragments of 25 and 4 kb (see Fig. 2 of ref. 4). Here, we describe the isolation of a biologically active complementary DNA clone of *onc*D and its complete nucleotide sequence analysis. Information derived from these studies has allowed us to determine that *onc*D was generated by a somatic rearrangement that resulted in the juxtaposition of two truncated loci, one of which codes for a non-muscle tropomyosin and the other for a putative tyrosine-specific protein kinase.

## Molecular cloning of *onc*D cDNA

Three overlapping λ recombinant clones representing about 20 kb of *onc*D were isolated from a genomic library prepared from 106-63 cells, a third-cycle NIH 3T3 transformant derived from *onc*D (D.M.-Z., B. Cabrer and M.B., unpublished). Several *Alu*-negative DNA fragments were subcloned in vector pUC18 and used as probes to identify *onc*D-specific transcripts in NIH 3T3 transformants. Two of these subclones, pDM6, which contained a 1.4-kb *Bam*HI—*Eco*RI insert, and pDM8, which carried a 2.7-kb *Kpn*I insert, identified a 2.5-kb poly(A)-containing messenger RNA in *onc*D-derived transformants. Significant levels of mRNA of this size were also found in normal NIH 3T3 cells as well as in *ras*-derived transformants (data not shown).

A cDNA library consisting of 10⁶ clones was prepared as described by Helfman *et al.*[13]. DH5 *Escherichia coli* cells were transformed with 25 ng of cDNA prepared by reverse transcription of polyadenylated mRNA isolated from 106-63 cells and ligated to pUC18 and pUC19 vectors using *Sal*I and *Eco*RI linkers. Nitrocellulose filters containing these colonies were hybridized with the 1.4-kb *Bam*HI-*Eco*RI insert of pDM6 and with the 2.7-kb *Kpn*I insert of pDM8 as described previously[14]. About 200 hybridizing colonies were selected for further analysis.

## Presence of non-muscle tropomyosin sequences

The existence of *onc*D-related polyadenylated mRNA in untransformed NIH 3T3 cells suggested that a significant fraction of the *onc* D-positive cDNA clones might represent normal mouse cellular transcripts. Partial nucleotide sequence analysis of the pDM6 and pDM8 inserts was used to identify a domain that did not hybridize with NIH 3T3 mRNAs. Comparison of pDM6 and pDM8 *onc*D sequences with those stored in gene databanks revealed that a stretch of 204 base pairs (bp) of pDM6 and 66 bp of pDM8 were almost identical (>96% homology) to a clone (hTM$_{NM}$-1) of a human non-muscle tropomyosin pseudogene chracterized previously[15] (Fig. 1). Moreover, these regions had open reading frames whose deduced amino-acid sequences also exhibited striking homology (over 90%, taking into account conserved amino-acid substitutions) to the amino terminus (residues 1-43 in the case of pDM6) and to an internal domain (residues 177-197 in the case of pDM8) of a tropomyosin molecule isolated from horse platelets[16] (Fig. 1). Finally, pDM6 and pDM8 sequences located at the 3' boundaries of their respective regions of homology to hHTM$_{NM}$-1, closely matched those of consensus donor splicing sites (not shown). These results suggested strongly that pDM6 and pDM8 contained the first coding exon and an internal exon, respectively, of a non-muscle tropomyosin gene. Tropomyosins are known to be expressed at high levels in most cell types, explaining the existence of abundant *onc*D-related transcripts in untransformed NIH 3T3 cells.

The 204-bp region of pDM6 that is homologous to hTM$_{NM}$-1 includes 70 bp of 5'-untranslated sequence[5] (Fig. 1). Noncoding sequences evolve faster than their translated counterparts. Thus, we reasoned that this 70-bp segment of pDM6 might have diverged sufficiently from its murine counterpart to allow identification of cDNA clones carrying *onc*D sequences from those containing mouse tropomyosin transcripts. The 70-bp 5'-untranslated sequences were isolated as part of a *Bam*HI-*Nco*I DNA fragment of pDM6 which extended from the 5' *Bam*HI boundary of the insert to the convenient *Nco*I site which contains the putative initiator codon (ATG) of the tropomyosin gene
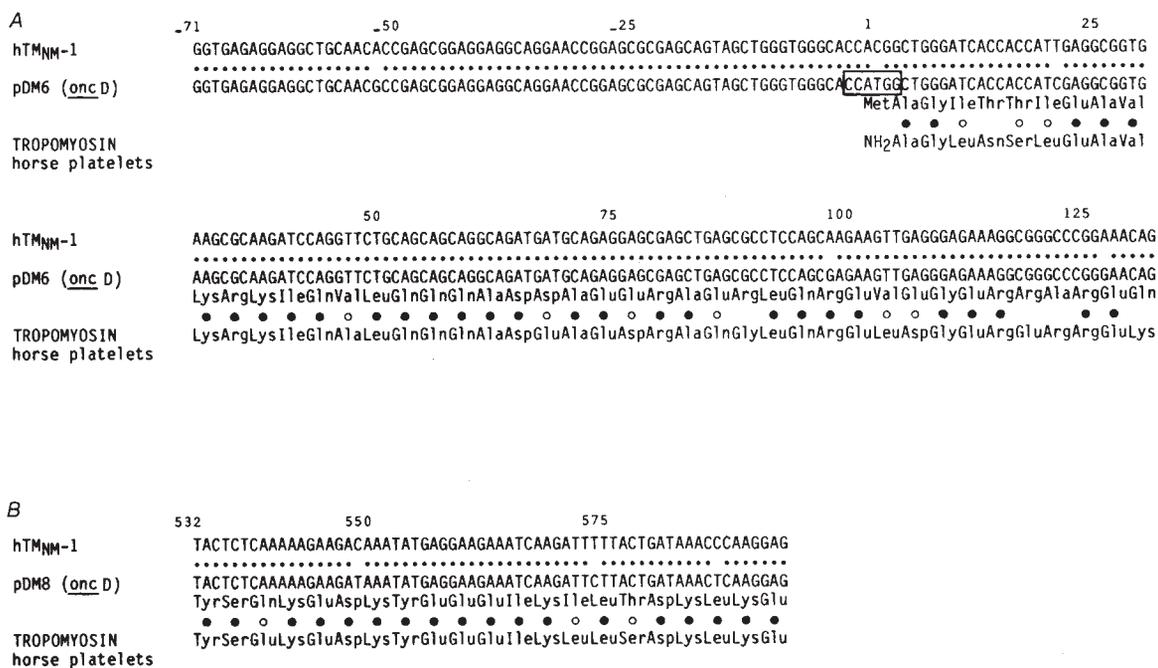
**Fig. 1** Nucleotide sequence homology between pDM6 (*A*) and pDM8 (*B*) subclones of *onc*D and hTM$_{NM}$-1, a clone of a human non-muscle tropomyosin pseudogene[15]. Dots indicate base pairs that are identical in the two sequences. Numbers correspond to numbering of nucleotides in the hTM$_{NM}$-1 sequence, with +1 as the deoxyadenosine residue of the putative initiator codon (ATG) which has been replaced by a non-functional ACG triplet in hTM$_{NM}$-1. The box indicates the convenient *Nco*I site used in generating the 209-bp *Bam*HI–*Nco*I probe used to identify *onc*D-specific cDNA clones (see text). The amino-acid sequences deduced from the open reading frames of pDM6 (nucleotides 1–132) and pDM8 are compared with the sequence of a tropomyosin molecule isolated from horse platelets[16]. ●, ○, Identical and conserved amino-acid residues, respectively.

(Fig. 1). Of the 200 positive cDNA clones, only 11 specifically hybridized with this probe; these colonies were picked, grown up and their plasmids isolated by standard techniques[17]. One of the plasmids, pDM10-1, contained a 2.3-kb *Sal*I–*Eco*RI DNA insert which is approximately the size of the mRNAs identified in 106-63 cells by *onc*D specific probes.

## Nucleotide sequence of *onc*D cDNA

The complete nucleotide sequence of the 2,301-bp *Sal*I–*Eco*RI insert of pDM10-1 was established by the method of Maxam and Gilbert[18] (see Fig. 2). pDM10-1 contains a 1,923-bp open reading frame capable of directing the synthesis of a polypeptide of 641 amino acids. The open reading frame is flanked by 232 bp of 5'-untranslated sequence and 143 bp of 3'-noncoding sequence in which no putative polyadenylation signal could be identified. It is possible that pDM10-1 does not contain all of the untranslated sequences of *onc*D. However, based on the size of the mRNA, only 100–200 bp are absent.

The most relevant features of this sequence are: (1) nucleotides −232 to −64 (position +1 corresponds to the deoxyadenosine of the initiator ATG codon) represent a stretch of G+A-rich sequences in which the sequence GGAGGAGCA is repeated 12 times (with one imperfection in two cases). This repeat is interrupted by the hexanucleotide GGA$^C_A$CA on three occasions and by repeated GGA triplets on two occasions. A similar arrangement of repetitive sequences (a GGGGCAGGA sequence interrupted by the hexanucleotide GCAGGA) exists in the simple IR3 (internal repeat 3) repeat array of Epstein–Barr virus[19]. The biological significance of this domain of the *onc* cDNA is being investigated. (2) Sequences located between positions −63 and +664 correspond to the first seven coding exons of a non-muscle tropomyosin gene, as determined by comparison of these sequences with those of genomic clones of the tropomyosin gene (A. MacLeod, personal communication)

and of *onc*D (our unpublished observations). The boundaries between each of the exons are indicated by vertical arrowheads in Fig. 2. Nucleotides −63 to −1 correspond to the sequences used to identify pDM10-1 and represent untranslated sequences of the first coding exon of the tropomyosin locus present in *onc*D. (3) The homology between pDM10-1 and genomic tropomyosin sequences ends abruptly at nucleotide +664, which corresponds to the last base of the tropomyosin exon 7. This result suggests that *onc*D is a hybrid gene generated by a genetic rearrangement that disrupted a non-muscle tropomyosin gene between coding exons 7 and 8. (4) Sequences located between nucleotides +1,015 and +1,773 share extensive homology with retroviral oncogenes coding for tyrosine-specific protein kinases such as v-*abl*, v-*erb*-B, v-*fes*/*fps*, v-*fgr*, v-*fms*, v-*ros*, v-*src* and v-*yes*[20]. By introducing appropriate gaps, the homology between the sequences coding for the catalytic domains of these retroviral protein tyrosine kinases and the *onc*D sequence was as high as 51%. Similar results have been obtained with cellular genes known to possess tyrosine-specific protein kinase activity, for example those encoding the insulin[21] and epidermal growth factor (EGF)[22] receptors. Finally, a lower degree of homology was observed when these *onc*D sequences were compared with retroviral oncogenes whose products exhibit serine/threonine-specific protein kinase activity, for example, v-*mos* and v-*raf*[20]. Taken together, these results strongly suggest that the carboxy-terminal moiety of *onc*D was derived from a locus that codes for a tyrosine-specific protein kinase.

## Protein tyrosine kinase gene

Figure 3 illustrates the striking homology between the carboxy-terminal moiety of the putative product of *onc*D and the common catalytic domain of retroviral and cellular tyrosine-specific protein kinases. The putative *onc*D product has 63 of the 70 amino-acid residues common to all known protein tyrosine kinases,
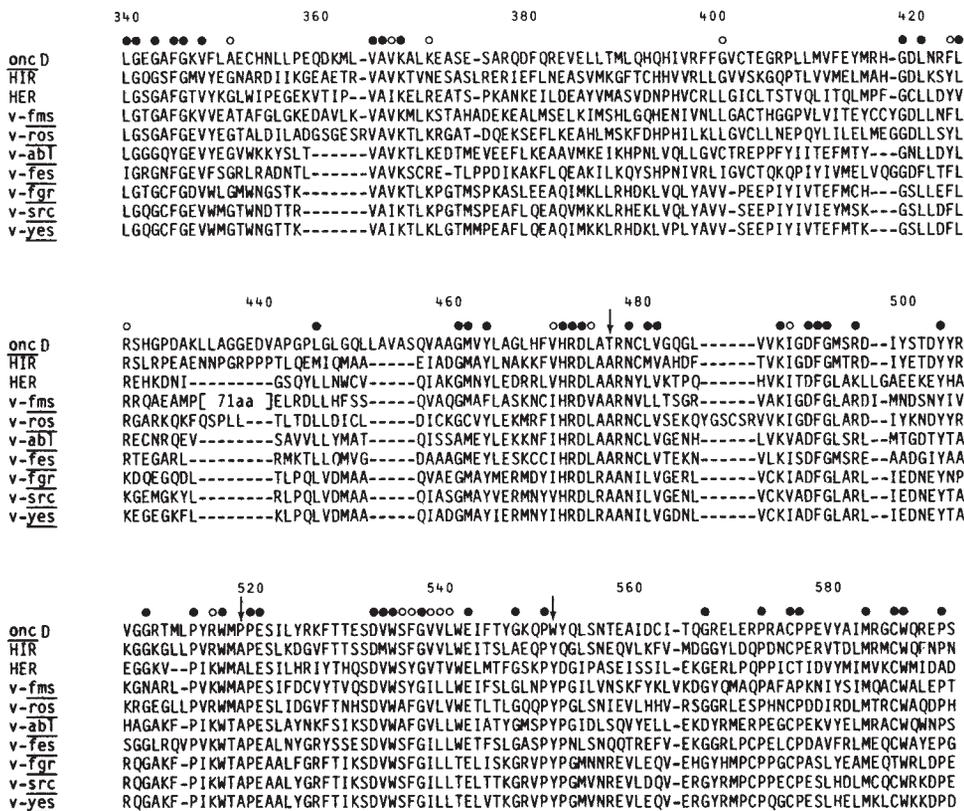
**Fig. 2** Molecular characterization of a cDNA clone of oncD. A, Schematic diagram of clone pDM10-1, showing the SalI and EcoRI cleavage sites that delimit the boundaries of the pDM10-1 insert, and the NcoI cleavage sites, one of which (near the 5' end) contains the initiator ATG codon. The solid line indicates untranslated sequences; the large box represents coding sequences, in which those derived from tropomyosin (□) and the putative receptor gene (▨) are indicated. The transmembrane domain (■, TM) and the region exhibiting sequence homology with tyrosine-specific protein kinases (■, TK) are superimposed. B, Nucleotide sequencing strategy. C, Nucleotide sequence of the 2,301-bp SalI–EcoRI insert of pDM10-1 and the deduced amino-acid sequence. Nucleotides are numbered from the deoxyadenosine residue of the initiator ATG codon and amino-acid residues from the presumed initiating methionine. The solid black underlining indicates the putative transmembrane domain. Amino-acid residues shared by other tyrosine-specific protein kinases are indicated by thin underlining. Boxed residues are those shared by other protein tyrosine kinases, but not by oncD. Sequences derived from the tropomyosin gene are separated from the putative receptor locus by horizontal arrows. Vertical arrowheads indicate the limits of the seven tropomyosin coding exons.

including 28 of the 31 residues shared between these protein tyrosine kinases and, the prototype serine/threonine-specific protein kinases[23]. The conserved residues include Tyr 503 (Tyr 416 in pp60$^{v\text{-src}}$), which lies within the catalytic domain and is the target of autophosphorylation reactions, and Lys 367 (Lys 295 in pp60$^{v\text{-src}}$), which forms part of the ATP-binding site. oncD Lys 367 is located 21 residues from the glycine-rich sequence Gly-Glu-Gly-Ala-Phe-Gly, as in other protein kinases[23]. Other highly conserved residues shared by known protein kinases and believed to be important for catalysis, for example, Arg-Asp-Leu (residues 472–474) and Asp-Phe-Gly (residues 491–493), are also present in the putative oncD protein (Fig. 3).

However, the predicted oncD product exhibits some unique features. Three residues common to all known protein tyrosine kinases, that is, Ala 389, Ala 430 and Tyr 463 in pp60$^{v\text{-src}}$ (ref. 22), have been replaced by Thr 476, Pro 518 and Trp 551, respectively, in oncD. Interestingly, the substitution of Ala 430 with valine drastically decreases the protein tyrsoine kinase activity of pp60$^{v\text{-src}}$ (ref. 24). In addition, the putative oncD protein contains five additional residues, Leu-Ala-Val-Ala-Ser (positions 450–454, Fig. 3), in the centre of the catalytic domain. These observations suggests that oncD contains a novel protein tyrsoine kinase gene.

The homology of the putative oncD product also extends to prototype cellular protein tyrosine kinases such as the EGF[22] and insulin[21] receptors (Fig. 3). Analysis of amino-acid sequences in domains other than the amino-terminal tropomyosin moiety (amino-acid residues 1–221) and the catalytic protein tyrosine kinase domain (amino-acid residues 339–591) revealed the presence of a long stretch of 25 hydrophobic amino acids (residues 237–262) surrounded by basic residues, a feature highly indicative of a transmembrane domain (Figs 2,3). Thus, the non-tropomyosin moiety of oncD might have been contributed by a gene coding for a growth factor receptor with protein tyrosine kinase activity.

Direct comparison of oncD sequences outside the tropomyosin domain with those of the genes coding for the EGF[22] and insulin[21] receptors clearly established that neither of these two loci was involved in the generation of oncD. A similar statement could be made for the c-abl locus whose partial sequence has already been published[25]. However, this is not the case for the human homologues of other retroviral protein tyrosine kinase oncogenes with which oncD shares extensive sequence homology. To determine whether oncD contains a truncated homologue of a known retroviral oncogene or a novel protein tyrosine kinase locus, human DNA was hybridized with probes specific for retroviral oncogenes known to code for tyrosine-specific protein kinases, including v-fes, v-fgr, v-fms, v-ros, v-src

and v-yes[20] (Fig. 4). The resulting DNA fragments were compared with those obtained using a probe specific for the receptor domain of oncD. Figure 4 shows a representative experiment. The pattern of PstI or PvuII fragments characteristic of the oncD receptor domain was clearly distinct from those obtained with the retroviral probes. These hybridization experiments, together with the amino-acid sequence comparisons discussed above, clearly establish that the putative protein tyrosine kinase receptor gene involved in the generation of oncD represents a novel human locus. Whether this gene codes for the receptor of a known growth factor is an interesting possibility that must await identification and biochemical characterization of the normal gene product.

```
         340              360              380              400              420
        •• • •• •  ○     •••○• ○                              ○                    • •  ○•
oncD    LGEGAFGKVFLAECHNLLPEQDKML-VAVKALKEASE-SARQDFQREVELLTMLQHQHIVRRFFGVCTEGRPLLMVFEYMRH-GDLNRFL
HIR     LGQGSFGMVYEGNARDIIKGEAETR-VAVKTVNESASLRERIEFLNEASVMKGFTCHHVVRLLGVVSKGQPTLVVMELMAH-GDLKSYL
HER     LGSGAFGTVYKGLWIPEGEKVTIP--VAIKELREATS-PKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQLMPF-GCLLDYV
v-fms   LGTGAFGKVVEATAFGLGKEDAVLK-VAVKMLKSTAHADEKEALMSELKIMSHLGQHENIVNLLGACTHGGPVLVITEYCCYGDLLNFL
v-ros   LGSSAFGEVYEGTALDILADGSGESRVAVKTLKRGAT-DQEKSEFLKEAHLMSKFDHPHILKLLGVCLLNEPQYLILELMEGGDLLSYL
v-abl   LGGGQYGEVYEGVWKKYSLT------VAVKTLKEDTMEVEEFLKEAAVMKEIKHPNLVQLLGVCTREPPFYIITEFMTY---GNLLDYL
v-fes   IGRGNFGEVFSGRLRADNTL------VAVKSCRE-TLPPDIKAKFLQEAKILKQYSHPNIVRLIGVCTQKQPYIVMELVQGGDFLTFL
v-fgr   LGTGCFGDVWLGMWNGSTK-------VAVKTLKPGTMSPKASLEEAQIMKKLRHDKLVQLYAVV-PEEPIYIVTEFMCH---GSLLEFL
v-src   LGQGCFGEVWMGTWNDTTR-------VAIKTLKPGTMSPEAFLQEAQVMKKLRHEKLVQLYAVV-SEEPIYIVIEYMSK---GSLLDFL
v-yes   LGQGCFGEVWMGTWNGTTK-------VAIKTLKLGTMMPEAFLQEAQIMKKLRHDKLVPLYAVV-SEEPIYIVTEFMTK---GSLLDFL

         440              460              480                          500
        ○              •             •• •    ○••••○ ↓ • ••           •○ ••• •        •
oncD    RSHGPDAKLLAGGEDVAPGPLGLGQLLAVASQVAAGMVYLAGLHFVHRDLATRNCLVGQGL------VVKIGDFGMSRD--IYSTDYYR
HIR     RSLRPEAENNPGRPPPTLQEMIQMAA-----EIADGMAYLNAKKFVHRDLAARNCMVAHDF------TVKIGDFGMTRD--IYETDYYR
HER     REHKDNI---------GSQYLLNWCV-----QIAKGMNYLEDRRLVHRDLAARNYLVKTPQ------HVKITDFGLAKLLGAEEKEYHA
v-fms   RRQAEAMP[ 71aa ]ELRDLLHFSS-----QVAQGMAFLASKNCIHRDVAARNVLLTSGR------VAKIGDFGLARDI-MNDSNYIV
v-ros   RGARKQKFQSPLL---TLTDLLDICL-----DICKGCVYLEKMRFIHRDLAARNCLVSEKQYGSCSRVVKIGDFGLARD--IYKNDYYR
v-abl   RECNRQEV--------SAVVLLYMAT-----QISSAMEYLEKKNFIHRDLAARNCLVGENH------LVKVADFGLSRL--MTGDTYTA
v-fes   RTEGARL---------RMKTLLQMVG-----DAAAGMEYLESKCCIHRDLAARNCLVTEKN------VLKISDFGMSRE--AADGIYAA
v-fgr   KDQEGQDL--------TLPQLVDMAA-----QVAEGMAYMERMDYIHRDLRAANILVGERL------VCKIADFGLARL--IEDNEYNP
v-src   KGEMGKYL--------RLPQLVDMAA-----QIASGMAYVERMNYVHRDLRAANILVGENL------VCKVADFGLARL--IEDNEYTA
v-yes   KEGEGKFL--------KLPQLVDMAA-----QIADGMAYIERMNYIHRDLRAANILVGDNL------VCKIADFGLARL--IEDNEYTA

         520              540              560              580
        •  • •○• ↓•○     •••○○•○○○ •  • •↓                •    •  ••     •  •• •
oncD    VGGRTMLPYRWMPPESILYRKFTTESDVWSFGVVLWEIFTYGKQPWYQLSNTEAIDCI-TQGRELERPRACPPEVYAIMRGCWQREPS
HIR     KGGKGLLPVRWMAPESLKDGVFTTSSDMWSFGVVLWEITSLAEQPYQGLSNEQVLKFV-MDGGYLDQPDNCPERVTDLMRMCWQFNPN
HER     EGGKV--PIKWMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIPASEISSIL-EKGERLPQPPICTIDVYMIMVKCWMIDAD
v-fms   KGNARL-PVKWMAPESIFDCVYTVQSDVWSYGILLWEIFSLGLNPYPGILVNSKFYKLVKDGYQMAQPAFAPKNIYSIMQACWALEPT
v-ros   KRGEGLLPVRWMAPESLIDGVFTNHSDVWAFGVLVWETLTLGQQPYPGLSNIEVLHHV-RSGGRLESPHNCPDDIRDLMTRCWAQDPH
v-abl   HAGAKF-PIKWTAPESLAYNKFSIKSDVWAFGVLLWEIATYGMSPYPGIDLSQVYELL-EKDYRMERPEGCPEKVYELMRACWQWNPS
v-fes   SGGLRQVPVKWTAPEALNYGRYSSESDVWSFGILLWETFSLGASPYPNLSNQQTREFV-EKGGRLPCPELCPDAVFRLMEQCWAYEPG
v-fgr   RQGAKF-PIKWTAPEAALFGRFTIKSDVWSFGILLTELISKGRVPYPGMNNREVLEQV-EHGYHMPCPPGCPASLYEAMEQTWRLDPE
v-src   RQGAKF-PIKWTAPEAALYGRFTIKSDVWSFGILLTELTTKGRVPYPGMVNREVLDQV-ERGYRMPCPPECPESLHDLMCQCWRKDPE
v-yes   RQGAKF-PIKWTAPEAALYGRFTIKSDVWSFGILLTELVTKGRVPYPGMVNREVLEQV-ERGYRMPCPQGCPESLHELMKLCWKKDPD
```

**Fig. 3** Amino-acid homology of the putative oncD protein with the catalytic domain of tyrosine-specific protein kinases. Amino-acid residues 339–592 of the oncD protein were aligned with those corresponding to the homologous regions of the human insulin receptor (HIR), the human EGF receptor (HER) and the retroviral oncogenes v-fms, v-ros, v-abl, v-fes, v-fgr, v-src and v-yes. ●, oncD residues shared by at least eight of the nine protein tyrosine kinases listed; ○, conserved amino-acid substitutions. Vertical arrows indicate those residues shared by all other protein tyrosine kinases, but not by oncD.

## Transforming activity of oncD cDNA

To establish whether pDM10-1 contains a functional cDNA clone of oncD, we investigated its potential transforming activity in gene transfer assays. For this purpose we constructed two plasmids, designated pDM14 and pDM16, in which the coding sequences of pDM10-1 were linked to a human metallothionein (IIA) promoter (derived from p84H[26], given by Michael Karin) and to a Moloney murine sarcoma virus (M-MSV) long terminal repeat (LTR) (derived from pm1SP[27], given by G. Vande Woude) (Fig. 5). In addition, pDM14 and pDM16 contained a polyadenylation signal derived from simian virus 40 (SV40)[28]. As shown in Fig. 5, pDM16 induced the appearance of $2 \times 10^4$ foci per μg of insert DNA when added to NIH 3T3 cells in gene transfer assays[29], a transforming activity comparable to that of DNA clones of acute transforming retroviruses and cellular ras oncogenes; the transforming activity of pDM14 ($6 \times 10^3$ focus-forming units (FFU) per μg) was somewhat lower. Whether this result reflects the more limited promoting activity of the metallothionein IIA regulatory sequences or is due to the fact that this construct does not contain any of the 232 bp of 5'-untranslated sequence of pDM10-1, remains to be determined. In this regard, it is interesting to note that pDM10-1 is capable of inducing low levels of NIH 3T3 transformation without the addition of a eukaryotic promoter (Fig. 5). We are presently investigating whether the G/A-rich domain of pDM10-1 has low levels of promoter activity. In any case, the results summarized in Fig. 5 establish that pDM10-1 contains a functional cDNA clone of oncD.

## A somatic rearrangement generated oncD

As oncD was identified by transfecting human tumour DNA into NIH 3T3 cells, it was important to establish whether the genetic rearrangement responsible for its malignant activation occurred during experimental manipulations or was already present in the original colon carcinoma (tumour 2033). Two oncD probes capable of recognizing DNA fragments encompassing the breakpoints of the two loci present in oncD were

generated. As shown in Fig. 6A, the tropomyosin-specific probe (2.7-kb KpnI insert of pDM8) should detect an 11-kb EcoRI DNA fragment in normal human DNA and a 23-kb EcoRI fragment in DNAs containing oncD. Similarly, a probe derived from the kinase domain of oncD cDNA (1.2-kb BalI–EcoRI insert of pDM10-1) should recognize a normal 14-kb BamHI DNA fragment of human DNA and a rearranged 7-kb BamHI DNA fragment of oncD DNA (Fig. 6B). Southern transfer analysis of DNAs isolated from colon carcinoma 2033 and from first- and third-cycle NIH 3T3 transformants derived from oncD[4], revealed the presence of the rearranged 23-kb EcoRI and 7-kb BamHI DNA fragments diagnostic of oncD. These results established that oncD was not activated during gene transfer but was present in the original human tumour DNA. The normal 11-kb EcoRI and 14-kb BamHI DNA fragments were also present in the original tumour DNA, demonstrating its heterozygous nature.

Finally, to test whether the rearrangement responsible for the activation of oncD was an inherited genetic abnormality or occurred during tumour development, DNA was isolated from normal colonic tissue adjacent to tumor 2033 (provided by S. A. Aaronson). This DNA exclusively exhibited the normal 11-kb EcoRI and 14-kb BamHI DNA fragments characteristic of normal human DNAs (Fig. 6). These results establish unequivocally that oncD was generated by a somatic rearrangement that brought together two independent loci during development of a colon carcinoma.

## Discussion

Here we have described the generation of a human transforming gene by a genetic rearrangement that resulted in the juxtaposition of two unrelated loci. One of the two genes involved in the generation of oncD appears to be a novel receptor gene with associated protein tyrosine kinase activity. That a protein tyrosine kinase receptor gene could be activated as an oncogene was first illustrated by Downward et al., who demonstrated that the v-erb-B oncogene was derived from the EGF receptor locus[30]. It has been proposed that truncation of the amino-
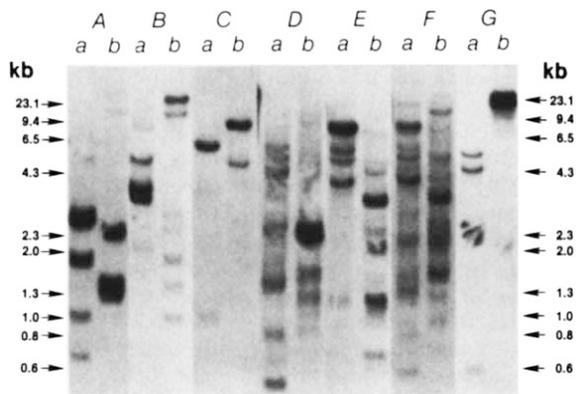
**Fig. 4** A novel human locus in oncD. The figure shows a Southern transfer analysis of normal human DNA digested with PstI (a) and PvuII (b) restriction endonucleases and hybridized with probes specific for: A, oncD (1.2-kb BalI-EcoRI insert of pDM10-1); B, v-fms (409-bp PstI insert of pSM7C, ATCC no. 41016); C, v-ros (844-bp EcoRI-PvuII insert of p-ros, ATCC no. 41051); D, v-fes (459-bp PstI insert of pST4, ATCC no. 41014); E, v-fgr (952-bp PvuII-KpnI insert of GR-FeSV; ref. 31); F, v-src (2,950-bp EcoRI insert of pEcoRIB, ATCC no. 41005); and G, v-yes (550-bp PstI-EcoRI insert of pY-73; ref. 39).

**Methods.** Restricted DNAs (20 μg) were applied to 1.2% (w/v) agarose gels, electrophoresed and transferred to nitrocellulose paper as described elsewhere[40]. Filters were hybridized for 48 h to $2 \times 10^7$ c.p.m. of the corresponding nick-translated probe under stringent conditions (50%(v/v) formamide, 42 °C) for the oncD probe and under relaxed conditions (30%(v/v) formamide, 42 °C) for the rest of the probes. Filters were exposed to Kodak XAR-5 film at −70 °C for 20 h with the help of intensifier screens. Co-electrophoresed λ HindIII DNA fragments served as relative molecular mass ($M_r$) markers.

terminal (extracellular) domain may lock the EGF receptor in its active state, thus mimicking constitutive activation by EGF[22]. However, it appears that an additional deletion in the carboxy-terminal end of the EGF receptor molecule is necessary for full neoplastic potential (ref. 22 and A. Ullrich, personal communication). In the case of oncD, it seems that a putative receptor gene has been decapitated while maintaining the transmembrane and catalytic domains without any other detectable rearrangements. However, the homology between other protein tyrosine kinases and oncD ends at residue 592 (Fig. 3), about 15 residues from the consensus carboxy-terminal boundary[22]. Identification of additional genetic alterations in the protein tyrosine kinase domain of oncD must await the molecular cloning of its normal allele.

v-fgr, the oncogene of the Gardner-Rasheed strain of feline sarcoma virus (GR-FeSV), also encodes a protein tyrosine kinase that carries a domain (128 amino acids) of γ-actin, another cytoskeletal protein[31]. Normal c-fgr, unlike the predicted oncD product, apparently is not a receptor[32,33]. Moreover, the 118 amino-terminal residues of the v-fgr protein are contributed by viral gag gene sequences[31]. However, it is possible that the respective cytoskeletal domains of the v-fgr and oncD products play a part in redirecting their subcellular localization, preventing them from interacting with their normal substrates. In this regard, it is worthwhile noting that tropomyosins do not contain leader signal sequences that would be expected to direct transport through lipid bilayers. Thus, it is possible that the oncD protein cannot be anchored to the cellular membrane as predicted by its receptor-like structure. Generation of specific antibodies as well as appropriate deletion mutants should help to resolve the role of the cytoskeletal domains in the transforming properties of the v-fgr and oncD proteins.

Molecular characterization of genomic and cDNA clones of tropomyosin genes has indicated that some tropomyosin isoforms may be encoded by a single structural gene that used
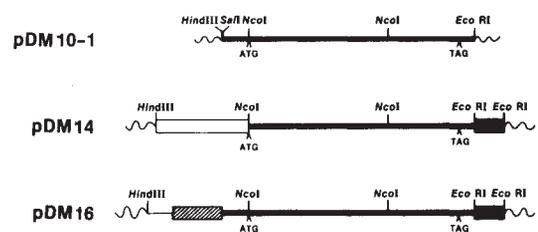


**Fig. 5** Transforming activity of oncD cDNA. The figure shows a schematic diagram of clones pDM10-1, pDM14 and pDM16. (~), pUC18 vector sequences; ▬, oncD cDNA sequences; □, promoter sequences of the human metallothionein IIA gene; ▨, M-MSV LTR; ▮, SV40 polyadenylation sequences. The restriction endonuclease sites used to construct pDM14 and pDM16 from pDM10-1 as well as the locations of the beginning (ATG) and end (TAG) of the oncD cDNA coding sequences are indicated. The transforming activities of pDM10-1, pDM14 and pDM16 (as supercoiled plasmids) in gene transfer assays using NIH 3T3 mouse fibroblasts as recipient cells[29] were, respectively, 5, $6 \times 10^3$ and $2 \times 10^4$ FFU per μg of transfected plasmid DNA. pUC18 alone did not elicit focus formation.

**Methods.** pDM14 was generated by replacing the 249-bp HindIII-NcoI fragment of pDM10-1 with an 840-bp HindIII-NcoI fragment of p84H containing the human methallothionein IIA promoter[27]. pDM16 was constructed by inserting a 750-bp EcoRI-SmaI fragment of pm1SP[26], containing the U3 and R regions of M-MSV and 275-bp of mink DNA 5′-flanking sequences (thin line), into the SalI site of pDM10-1. The EcoRI and SalI sites were blunt-ended before ligation. pDM14 and pDM16 also contained a 237-bp BclI-BamHI DNA fragment carrying the SV40 polyadenylation signal. The BclI and BamHI cleavage sites were replaced by EcoRI sites.
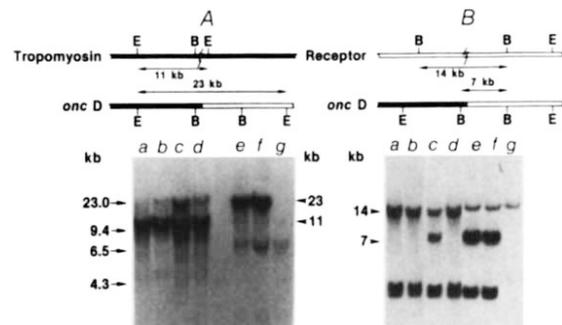


**Fig. 6** oncD was generated by a somatic rearrangement in a human colon carcinoma. *Top*, schematic representation of DNA fragments derived from : A, EcoRI and B, BamHI digestion of normal human DNAs and DNAs containing oncD. Tropomyosin sequences are represented by a solid black line whereas those of the protein tyrosine kinase gene are represented by an open line. The zig-zag line indicates the approximate location of the breakpoint that led to the generation of oncD. Only those EcoRI (E) and BamHI (B) cleavage sites needed to define the diagnostic DNA restriction fragments are indicated. *Bottom*, Southern transfer analysis of DNAs isolated from: a, normal human lymphocytes; b, colon carcinoma 1853; c, colon carcinoma 2033, the tumour in which oncD was identified; d, normal colonic tissue adjacent to tumour 2033; e, first-cycle NIH 3T3 transformant derived from oncD; f, third-cycle NIH 3T3 transformant derived from oncD; g, NIH 3T3 cells. Analysis was carried out as decribed in Fig. 5 legend, except that 0.7% (w/v) agarose gels were used. Nitrocellulose filters were hybridized under stringent conditions to $6 \times 10^7$ c.p.m. of nick-translated probes specific for a, tropomyosin sequences (2.7-kb KpnI insert of pDM8) or B, protein tyrosine kinase sequences (1.2-kb BalI-EcoRI fragment of pDM10-1). Filters were exposed to Kodak XAR-5 film at −70 °C for 20 h with the help of intensifier screens. Co-electrophoresed λ HindIII DNA fragments serves as $M_r$ markers. DNA fragments which defined the genetic rearrangement that generated oncD are indicated by arrowheads; these include the 11-kb EcoRI and 7-kb BamHI normal human DNA fragments and the 23-kb EcoRI and 14-kb BamHI oncD-specific DNA fragments.

alternative RNA splicing mechanisms. In *Drosophila*, embryonic and thoracic tropomyosins are encoded by the same gene via alternative splicing of the last exon, which codes for the different 27 carboxy-terminal residues that distinguish the two tropomyosin isoforms[34]. A similar mechanism appears to be responsible for the synthesis of $\alpha$-tropomyosin isoforms of smooth and striated muscle of rats[35]. Finally, alternative splicing of both internal and carboxy-terminal exons is also responsible for the synthesis of a smooth muscle-like tropomyosin expressed in human non-muscle cells and of the human skeletal muscle $\beta$-tropomyosin[36]. Interestingly, *onc*D contains the entire tropomyosin coding sequences except for the carboxy-terminal 27 amino-acid residues presumably coded for by the missing coding exon 8 (ref. 15 and our unpublished observations). c-*abl*, the other protein tyrosine kinase oncogene implicated in human neoplasia[37], becomes activated by a chromosomal translocation that causes the replacement of its first coding exon with the 5' domain of another locus[38]. Recently, Y. Ben-Neriah and D. Baltimore have shown that the normal c-*abl* locus codes for at least four different proteins by alternative splicing of the first coding exon (personal communication). It is possible that complex transcriptional processes required for the synthesis of multiple proteins by a single gene may facilitate the occurrence of genetic rearrangements and the formation of fusion products capable of inducing malignant transformation.

Molecular characterization of novel transforming genes such as *onc*D will not only expand our current understanding of the molecular biology of neoplasia, but will prove to be a fruitful source of 'mutant' genes that should help us to elucidate the pathways involved in the control of cell growth and proliferation.

## *onc*D to be designated *trk*

Most oncogenes have been designated according to their origin (that is, *src* for sarcoma, *ras* for rat sarcoma, *abl* for Abelson murine leukaemia virus, etc.). Although *onc*D was initially identified in a colon carcinoma[4], apparently this oncogene is not frequently activated in this type of tumour (D.M.-Z. and M.B., unpublished observations). Therefore, we suggest that this rearranged oncogene be designated by a conventional three-letter abbreviation, *trk* (pronounced 'track'), that reflects its molecular structure: tropomyosin(t)-receptor(r)-kinase(k).

1. Varmus, H. E. *A. Rev. Genet.* **18**, 553-612 (1984).
2. Barbacid, M. in *Important Advances in Oncology 1986* (eds Devita, V., Hellman, S. & Rosenberg, S.) (Lippincott C., Philadelphia, in the press).
3. Balmain, A. *Br. J. Cancer* **51**, 1-7 (1985).
4. Pulciani, S. *et al. Nature* **300**, 539-542 (1982).
5. Cooper, C. S. *et al. Nature* **311**, 29-33 (1984).
6. Padua, R. A., Barrass, N. & Currie, G. A. *Nature* **311**, 671-673 (1984).
7. Eva, A. & Aaronson, S. A. *Nature* **316**, 273-275 (1985).
8. Shimizu, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5641-5645 (1985).
9. Fasano, O., Birnbaum, D., Edlund, L., Fogh, J. & Wigler, M. *Molec. cell. Biol.* **4**, 1695-1705 (1984).
10. Takahashi, M., Ritz, J. & Cooper, G. M. *Cell* **42**, 581-588 (1985).
11. Cooper, G. M. & Lane, M. A. *Biochim. biophys. Acta* **738**, 9-20 (1984).
12. Notario, V., Sukumar, S., Santos, E. & Barbacid, M. *Cold Spring Harb. Conf. Cell Proliferation Cancer* **11**, 425-432 (1984).
13. Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. & Hughes, S. H. *Proc. natn. Acad. Sci. U.S.A.* **80**, 31-35 (1983).
14. Hanahan, D. & Meselson, M. *Gene* **10**, 63-67 (1980).
15. MacLeod, A. R. & Talbot, K. *J. molec. Biol.* **167**, 523-537 (1983).
16. Lewis, W. G., Cote, G. P., Mak, A. S. & Smilie, L. B. *FEBS Lett.* **156**, 269-273 (1983).
17. Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning: A Laboratory Manual*, 90 (Cold Spring Harbor Laboratory, New York, 1982).
18. Maxam, A. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 560-564 (1977).
19. Baer, R. *et al. Nature* **310**, 207-211c (1984).
20. Weiss, R., Teich, N., Varmus, H. & Coffin, J. (eds) *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, New York, 1985).
21. Ullrich, A. *et al. Nature* **313**, 756-761 (1985).
22. Ullrich, A. *et al. Nature* **309**, 418-425 (1984).
23. Hunter, T. & Cooper, J. A. *A. Rev. Biochem.* **54**, 897-930 (1985).
24. Bryant, D. & Parsons, J. T. *J. Virol.* **45**, 1211-1216 (1983).
25. Groffen, J., Heisterkamp, N., Reynolds, F. H. & Stephenson, J. R. *Nature* **304**, 167-169 (1983).
26. Karin, M. *et al. Nature* **308**, 513-519 (1984).
27. Blair, D. G. *et al. Science* **212**, 941-943 (1981).
28. Fitzgerald, M. & Shenk, T. *Cell* **24**, 251-260 (1981).
29. Graham, F. L. & van der Eb, A. J. *Virology* **52**, 456-467 (1973).
30. Downward, J. *et al. Nature* **307**, 521-526 (1984).
31. Naharro, G., Robbins, K. C. & Reddy, E. P. *Science* **223**, 63-66 (1984).
32. Tronick, S. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 6595-6599 (1985).
33. Parker, R. C., Mardon, G., Lebo, R. V., Varmus, H. E. & Bishop, J. M. *Molec. cell. Biol.* **5**, 831-838 (1985).
34. Basi, G. S., Boardman, M. & Sorti, R. *Molec. cell. Biol.* **4**, 2828-2836 (1984).
35. Ruiz-Opazo, N., Weinberger, J. & Nadal-Ginard, B. *Nature* **315**, 67-70 (1985).
36. MacLeod, A. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7835-7839 (1985).
37. Heisterkamp, N. *et al. Nature* **306**, 239-242 (1983).
38. Shtivelman, E., Lifshitz, B., Gale, R. P. & Canaani, E. *Nature* **315**, 550-554 (1985).
39. Kitamura, N., Kitamura, A., Toyoshima, K., Hirayama, Y. & Yoshida, M. *Nature* **297**, 205-208 (1982).
40. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).

# ——————— LETTERS TO NATURE ———————

# Inflation and shadow matter

Lawrence M. Krauss*, Alan H. Guth*, David N. Spergel, George B. Field & William H. Press

Harvard-Smithsonian Center for Astrophysics, 60 Garden St, Cambridge, Massachusetts 02138, USA
* Present addresses: Department of Physics, Yale University, New Haven, Connecticut 06511, USA (L.M.K.); Center for Theoretical Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA (A.H.G.).

**Recent work on superstring theories has prompted interest in 'shadow matter', exotic matter which interacts only gravitationally with normal matter. Green and Schwarz[1,2] have shown that the only anomaly-free type 1 superstring theories are those with gauge group $SO_{32}$. Gross et al.[3] have constructed a superstring theory resulting in a gauge group $E_8 \times E_8$. Such a theory could result, at low energies, in the existence of two sectors: an 'observed' sector associated with all familiar particles and interactions, and another 'hidden' sector, previously discussed in various low-energy supersymmetry models (see, for example, ref. 4 and references therein). The particles of the hidden $E_8$ sector couple only through gravitational interactions with ordinary matter. Kolb, Seckel and Turner[5] have explored some of the astrophysical and cosmological implications of the existence of such shadow matter. We demonstrate here that if, in the early Universe, an inflationary phase[6-8] is associated with the breaking of one of the symmetries in the $E_8 \times E_8$ theory, this strongly constrains the physics of both sectors if shadow matter is to be the missing mass in the Universe.**

We begin by assuming that inflation occurs because of symmetry breaking in the observed $E_8$ sector. After inflation has ended, the abundance of all observed particles is determined by particle production and reheating as the scalar field responsible for inflation oscillates and relaxes to its true minimum by particle production[9,10]. As is well known for the case of gravitinos in low-energy supergravity models, the production of particles which are only gravitationally coupled to ordinary matter will be suppressed during reheating compared with ordinary matter, by powers of the reheating temperature over the Planck mass. This allows one to place upper bounds on the