The EMBO Journal vol.7 no.1 pp.147 - 154, 1988

Activation of the receptor kinase domain of the *trk* oncogene by recombination with two different cellular sequences

Sara C.Kozma[1], Shelagh M.S.Redmond, Fu Xiao-Chang, Susanne M.Saurer, Bernd Groner and Nancy E.Hynes

Ludwig Institute for Cancer Research, Bern Branch, Inselspital, CH-3010 Bern, Switzerland

[1]Present address: Friedrich Miescher-Institute, Postfach 2543, CH-4002 Basel, Switzerland

Communicated by H.Bauer

A new chimeric oncogene, *trk*-2^h, has been generated by recombination of two segments of MDA-MB231 human breast carcinoma cell line DNA after transfection in NIH/3T3 cells. The rearranged DNA segments form a fused transcriptional unit. Sequences at the 3′ end are homologous to the tyrosine kinase receptor moiety found in the *trk* oncogene which resembles a truncated growth factor receptor lacking part of its extracellular domain (Martin-Zanca *et al.*, 1986). The 5′ sequence of the *trk*-2^h oncogene is contributed by a gene which is expressed in all human cells tested, and is not related to any known gene. Transfection of the receptor kinase domain DNA fragment into NIH/3T3 cells generated another oncogene, *trk*-3^{mh}, which contains a mouse-specific sequence fused 5′ to the receptor kinase. All three *trk* recombinants have the receptor kinase moiety fused to an activating amino terminus at the same nucleotide in their transcriptional product.

Key words: genomic DNA transfection/tumorigenicity assay/truncated receptor protein/tyrosine-kinase oncoprotein

Introduction

Transforming genes from human and animal tumors have been identified by genomic DNA transfections in NIH/3T3 cells. The assay has identified oncogenes activated by point mutations such as those of the *ras* gene family (Reddy *et al.*, 1982; Tabin *et al.*, 1982) or the *neu* oncogene (Schechter *et al.*, 1984). In addition, oncogenes activated by rearrangements have also been detected. The rearrangements are either somatic and are observed in the original tumor cell DNA, or they occur after the transfection. The *trk* oncogene was generated in a human colon carcinoma by a somatic rearrangement of a receptor kinase sequence with the tropomyosin gene (Martin-Zanca *et al.*, 1986). Other activating rearrangements such as those forming the *ret* chimeric oncogene (Takahashi and Cooper, 1987) or involving the *raf* oncogene (Stanton and Cooper, 1987; Tahira *et al.*, 1987) occurred after transfection. These rearrangements result in the production of fusion oncoproteins with the amino terminus of an apparently normal cellular protein fused to the carboxy terminus of a tyrosine (*ret*, *trk*) or serine/threonine (*raf*) kinase.

We have detected two different transforming sequences by a tumorigenicity assay using NIH/3T3 cells transfected with DNA isolated from the human breast carcinoma cell line MDA-MB231. We have previously characterized one of these transforming sequences as the human Ki-*ras* oncogene in which a point mutation at codon 13 led to an amino acid change from glycine to aspartic acid in the transforming protein (Kozma *et al.*, 1987). The second transforming sequence from the MDA-MB231 DNA was retained in four consecutive cycles of transfection and tumorigenicity experiments in NIH/3T3 cells. The oncogene was molecularly cloned and its cDNA was sequenced. The oncogene encodes the receptor kinase domain of the *trk* fusion oncoprotein at its carboxy terminus. Unlike the original *trk* fusion oncogene which contains the first seven exons of the non-muscle tropomyosin gene fused to the receptor kinase domain, the novel *trk*-related oncogene which we describe here contains at its 5′ end human sequences which are not related to a known gene. Oncogenic potential is probably conferred to the *trk* receptor kinase by the human DNA sequences which became linked during the transfection procedure. This activating recombination event could be experimentally mimicked a second time by transfection of the receptor sequence portion of the gene in NIH/3T3 cells. A new transforming sequence was generated in which the 5′ activating DNA was provided by yet another sequence of mouse DNA.

Results

Isolation of a transforming sequence from the MDA-MB231 genomic DNA following transfection in NIH/3T3 cells

Genomic DNA from the human breast tumor cell line MDA-MB231 was transfected into NIH/3T3 cells and the transfectants were assayed for tumorigenicity in nude mice (Kozma *et al.*, 1987). Two animals developed tumors within 4 weeks. Hybridization to human genomic probes revealed a similar pattern of human repetitive sequences in both primary tumors (data not shown). We used the DNA of one primary tumor, NIH/MDA-1, in a secondary transfection and tumorigenicity assay. Secondary tumors appeared within 3 weeks in six out of six mice injected. Tertiary and quarternary cycle transfectants gave rise to tumors with the same efficiency. A blot analysis with DNA from different transformants using human genomic DNA as a probe revealed that specific fragments were retained through successive transfection cycles (Figure 1). This implies that these fragments are associated with a human transforming gene. All *Bam*HI-digested DNA from second, third and fourth round transformants contained three fragments of 6.7, 3.8 and 2.8 kb (Figure 1B) which hybridized to human repetitive DNA sequences. A tertiary transformant had two human DNA-containing *Eco*RI fragments of 18 kb and 12 kb (Figure 1A, lane f), which were cloned into the λ EMBL4 vector. The two genomic DNA inserts were released from the vector DNA by *Eco*RI digestion, ligated and transfected into NIH/3T3 cells. The transfectants

gave rise to tumors in nude mice, a result which shows that the oncogene is contained within these two fragments.

Structure of the transforming genomic sequence and of the corresponding transcript

The restriction maps of the genomic clones were established (Figure 2). DNA blotting analyses using probes from the 18- and 12-kb clones revealed that the transforming sequences arise from the rearrangement of two genomic fragments which are not linked in the MDA-MB231 DNA or in placenta DNA (not shown). The rearrangement occurred within the 1.1-kb *Eco*RI−*Hin*dIII fragment at one end of the 12-kb sequence (probe 4, Figure 2). The junction between the two human DNA fragments is detected in all the NIH/MDA transformants and in the tumors obtained from cells transfected with *in vitro* ligated 18- and 12-kb clones (not shown).

Using the entire 12-kb clone as a probe in an RNA blotting analysis, a 1.8-kb transcript was detected in all transformants, but neither in MDA-MB231 cells nor in NIH/3T3 cells. Subfragments of the genomic clones were then used to delimit the transcribed sequence. Probe 5 from the 12-kb genomic clone detected a 1.8-kb transcript in the transformants (Figure 3A, lanes b and c). Probe 1 from the 18-kb clone also detected the 1.8-kb mRNA transcript in the transformants (Figure 3B, lane b) and did not detect a transcript in control RNA from NIH/3T3 cells (Figure 3B, lane e). In addition, probe 1 detected a relatively abundant 1.2-kb mRNA in the MDA-MB231 cells (Figure 3B, lane a). Probes 4 and 7 were negative in an RNA analysis. DNA sequences 5′ to probe 1 did not detect transcripts in the transformants or in MDA-MB231 cells (data not shown).

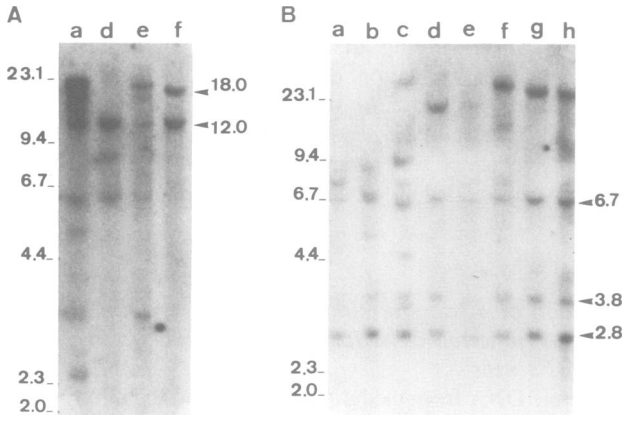

Fig. 1. Southern blot analysis of genomic DNAs isolated from NIH/MDA transformants. The DNAs were isolated from the following sources: NIH/MDA-2 transformants (**a**−**c**), NIH/MDA-3 transformants (**d**−**f**) and NIH/MDA-4 transformants (**g** and **h**). The DNAs were digested with *Eco*RI (panel **A**) or with *Bam*HI (panel **B**) and hybridized with ^{32}P-labeled human placenta DNA. The human DNA containing genomic clones was isolated from the NIH/MDA-3 transformant shown in lane f.

The direction of transcription was determined using probe 6 cloned in both orientations in an SP6 vector. The 5′ end of the gene is within probe 1 in Figure 2. From these results it appears that a large fraction of the 1.8-kb mRNA transcript is encoded by the 12-kb clone, while the 18-kb clone, in particular probe 1, provides sequences at the 5′ end. The same 5′ sequences are also present in a 1.2-kb transcript found in MDA-MB231 RNA. This 1.2-kb transcript was detected in all the human RNA samples that we examined, including normal breast epithelium, normal fibroblast and placental RNA (not shown).

cDNA cloning and sequencing of the oncogene-specific transcript

We used poly(A)$^+$ mRNA from NIH/MDA-2 cells to prepare a cDNA library in λgt10 (Huynh *et al.*, 1985). Six clones were isolated from 9 × 10^5 recombinants after screening the library consecutively with probes 3, 5 and 6, shown in Figure 2. The longest cDNA obtained was 1.6-kb in length. In an RNA blot analysis the cDNA hybridized to the 1.8-kb transcript from the transformants (Figure 3C, lanes b and c) and to the 1.2-kb transcript from MDA-MB231 cells (Figure 3C, lane a), but did not hybridize to NIH/3T3 RNA (Figure 3C, lane e). This cDNA was subcloned in bacteriophage M13 and sequenced (Figure 4). The sequence 3′ of nucleotide 125 is homologous with the receptor kinase part of the *trk* oncogene (Martin-Zanca *et al.*, 1986).

To avoid the addition of a new name to the list of oncogenes, we propose to consider *trk* as the abbreviation for *t*ransforming *r*eceptor *k*inase (rather than *t*ropomyosin *r*eceptor *k*inase, as proposed by Martin-Zanca *et al.*, 1986). We will refer to the previously described *trk* oncogene as *trk*-1^h, whereby the 1 stands for the first chimera described and the h for human, and will designate the chimera found in NIH/MDA transformants as *trk*-2^h. In both oncogene transcripts the receptor kinase domain starts at the same nucleotide, which corresponds to position 664 in *trk*-1^h and position 125 in *trk*-2^h.

The receptor kinase domain of both *trk*-1^h and *trk*-2^h begins at the same amino acid, residue 42 in *trk*-2^h. A putative transmembrane domain spans residues 57−82. This

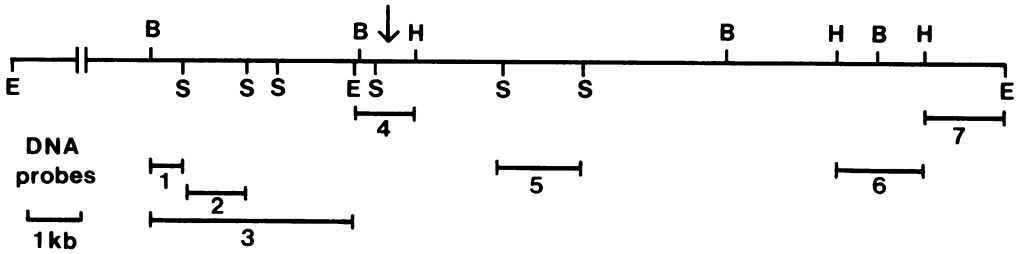

Fig. 2. Scheme of the genomic DNA clones isolated from NIH/MDA-3. A partial map of the 18-kb *Eco*RI fragment is shown on the left and the 12-kb *Eco*RI fragment is shown on the right. The junction between the two human DNA sequences is marked by an arrow. The restriction endonuclease sites shown are: *Eco*RI (E), *Sma*I (S), *Bam*HI (B) and *Hin*dIII (H). The indicated DNA probes were tested on poly(A)$^+$ RNA isolated from the NIH/MDA transformants and MDA-MB231 cells: probes 1, 2, 3, 5 and 6 hybridized with the 1.8-kb oncogene specific transcript; probes 1−3 hybridized with the normal 1.2-kb transcript present in MDA-MB231 cells; probes 4 and 7 as well as DNA sequences 5′ to probe 1 were negative with all RNAs tested.

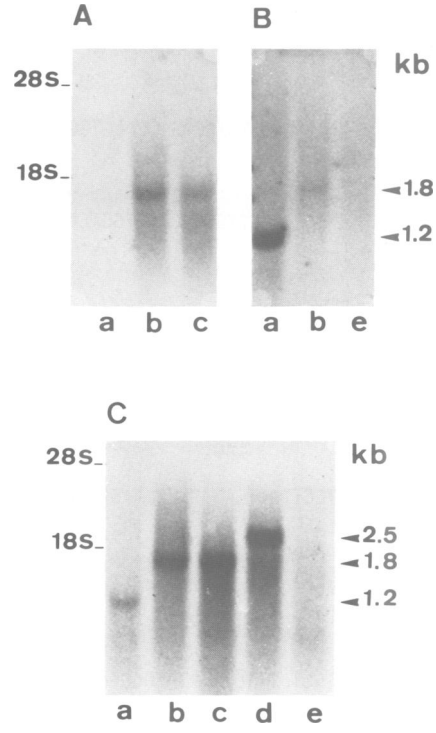

Fig. 3. RNA blot analysis of poly(A)$^+$ RNA isolated from NIH/MDA transformants. Poly(A)$^+$ RNA was isolated from: MDA-MB231 cell line (**a**), NIH/MDA-2 transformant (**b**), NIH/MDA-4 transformant (**c**), a transformant obtained by transfection of the 12-kb genomic clone (**d**) and NIH/3T3 cells (**e**). The hybridization was carried out with probe 5 from the 12-kb genomic clone (panel **A**), probe 1 from the 18-kb genomic clone (panel **B**) and the *trk*-2^h cDNA probe (panel **C**).

is followed by a region which shows similarities with other tyrosine kinases (Hunter and Cooper, 1986). The published trk-1^h sequence and the trk-2^h sequence are different at some positions (indicated in Figure 4). Most of these differences do not affect the amino acid sequence. The exceptions are the changes at positions 1224 and 1225, which lead to a replacement of, respectively, histidine for glutamine and glycine for arginine.

The most significant difference between the sequence of *trk*-2^h shown in Figure 4 and that reported for *trk*-1^h (Martin-Zanca *et al.*, 1986) is the presence of an additional cytosine at position 1234 and a cytosine instead of a thymidine at position 1285 in the *trk*-2^h sequence. These changes alter the open reading frame at the carboxy terminus of the protein. Barbacid's group have recently re-examined their sequence and have confirmed the existence of these two cytosines in *trk*-1^h (personal communication). The new predicted carboxy terminal sequence enhances the homology of the *trk* receptor kinase with other tyrosine kinases. In particular, there is now a leucine residue at position 424, the carboxy boundary of the catalytic domain. All tyrosine kinases have a hydrophobic residue (leucine or phenylalanine) at this site (Hunter and Cooper, 1986). The single tyrosine residue at 434 following the catalytic domain of *trk*-2^h may correspond to tyrosine 527 of c-*src*.

The open reading frame of the receptor kinase is preceded by an open reading frame at the 5′ end of the *trk*-2^h cDNA. This sequence is not homologous with tropomyosin, which is present in the *trk*-1^h oncogene. The 5′ sequence of the *trk*-2^h cDNA (nucleotides 20−124) is incomplete, since it does not contain an ATG start codon in the open reading frame. The sequence was extended in the 5′ direction by a primer extension procedure. Primer 1 shown in Figure 4 was annealed to poly(A)$^+$ RNA from NIH/MDA-4 and elongated in the presence of dideoxynucleotides by reverse transcriptase. The sequence −30 to +20 shown in Figure 4 was determined. This sequence contains two 9-mer inverted repeats (nucleotides −20 to −22 and +6 to +14) which could fold into a stem and loop structure and explain the incomplete length of the cDNA. Most of the extended primers stopped at the C at position −20, suggesting that this is the major RNA start site. The sequence also contained an ATG in frame with the receptor kinase domain. To determine if the ATG at position 1 corresponds to the translational start of the molecule, we sequenced probe 1 from the 18-kb genomic clone (Figure 2). It confirmed the sequence found by primer extension. The ATG codon is preceded by the sequence CCAAG. This sequence contains three out of five nucleotides of the consensus translational start sequence CC^A_GCC and respects the key feature of an adenosine at position −3 (Kozak, 1984). At the 3′ end of the *trk*-2^h cDNA there are 208 untranslated nucleotides. No consensus poly(A)$^+$ addition site could be detected (Wickens and Stephenson, 1984).

The deduced translational product of *trk*-2^h consists of 439 residues, 398 of which encompass the receptor kinase. The predicted mol. wt of the *trk*-2^h protein is 48.6 kd and corresponds reasonably well with the observed mol. wt of 44 kd (A.Ziemiecki *et al.*, in preparation). The amino terminus of *trk*-2^h is rich in lysine and arginine residues and contains neither putative glycosylation sites nor a hydrophobic leader sequence.

Cloning and sequencing of the transcript homologous to the 5′ activating sequence of trk-2^h

A placenta cDNA library (Clontech) was screened with probes 1 and 2 from the 18-kb genomic clone and a number of positively hybridizing plaques were isolated. A 0.9-kb placenta cDNA insert was subcloned in M13 and sequenced. The entire cDNA sequence is shown in Figure 5. The cloned placenta cDNA was also incomplete, stopping at nucleotide 15. Primer 1 was used to extend and sequence MDA-MB231 poly(A)$^+$ RNA. Its sequence is the same at the 5′ end as the *trk*-2^h sequence shown in Figure 4. The incomplete length of the placenta cDNA might be due to the 5′ stem and loop structure mentioned earlier. The sequences of the two cDNAs diverge 3′ of nucleotide 124. Sequencing genomic DNA probes 1 and 2 (Figure 2) revealed that the 124 coding nucleotides which are present at the 5′ end of *trk*-2^h and the placenta cDNA are contained in two exons. The first exon contains the 5′ untranslated sequences plus the ATG, while the second exon contains the remaining 121 nucleotides. The placenta cDNA has an open reading frame coding for 266 amino acids. The most striking feature of the putative protein is its high basic amino acid content (23% lysine plus arginine). Data bank searches on the nucleotide and amino acid level did not show any similarities between the placenta cDNA sequence and stored sequences. Thus, the receptor kinase domain of the *trk*-2^h oncogene is fused to a human sequence not yet described.

Generation of a mouse−human chimeric oncogene by transfection of the receptor kinase domain of trk-2^h

In order to determine the oncogenic potential of the two human DNA-containing *Eco*RI fragments isolated from the

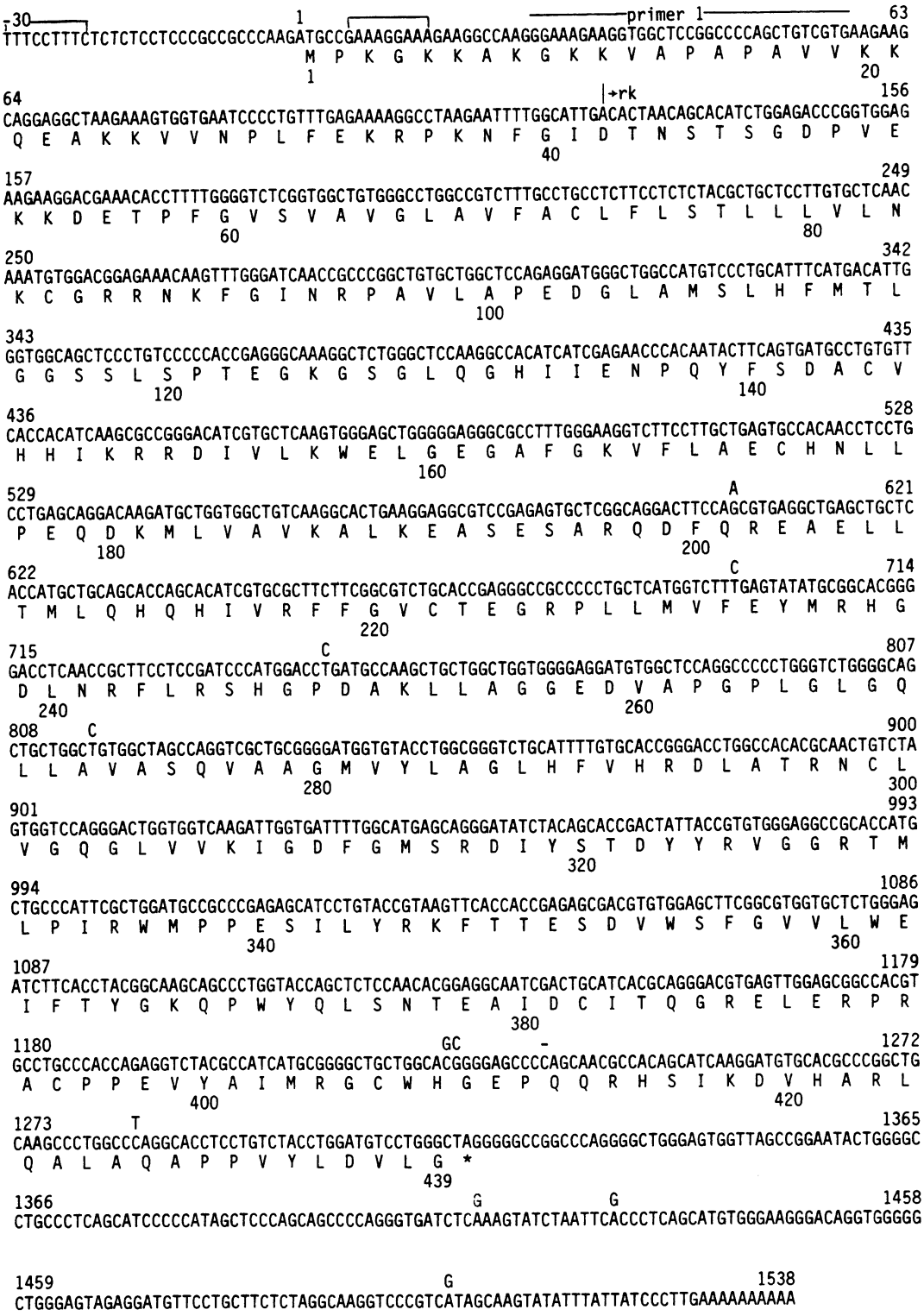

Fig. 4. Nucleotide sequence and deduced amino acid sequence of the *trk*-2^h oncogene. The entire sequence was obtained by cDNA sequencing combined with primer extension and sequencing of NIH/MDA-4 and MDA-MB231 poly(A)$^+$ RNAs. Differences between the *trk*-1^h (Martin-Zanca *et al.*, 1986) and our sequence are indicated above the line. The 28 amino acids at the carboxy terminus differ from the published sequence. The corrected *trk*-1^h sequence (D.Martin-Zanca, personal communication) and the *trk*-2^h sequence shown here are the same. Nucleotides are numbered above the line. Position 1 marks the first ATG codon. A 9-mer inverted repeat sequence is marked between positions −30, −22 and +6, +14. Amino acids are numbered below the line. Primer 1 was used to extend the sequence to the 5′ end. The majority of the extended products stopped at the C at position −20. The receptor kinase domain begins at amino acid 42 (indicated by →rk). The region with similarity to other tyrosine kinases spans residues 159−424.

NIH/MDA transformant, we individually transfected the 18- and 12-kb genomic fragments shown in Figure 2. The 18-kb fragment contains part of the sequence transcribed in the 1.2-kb human mRNA. We assume that the 12-kb genomic clone contains only the receptor kinase domain of the oncogene for two reasons. First, the 12-kb fragment hybridizes

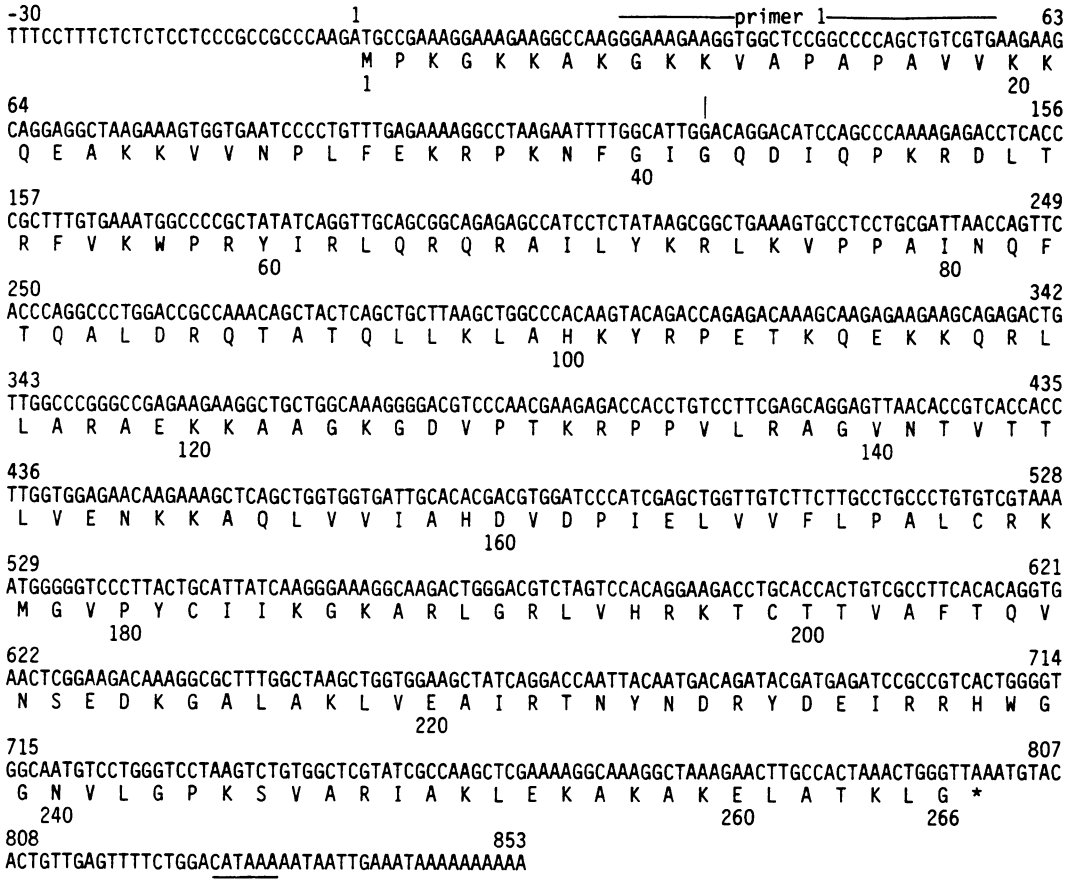

Fig. 5. Nucleotide sequence and deduced amino acid sequence of the placenta cDNA. The entire sequence was obtained by cDNA sequencing combined with primer extension and sequencing of NIH/MDA-4 and MDA-MB231 poly(A)$^+$ RNAs. Nucleotides are numbered above the line. Position 1 marks the first ATG codon. Amino acids are numbered below the line. The junction between the 5′ activating sequences and the receptor kinase of *trk*-2^h is indicated at nucleotide 125. A variant form of the poly(A)$^+$ addition signal (Wickens and Stephenson, 1984) is underlined at position 826–831.

to the 1.8-kb *trk*-2^h mRNA and not to the ubiquitous 1.2-kb normal human transcript. Second, probe 4 (Figure 2), which contains the junction of the two human genomic DNA fragments, does not hybridize to poly(A)$^+$ RNA from the NIH/MDA transformants or from MDA-MB231 cells.

Cells transfected with the 12-kb genomic sequence were tumorigenic with a low efficiency. Only one cell clone out of $\sim 1.3 \times 10^3$ which were cotransfected with pSV2neo and the 12-kb genomic DNA sequence gave rise to tumors in nude mice. Cells transfected with the 18-kb genomic DNA sequence were not tumorigenic. A blot analysis on poly(A)$^+$ RNA isolated from the 12-kb genomic DNA-induced tumors revealed a transcript of 2.5-kb which hybridized with the *trk*-2^h cDNA probe (Figure 3C, lane d). The receptor kinase moiety of *trk* is encoded within $\sim$1.5-kb. Therefore, we assume that the 2.5-kb transcript arose from a rare recombination event between a mouse gene from the NIH/3T3 recipient cells and the transfected human receptor kinase domain. Primer extension and sequencing of tumor poly(A)$^+$ RNA was performed using a synthetic oligonucleotide corresponding to the 5′ end of the receptor kinase sequence (primer 2, Figure 6). A partial sequence of 115 nucleotides located 5′ of the junction with the receptor kinase domain has been determined (Figure 6). This sequence contains an open reading frame of 38 amino acids in frame with the receptor kinase domain. Data bank searches did not reveal homology between this sequence and known DNA or protein sequences. We will refer to this chimeric oncogene as *trk*-3^{mh}, whereby mh stands for mouse–human chimera. The *trk*-3^{mh} has the same junction between the 3′ receptor kinase sequences and the 5′ activating sequences as seen for the other two *trk* oncogenes. A comparison of the sequences is shown in Figure 6.

Comparison of NIH/3T3 transformants obtained with trk recombinants activated by different 5′ sequences

Cell cultures established from tumors containing the *trk*-2^h and the *trk*-3^{mh} oncogenes display different morphologies (Figure 7). The *trk*-3^{mh} transformants are rounded up and very refractile (panel B), as are the cells transformed by *trk*-1^h (obtained from D.Martin-Zanca and M.Barbacid) (panel C). The *trk*-2^h transformants (panel A) exhibit a less transformed phenotype and round up only when they are crowded on the culture dish. The different morphologies may be due to properties of the different amino termini.

Discussion

We have molecularly cloned and sequenced an oncogene from transformed NIH/3T3 cells which were transfected with genomic DNA from the human tumor cell line MDA-MB231. The 3′ part of the transforming sequence corresponds to the tyrosine kinase domain of the previously described *trk* oncogene. The transforming potential of the

```
trk-3mh                                                    -----ATCATCTCAATGCGTCTCAGACTACACTTGAAGCAG
                                                                I  I  S  M  R  L  R  L  H  L  K  Q
                                                                ————————primer 2————————
                                                           |→rk
CTCCGAAGTAGAGCGTCAGCGCTGGGTGACAGCCTTAGAACTGCCAAGGCCAAGGCTGTGAAGATGCTGGCCGAATCAGACACTAACAGCACATCTGGAGACCCGGTGGAGAAGAAGGACGAAACACCT
 L  R  S  R  A  S  A  L  G  D  S  L  R  T  A  K  A  K  A  V  K  M  L  A  E  S  D  T  N  S  T  S  G  D  P  V  E  K  K  D  E  T  P

                                                           |→rk
trk-2h                                          -----AGGCCTAAGAATTTTGGCATTGACACTAACAGCACATCTGGAGACCCGGTGGAGAAGAAGGACGAAACACCT
                                                      R  P  K  N  F  G  I  D  T  N  S  T  S  G  D  P  V  E  K  K  D  E  T  P

                                                           |→rk
trk-1h                                          -----AAGACAATTGATGACCTGGAAGACACTAACAGCACATCTGGAGACCCGGTGGAGAAGAAGGACGAAACACCT
                                                      K  T  I  D  D  L  E  D  T  N  S  T  S  G  D  P  V  E  K  K  D  E  T  P
```

Fig. 6. Partial nucleotide and amino acid sequence of the *trk*-3^{mh} oncogene. The sequence was obtained by extension of the indicated primer 2 and dideoxy sequencing of poly(A)$^+$ RNA isolated from the *trk*-3^{mh} transformants. The junction between the *trk*-3^{mh} receptor kinase domain and the activating amino terminal sequences is shown by →rk. The junction is the same as seen for the *trk*-1^h and *trk*-2^h proteins.

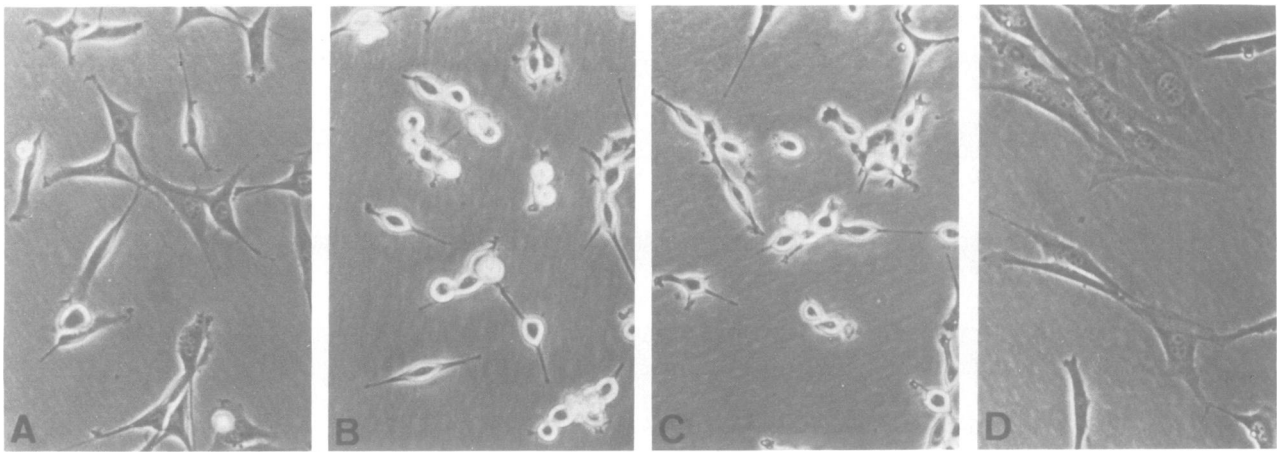

Fig. 7. Cells containing the *trk*-2^h oncogene (**A**), the *trk*-3^{mh} oncogene (**B**) and the *trk*-1^h oncogene (**C**) (provided by D.Martin-Zanca and M.Barbacid) display different morphologies. Panel **D** shows untransformed NIH/3T3 cells (×188).

trk-2^h oncogene was activated by a DNA rearrangement after transfection. The *trk*-1^h oncogene which was generated by a somatic rearrangement in a colon carcinoma (Martin-Zanca *et al.*, 1986) has the same junction between the 5′ activating sequences and the receptor kinase domain as has the *trk*-2^h oncogene. Perhaps the DNA near this break-point contains sequences which favor recombination events. Tahira *et al.* (1987) have shown that the c-*raf* gene which is usually activated between exons seven and eight by recombination with other DNA sequences contains inverted repeats near the genomic break-point. Such sequences can produce secondary structures which could play a role in recombination.

There are two mechanisms which could explain why the *trk*-related oncoproteins are transforming. The amino terminal truncation of the c-*trk* protein may be responsible, or the truncation and the production of a novel fusion protein may be necessary for activation. The many different examples of chimeric transforming proteins suggest that a fusion event may be important for activation. Many transforming retroviruses contain oncogenes encoding members of the tyrosine kinase family. The majority of the oncoproteins are chimeras containing parts of the viral structural proteins fused to proteins with tyrosine kinase activity. These retrovirally encoded fusion proteins include v-*fgr* of Gardner–Rasheed feline sarcoma virus (Naharro *et al.*, 1984), v-*erbB* of avian erythroblastosis virus (Debuire *et al.*, 1984) and c-*erbB* activated by ALV insertion (Nielson *et al.*, 1985), v-*abl* of Abelson leukemia virus (Wang *et al.*, 1984), v-*kit* of the Hardy–Zuckerman 4 feline sarcoma virus (Besmer *et al.*, 1986) and v-*fms* of the McDonough feline sarcoma virus (Hampe *et al.*, 1984). Oncogenes generated by recombination either somatically or after transfection also include members of the tyrosine or serine/threonine kinase family. These oncogenes also encode fusion proteins such as the *bcr/abl* oncogene present in chronic myelogenous leukemia cells (Konopka *et al.*, 1984), the *met* (Park *et al.*, 1987), the *ret* (Takahashi and Cooper, 1987) and the fusion *raf* (Stanton and Cooper, 1987; Tahira *et al.*, 1987) oncogenes.

It is not yet evident what the amino terminal activating sequences which have been described have in common. The *trk*-1^h and the v-*fgr* genes code for tyrosine kinases attached to domains of, respectively, tropomyosin (Martin-Zanca *et al.*, 1986) and the γ-actin (Naharro *et al.*, 1984), both of which are cytoskeletal proteins. The *met* protein appears to contain a domain at its amino terminus which is related to laminin, a glycoprotein present as a matrix component of basement membrane (Chan *et al.*, 1987). The activating sequences of the fusion *raf*, *ret*, *trk*-2^h and *trk*-3^{mh} are from genes whose proteins have not been described. The amino terminal residues in *trk*-2^h and *trk*-3^{mh} are not related but both have a high percentage of the basic amino acids arginine and lysine. We can speculate that these sequences are important for the stability, conformation and/or localization of the transforming proteins. A novel localization could bring the tyrosine kinase into contact with substrates with which

it normally does not associate. This may be the case for the family of the *trk* fusion oncoproteins. The protein product of the non-rearranged c-*trk* locus has not yet been described but, based upon the similarity of its structure with other receptor proteins, it most likely resides in the plasma membrane.

The definition of structural changes of receptor kinases which result in oncogenic activation will be an important question for future oncogene research. If there are numerous cellular sequences which by recombination with a receptor kinase domain can cause oncogenic activation, the number of conceivable oncogenes might become very large. An experimental approach to this question is possible by provision of different kinase domains with the 5′ segments present in *trk*-1^h, -2^h and -3^{mh}. Such gene constructs might become useful to assess the contribution of cellular 5′ sequences to the oncogene function and possibly define a specificity with respect to the different tyrosine kinase domains.

Materials and methods

Transfection and tumorigenic assay

Transfection of genomic DNA isolated from the MDA-MB231 human breast cell line (Cailleau *et al.*, 1974) into NIH/3T3 fibroblasts and the tumorigenicity assay in nude mice have been described previously (Hynes *et al.*, 1985; Kozma *et al.*, 1987). Cell cultures were established from the tumors by mincing tumor pieces in culture medium containing G418. Tumors and derived cell cultures were called NIH/MDA 1–4 indicating the transfection of origin.

DNA and RNA analysis

^{32}P-Labeled human placenta DNA was used as a probe in Southern blot analyses (Southern, 1975) carried out with DNA isolated from the NIH/MDA transformants. The hybridization was at 42°C in 50% formamide, 10% dextran sulfate, 1 M NaCl and 1% SDS. The filters were washed extensively at 65°C in 0.1 × SSC, 0.2% SDS before autoradiography.

Total RNA was prepared by the guanidinium–cesium chloride method (Glisin *et al.*, 1974) and poly(A)$^+$ RNA was selected on oligo(dT)-cellulose. Poly(A)$^+$ RNA was denatured with glyoxal, separated on agarose gels (McMaster and Carmichael, 1977) and transferred to nitrocellulose in 20 × SSC. Hybridizations with ^{32}P-labeled DNA probes were performed at 42°C in 50% formamide, 1 × Denhardt solution, 0.2% SDS, 4 × SSC, 0.1% sodium pyrophosphate, 25 μg/ml yeast tRNA and 30 μg/ml salmon sperm DNA. The filters were washed twice at 42°C for 30 min in 2 × SSC, 0.2% SDS followed by 0.5 × SSC and 0.2% SDS before autoradiography.

Molecular cloning

A genomic library was constructed in the λ vector EMBL4 (Frischauf *et al.*, 1983) using size-selected (10–20 kb) *Eco*RI-digested NIH/MDA-3 cell DNA. A library of 6 × 10^5 recombinants was screened by plaque hybridization on Gene Screen Plus filters (New England Nuclear) using ^{32}P-labeled human DNA as a probe and the hybridization conditions described above.

A cDNA library was constructed in λgt10 (Huynh *et al.*, 1985) using the poly(A)$^+$ RNA from NIH/MDA-2 cells. A library of 10^6 recombinants was screened. In order to obtain phage recombinants with large inserts, plaque hybridization was performed successively with three DNA probes (3, 5 and 6 of Figure 2) isolated from the genomic clones. A placenta cDNA library constructed in λgt11 was purchased from Clontech (Palo Alto, CA). Hybridization with DNA probes was carried out on Gene Screen Plus filters at 65°C in 10% dextran sulfate, 1 M NaCl, 1% SDS. The filters were washed between hybridizations according to the instructions provided by the supplier.

DNA sequencing

The cDNAs were subcloned into M13 and the DNA sequences were determined in both orientations by the dideoxy method (Sanger *et al.*, 1977) using ^{35}S-label (Biggins *et al.*, 1983). The 5′ end of the *trk*-2^h cDNA and the placenta cDNA was determined on NIH/MDA-3 and MDA-MB231 RNA as described by Geliebter *et al.* (1986). The oligonucleotide indicated as primer 1 in Figures 4 and 5 was labeled at the 5′ end and annealed to poly(A)$^+$ RNA from both cell types. The primer was extended with reverse transcriptase in the presence of the four deoxynucleotides to determine the RNA start site and in the presence of dideoxynucleotides to determine the sequence. A partial sequence of *trk*-3^{mh} was determined using the same method. The oligonucleotide labeled primer 2 in Figure 6 was annealed to poly(A)$^+$ RNA from NIH/3T3 cells transformed by the *trk*-3^{mh} oncogene. The DNA sequence around the junction between the receptor kinase and the mouse-specific sequences was determined. Both primers were synthesized by Dr R.Simpson of the LICR, Melbourne Branch. The NBRK DNA and Protein Data Bases and the Microgenie Program were used to search for sequence homologies.

Acknowledgements

We thank U.Günthert for help with the cDNA library preparation, R.Simpson for synthesizing the oligonucleotide primers and R.Jaggi for assistance with the nude mice. We thank G.Mitra, D.Martin-Zanca and M.Barbacid for communicating results before publication. We are grateful to A.Ziemiecki for providing us with the *trk* transfectant cell pictures. We profited from fruitful discussion with D.Martin-Zanca, M.Barbacid, R.Oskam and T.Hunter. We gratefully acknowledge the assistance of C.Wiedmer and T.Diabaté in the preparation of the manuscript and figures.

References

Besmer,P., Murphy,J.E., George,P.C., Qui,F., Bergold,P.J., Lederman,L., Snyder,H.W., Brodeur,D., Zuckerman,E.E. and Hardy,W.D. (1986) *Nature*, **320**, 415–421.

Biggins,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.

Cailleau,R., Young,R., Olivé,M. and Reeves,W.J.,Jr (1974) *J. Natl. Cancer Inst.*, **53**, 661–674.

Chan,A.M.-L., King,H.W.S., Tempest,P.R., Deakin,E.A., Cooper,C.S. and Brookes,P. (1987) *Oncogene*, **1**, 229–233.

Debuire,B., Henry,C., Benaissa,M., Biserte,G., Claverie,J.M., Saule,S., Martin,P. and Stehelin,D. (1984) *Science*, **224**, 1456–1459.

Frischauf,A.M., Lehrach,H., Poustka,A. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827–842.

Geliebter,J., Zeff,R.A., Melvold,R.W. and Nathenson,S.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3371–3375.

Glisin,V., Crkvenjakov,R. and Byus,C. (1974) *Biochemistry*, **13**, 2633–2637.

Hampe,A., Gobet,M., Sherr,C.J. and Galibert,F. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 85–89.

Hunter,T. and Cooper,J.A. (1986) In Boyer,P.D. and Krebs,E.G. (eds), *The Enzymes*. Academic Press, New York., pp. 191–246.

Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D. (ed.), *DNA Cloning Techniques*. IRL Press, Oxford, pp. 39–78.

Hynes,N.E., Jaggi,R., Kozma,S.P., Ball,R., Muellener,D., Wetherall,N.T., Davis,B.W. and Groner,B. (1985) *Mol. Cell. Biol.*, **5**, 268–272.

Konopka,J.B., Watanabe,S.M. and Witte,O.N. (1984) *Cell*, **37**, 1035–1042.

Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.

Kozma,S.C., Bogaard,M.E., Buser,K., Saurer,S.M., Bos,J.L., Groner,B. and Hynes,N.E. (1987) *Nucleic Acids Res.*, **15**, 5963–5971.

Martin-Zanca,D., Hughes,S.H. and Barbacid,M. (1986) *Nature*, **319**, 743–748.

McMaster,G. and Carmichael,G. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 4835–4838.

Naharro,G., Robbins,K.C. and Reddy,E.P. (1984) *Science*, **223**, 63–66.

Nilsen,T.W., Maroney,P.A., Goodwin,R.G., Rottman,F.M., Crittenden,L.B., Raines,M.A. and Kung,H.-J. (1985) *Cell*, **41**, 719–726.

Park,M., Dean,M., Keul,K., Braun,M.J., Gouda,M.A. and Vande Woude,G.F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6379–6383.

Reddy,E.P., Reynolds,R.K., Santos,E. and Barbacid,M. (1982) *Nature*, **300**, 149–152.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.

Schechter,A.L., Stern,D.F., Vaidyanathan,L., Decker,S.J., Drebin,J.A., Greene,M.I. and Weinberg,R.A. (1984) *Nature*, **312**, 513–516.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.

Stanton,V.P. and Cooper,G.M. (1987) *Mol. Cell. Biol.*, **7**, 1171–1179.

Tabin,C.J., Bradley,S.M., Bargmann,C.I., Weinberg,R.A., Papageorge,A.G., Scolnick,E.M., Dhar,R., Lowy,D.R. and Chang,E.H. (1982) *Nature*, **300**, 143–149.

Tahira,T., Ochiai,M., Hayashi,K., Nagao,M. and Sugimura,T. (1987) *Nucleic Acids Res.*, **15**, 4809–4820.

Takahashi,M. and Cooper,G.M. (1987) *Mol. Cell. Biol.*, **7**, 1378–1385.
Wang,J.Y.J., Ledley,F., Goff,S., Lee,R., Groner,Y. and Baltimore,D. (1984) *Cell*, **36**, 349–356.
Wickens,M. and Stephenson,P. (1984) *Science*, **226**, 1045–1051.

Received on September 10, 1987; revised on October 19, 1987

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00100.