MOLECULAR AND CELLULAR BIOLOGY, Mar. 1987, p. 1226–1232
0270-7306/87/031226-07$02.00/0
Copyright © 1987, American Society for Microbiology

Vol. 7, No. 3

Rat c-*raf* Oncogene Activation by a Rearrangement That Produces a Fused Protein

FUYUKI ISHIKAWA,[1] FUMIMARO TAKAKU,[2] MINAKO NAGAO,[1*] AND TAKASHI SUGIMURA[1]

Carcinogenesis Division, National Cancer Center Research Institute, Chuo-ku, Tokyo 104,[1] and The Third Department of Internal Medicine, Faculty of Medicine, University of Tokyo, Bunkyo-ku, Tokyo 113,[2] Japan

Received 26 June 1986/Accepted 4 December 1986

In a previous study, activated rat c-*raf* was detected by an NIH 3T3 cell transfection assay, and a rearrangement was demonstrated in the 5′ half of the sequence of the gene. In the present study, the cDNAs of normal and activated rat c-*raf* were analyzed. Results showed that the activated c-*raf* gene is transcribed to produce a fused mRNA, in which the 5′ half of the sequence is replaced by an unknown rat sequence. This mRNA codes a fused c-*raf* protein. The normal and activated c-*raf* cDNAs were each connected to the long terminal repeat of Rous sarcoma virus and transfected into NIH 3T3 cells. Only the activated form had transforming activity. We conclude that the rearrangement is responsible for the activation of c-*raf*.

More than 40 oncogenes have been identified by retroviral transduction or NIH 3T3 cell transfection assay (3). Many of these were found to encode protein kinases that phosphorylate serine, threonine, or tyrosine residues. Some of the tyrosine kinases were shown to be receptors of a growth factor and receptor-related proteins (11, 37). The physiological functions of others are still unknown, but they are thought to be important in intracellular signal transduction (16, 40).

v-*raf* was identified as an oncogene of a murine transforming retrovirus, 3611-MSV (33). c-*raf* is its cellular counterpart. Recently it was reported that activated c-*raf* was identified in transformants induced by a human gastric cancer and a glioblastoma line (13, 38). The genomic and cDNA structures of human c-*raf* were reported (6, 7), but little is known about its physiological function except that viral *raf* protein fused with *gag* shows serine-threonine kinase activity in vitro (26).

Previously we found that rat c-*raf* is activated during transfection of rat hepatocellular carcinoma DNA into NIH 3T3 cells (19), that the activated c-*raf* has a rearrangement, and that its 5′ half is replaced by another sequence (18). Here we report that cDNA structures of the normal and activated rat c-*raf* and show that the rearrangement is responsible for the acquirement of transforming activity.

MATERIALS AND METHODS

cDNA cloning. Poly(A)$^+$ RNA was prepared as described previously (12) from the liver of a normal Fischer 344 rat 2 months old and a primary NIH 3T3 cell transformant, IQ7-2, which was shown to have active rat c-*raf* (19). cDNA libraries were constructed from poly(A)$^+$ RNA from both sources with λ gt10 as described previously (17). Normal rat c-*raf* cDNAs were cloned by screening the cDNA library of normal Fischer 344 rat liver with the v-*raf*-specific probe v-*raf* XB (32). cDNAs of activated rat c-*raf* were cloned by screening the IQ7-2 cDNA library with v-*raf* XB and the pRAH insert as probes. pRAH is a subclone of the activated rat c-*raf* genomic sequence containing the recombination point (18). The cDNA inserts obtained were subcloned in pGEM2 (Promega Biotec, Madison, Wis.).

DNA sequencing. The DNAs to be sequenced were subcloned in mp18, and the inserts were progressively deleted with exonucleases III and VII as described previously (42). Each deletant was sequenced by the dideoxy-termination method (34). All sequences were determined by analyzing both complementary strands.

Computer-assisted analysis of DNA and protein sequences. Sequence homology was searched with the GenBank (2) database using the programs of SEQF and SEQFP, IDEAS (21).

DNA transfection. Samples of 0.01 to 10 μg of purified plasmids were supplemented with carrier salmon sperm DNA to 60 μg and transfected into NIH 3T3 cells by the calcium phosphate precipitation method (41). Transformants were scored after 14 to 21 days.

RESULTS

Molecular cloning of normal and activated rat c-*raf* cDNAs. We have reported that c-*raf* is actively expressed in normal rat liver (18). To obtain normal rat c-*raf* cDNA, we constructed a cDNA library of normal Fischer 344 rat liver. Eight of 100,000 clones of this cDNA library hybridized with the v-*raf*-specific probe, and the clone with the longest insert was subcloned in pGEM2 and named pR3.

To obtain activated c-*raf* cDNA clones, we screened the cDNA library made from the primary NIH 3T3 transformant IQ7-2. We have reported that the activated rat c-*raf* of IQ7-2 has a rearrangement and that its 5′ half sequence is replaced by an unrelated sequence. The insert of pRAH is derived from the genomic DNA of activated rat c-*raf*. The insert contains the recombination point, but not a sequence homologous to v-*raf* XB. RNA blot analysis with pRAH as a probe indicated that the insert hybridized with a 3.0-kilobase (kb) transcript specific to IQ7-2. This 3.0-kb transcript seemed also to be hybridized with v-*raf* XB (18), so we could clone the activated c-*raf* cDNA by the criterion of hybridization with both v-*raf* XB and pRAH and exclude normal NIH 3T3, mouse c-*raf* cDNA clones.

About 150 of 90,000 clones of the IQ7-2 cDNA library hybridized with both v-*raf* XB and pRAH. Eight of them

* Corresponding author.

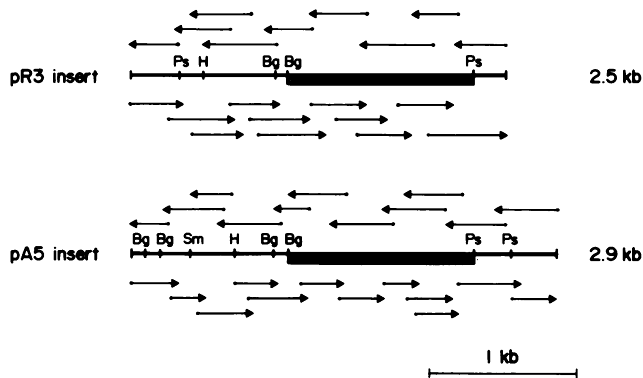

FIG. 1. Restriction maps of the normal rat c-*raf* cDNA pR3 insert and activated rat c-*raf* cDNA pA5 insert. The hatched boxes indicate the fragments that were shown to be homologous to v-*raf* XB by Southern blot analysis. The strategies of sequencing are indicated above and below the maps. Restriction enzymes: Bg, *Bgl*I; H, *Hind*III; Ps, *Pst*I; Sm, *Sma*I.

were cloned, and the clone with the longest insert was subcloned in pGEM2 and named pA5. The high cloning efficiency of the activated form relative to the normal form was thought to be due to the amplification and overexpression of the activated rat c-*raf* in IQ7-2.

Structures of normal and activated rat c-*raf* cDNAs. pR3 and pA5 have 2.5- and 2.9-kb inserts, respectively. The restriction maps of the inserts in Fig. 1 show that the activated and normal forms have different sequences in one half. This, together with our previous data (18), clearly indicates that the activated rat c-*raf* has a rearrangement and forms a fused mRNA.

We determined the complete nucleotide sequences of the inserts of pR3 and pA5 (Fig. 2). The sequencing strategies are indicated in Fig. 1. pR3 has a 2,525-base-pair (bp) insert without a poly(A) tail, and pA5 has an insert of about 2.9 kb with a poly(A) tail. RNA blot analysis revealed that the normal and activated c-*raf* transcripts are 3.2 and 3.0 kb, respectively (18). So, pA5 is a nearly full-length cDNA, whereas pR3 is a partial one, missing about 0.7 kb.

Comparison of the two sequences showed that the 3'-half nucleotide sequences of the pR3 and pA5 inserts are identical. The breakpoints of this identity are located between nucleotides 872 and 873 in pR3 and 898 and 899 in pA5. The 5' sequences of the two beyond these points are quite different, and the sequence of pA5 is transcribed from the sequence recombined to c-*raf*.

The pR3 insert has a long open reading frame starting at nucleotide 42 and ending at nucleotide 1985. We could not find a stop codon in frame with this open reading frame from nucleotide 1 to 41. So we could not define the translation start point as definitely nucleotide 42. However, recently Bonner et al. reported the nucleotide sequence of a human c-*raf* cDNA clone (7) and determined a unique long open reading frame with an in-frame stop codon 12 nucleotides upstream of the initiating ATG. Our tentative open reading frame has the same nucleotide number and shows high homology with the human sequence, and this homology diminishes in the region upstream of nucleotide 42 of the pR3 insert. Moreover, the predicted molecular weight of the protein deduced from this open reading frame is consistent with the value of the rat c-*raf* protein observed with immunoblot analysis (27). Thus we concluded that the open reading frame of pR3 is the coding sequence of the normal rat c-*raf* product.

In pA5, we could find only one long open reading frame, starting at nucleotide 206 and ending at nucleotide 2011. There is an in-frame stop codon TAG 195 bp upstream of this ATG codon. We assumed that this open reading frame is actually translated.

Deduced amino acid sequences of normal and activated rat c-*raf* proteins. Figure 3 shows the deduced amino acid sequences of the normal and activated rat c-*raf* proteins. The putative molecular weights of the normal and activated products are 72,927 and 69,209 respectively. These values may represent the actual molecular weights of the proteins, because these proteins are unlikely to be subject to post-translational modification as described below.

The normal c-*raf* product of rats differs from the reported human sequence in 11 amino acids but has the same total number of residues, 648 amino acids (7). The *raf* protein is a member of the kinase family and has two sequences conserved in this family. One is a consensus sequence of the ATP-binding site, Gly-X-Gly-XX-Gly-$(X)_{10-22}$-Lys (20), which starts at amino acid residue 356. The other is a conserved kinase domain sequence (16), and a putative autophosphorylated serine is identified at residue 499 (7).

The activated c-*raf* product has 602 amino acid residues. Its carboxyl half is identical with that of the normal product. This identity starts at amino acid residues 278 and 232 of the normal and activated products, respectively. The sequences of the amino halves of the two c-*raf* products beyond these points show no homology. Thus in the amino terminal, about half the sequence of the normal c-*raf* product is replaced by an unrelated sequence to form a fused protein of activated c-*raf*. The ATP-binding site and kinase domain are also found in the deduced amino acid sequence of activated rat c-*raf*. The rearrangement in the activated form leaves these two domains intact, at least as a primary structure.

We examined the local hydrophobicities of these two deduced amino acid sequences by the method of Kyte and Doolittle (24), but we found no markedly hydrophobic region. Since a hydrophobic sequence is a hallmark of signals of transfer to the membrane and anchoring in the membrane (5, 30), these normal and activated c-*raf* proteins are unlikely to be bound to the membrane with these signal sequences. The transforming protein of Rous sarcoma virus, $pp60^{src}$, is known to be myristylated and to become bound to the membrane (14, 36). In that case, the amino-terminal glycine next to the initiating methionine is myristylated. In the normal and activated forms of the c-*raf* product, the amino acid residue next to the initiating methionine is not glycine, and these proteins are also unlikely to be myristylated at their amino termini. These observations suggest that both forms of the c-*raf* product are soluble proteins present in the cytosol and that the localization of the protein is not changed by activation. The presence of the normal rat c-*raf* product in the cytosol was shown by Mölders et al. (27).

Upstream sequence of activated rat c-*raf* cDNA is derived from an unknown rat sequence. As previously reported, no rearrangement of activated rat c-*raf* was detected in the original tumor, and so probably it occurred during the transfection assay (18). Thus the DNA sequence upstream of the recombination junction may be derived from NIH 3T3 or transfected Fischer 344 rat tumor DNA. To determine which is the case, we used a 0.7-kb *Eco*RI-*Hind*III fragment of pA5 as a probe (Fig. 4). This fragment is located upstream of the recombination junction. On hybridization of the gel-fractionated *Hind*III digest of DNA of primary transformant

```
pR3        1 GCCCGGCTCCCTCAGGTATAAGTATTGTTGAAATTGCATCAATGAGCACATACAGGGAGCTTGGAAGACGATC     74
pA5     1 GCCGTGCGCATAGAGGCCGGTGCGCG----TTGCT-G-TTAACGCGG-AC-ATA-TTCCCA-GG--CG-C-C-GGAG-CTCC--C-GGCA-GC-CGCGGG   100

  75 AGCAATGGTTTCGGACTCAAAGATGCTGTGTTTGATGGCTCCAGTTGCATCTCCCCTACAATTGTTCAGCAGTTTGGCTATCAGCGTCGGGCCTCTGATG   174
 101 --ACTGC-AGCGA-G-G-CG-CGG-GC-GC-CA-GC-T--GGGTCC--GCA--T---TGCTCCT-CGCTTCTCC-TCAGCCGCTGC-GCCA-GA-CCCG-   200

 175 ATGGCAAACTCACGGATTCTTCTAAGACAAGCAATACTATCCGGGTTTTCTTGCCGAATAAGCAAAGGACTGTGGTCAATGTGCGGAATGGGATGAGCTT   274
 201 CC-A-CTGGCGG---TG-TGCAGC-AGTGCTGG-GCGCCCGGA-C-GAA-AA--T-CC----TCG-CCCAGAACAAACT--A-AA-TTCCT-GCTGAACA   300

 275 GCACGACTGCCTTATGAAAGCTCTGAAGGTTAGAGGCCTGCAGCCAGAGTGCTGTGCAGTGTTCAGACTTCTTCAGGAGCACAAAGGTAAGAAAGCACGC   374
 301 ---GTC-GAAA-CGACTGCCTGAA-GG-CGGCAC-AGAAATTTAAG-T-GAGA---AGCAACAATACT--GAGATA---A-G-G-CTATCCC-GAGT-AG   400

 375 TTAGATTGGAACACCGACGCCGCCTCTCTGATTGGAGAAGAACTGCAAGTGGATTTTTTGGATCACGTTCCACTCACAACTCACAACTTTGCTCGGAAAA   474
 401 GAGAGGCTTGTT-AT--AA--CGGGAGTGTCAGAACTTGAGG---G-GC-T--GAAGC-AA--A--CAAGT-AAAGT-TTAACTG-GAAAAACAAAG--C   500

 475 CGTTCCTGAAGCTTGCATTCTGTGACATCTGTCAAAAGTTCCTGCTAAATGGATTTCGATGTCAGACTTGTGGCTACAAGTTTCATGAGCACTGTAGCAC   574
 501 TTGAAACTGCT-AA-ACCG-AA-CTAGG-AT---G-GCCAGT-TAC--GA-C-AAGGA-GAGTTAGAAGC--AAA-A-GAGA-TTAATCAGAACC-ATGA   600

 575 CAAAGTACCTACTATGTGTGTGGACTGGAGTAATATCAGACAGCTCTTGCTGTTTCCAAATTCCACTGCAAGTGACAGTGGAGTCCCAGCACCACCCTCT   674
 601 G-GGT--T--CAGGAAGT--AAT-T-TA-CAG-GGATGTTA-A-GTC-AAACGAAAA-CT-AAAGAAAGC-A-ACA-CGAAG-GTGA-CTT-AGTTAAAG   700

 675 TTCACAATGCGTCGGATGCGAGAATCTGTTTCCCGGATGCCTGCTAGTTCCCAGCACAGATACTCCACACCCCATGCCTTCACTTTCAACACCTCCAGCC   774
 701 C-GGATGAA-T--AAGCTTCT--TGTCAC-GTGAA-TAC-GA-AA-AACG-TTAG-ACA-G-AAAGGA-TTG-TACA-AATCAAAATTCATGGCTA-A-A   800

 775 CTTCCTCTGAAGGTTCCCTGTCCCAGAGGCAGAGGTCAACGTCCACTCCCAATGTCCACATGGTCAGCACCACCCTGCCTGTGGACAGCAGGATGATTGA   874
 801 -AGAG-TGA--ACCAAAACTGATG--CTATT-GCTCT-GGAAGAGAAAGGGAAATG-A--TC-GGAACTT-AGTGTA--C-T--A-A--AA-A-GAA--   900

 875 GGATGCAATTCGAAGTCACAGTGAATCAGCCTCACCTTCAGCCCTGTCCAGCAGCCCCAACAACCTGAGCCCAACAGGCTGGTCACAGCCCAAAACCCCT   974
 901 ----------------------------------------------------------------------------------------------------  1000

                               v-raf start
 975 GTGCCAGCACAAAGAGAGAGGGCGCCAGGATCTGGGACCCAGGAAAAAAACAAAATTAGGCCTCGTGGGCAGAGAGATTCAAGTTATTACTGGGAAATAG  1074
1001 ----------------------------------------------------------------------------------------------------  1100

1075 AAGCCAGTGAGGTGATGCTGTCTACTCGGATTGGCTCGGGCTCCTTTGGCACTGTGTACAAGGGCAAGTGGCATGGAGATGTTGCAGTAAAGATCCTAAA  1174
1101 ----------------------------------------------------------------------------------------------------  1200

1175 GGTGGTTGACCCAACTCCAGAGCAACTTCAGGCCTTCAGGAACGAGGTGGCTGTTTTGCGCAAAACACGGCATGTTAATATCCTGCTGTTCATGGGGTAC  1274
1201 ----------------------------------------------------------------------------------------------------  1300

1275 ATGACAAAGGACAACCTGGCGATTGTGACCCAGTGGTGTGAAGGCAGCAGTCTCTACAAACACCTGCATGTCCAGGAGACCAAATTCCAGATGTTCCAGC  1374
1301 ----------------------------------------------------------------------------------------------------  1400

1375 TAATTGACATTGCCCGGCAGACAGCTCAGGGAATGGACTATTTACATGCAAAGAACATCATCCACAGAGACATGAAATCCAACAATATATTTCTCCATGA  1474
1401 ----------------------------------------------------------------------------------------------------  1500

1475 AGGCCTCACGGTGAAAATCGGAGATTTTGGTTTGGCAACAGTGAAGTCGCGCTGGAGTGGTTCTCAGCAGGTTGAACAGCCCACTGGCTCTGTGCTGTGG  1574
1501 ----------------------------------------------------------------------------------------------------  1600

1575 ATGGCCCCAGAAGTAATCCGAATGCAGGATAACAACCCGTTCAGCTTCCAGTCCGATGTCTACTCCTATGGCATTGTGCTGTATGAGCTGATGACTGGGG  1674
1601 ----------------------------------------------------------------------------------------------------  1700

1675 AGCTTCCCTACTCCCACATCAACAACCGAGACCAGATCATCTTCATGGTGGGCCGTGGGTACGCCTCCCCAGATCTTAGCAGGCTCTACAAGAACTGCCC  1774
1701 ----------------------------------------------------------------------------------------------------  1800

1775 CAAGGCAATGAAGAGGTTGGTGGCTGACTGTGTGAAGAAAGTCAAAGAAGAAAGGCCTTTGTTTCCTCAGATCCTGTCTTCCATTGAGCTGCTTCAGCAC  1874
1801 ----------------------------------------------------------------------------------------------------  1900

1875 TCTCTGCCGAAAATCAACAGGAGCGCCTCTGAGCCTTCCCTGCATCGGGCAGCTCACACTGAGGACATCAATGCTTGTACGCTGACCACATCCCCAAGGC  1974
1901 ----------------------------------------------------------------------------------------------------  2000

1975 TACCAGTCTTCTAGCTGACGTTATAGCTGTTCTTAGGCCACCAGGGGACGAAGAAGAGTCAGCAGGCACCACTTTCTGTTTCCTTGGGGGCAGAATGCAT  2074
2001 ----------------------------------------------------------------------------------------------------  2100

2075 GTTTCCGGAAAAGCTGCTGCTAAGGACCTAGACTACTCACAGGGCCTTAACTTCATATTGCCTTCTTTTCTACCCTTCCTGCCCTGGAAATGGAAGCTGT  2174
2101 ----------------------------------------------------------------------------------------------------  2200

2175 CCGCCAAGCCAGCCTGCTCCAGAGGTATACAAGTCAGCGAGTATTTTTAGGGCAAATGGCCTTGGAGAGAGAAGGCAGGGCACTCCGGCTACTGCAGGGA  2274
2201 ----------------------------------------------------------------------------------------------------  2300

2275 CATGCAGTTGGGAACTTGGCTCATTGAGCTGTACAGACAGTGGTGCAGTGCCAGTTTTGCACATGGAGTCCTGGCCACCTGGGGGAGCCTGCTTTGGTAC  2374
2301 ----------------------------------------------------------------------------------------------------  2400

2375 TACAGAACTTCACTTTGTGGACACACCTTCCTCTTACTGAGTCTAAGATGTCCTGTGCAGAGGATGCTTTCCAAGCACGGTGCTCCACCTTCTGGCAGCC  2474
2401 ----------------------------------------------------------------------------------------------------  2500

2475 TCCCACACGCTGAATCTGTCTTCCAGGAGCTGCCCTATGGGGTGCTGCAGC    2525
2501 ---------------------------------------------------CCAGCCCTATCTCTATAGTCACATCCTTGTCTGTAAGAAAGCCAGGAAT  2600

2601 ACAGGTTTTCTTAATGATTTTGGGTTTTAATTTTGTTTTTATTGAGCCTGATAAAATACAGTTATCTGATGGTTCCTCAATTATGTTATTTTAATAAAAT  2700

2701 AAATTAAATTTAAAAAAAA(A)n
```

```
Normal       1  MEHIQGAWKTISNGFGLKDAVFDGSSCISPTIVQQFGYQRRASDDGKLTDSSKTSNTIRV  60
Activated                                                  1  MAAVLQQVLERPEL  14

            61  FLPNKQRTVVNVRNGMSLHDCLMKALKVRGLQPECCAVFRLLQEHKGKKARLDWNTDAAS  120
            15  NKLP-STQNKLEKFLAEQQSEIDCLKGRHEKFKVESEQQYFEI-KRLSQSQERLVNETRE  74

           121  LIGEELQVDFLDHVPLTTHNFARKTFLKLAFCDICQKFLLNGFRCQTCGYKFHEHCSTKV  180
            75  CQNLR-ELEK-NNQVKVLTEKNKELETAQDRNLGI-SQFTRAKEELEAEKRDLIRTNERL  134

           181  PTMCVDWSNIRQLLLFPNSTASDSGVPAPPSFTMRRMRESVSRMPASSQHRYSTPHAFTF  240
           135  SQEVEYLTEDVKR-NEKLKESNTTKGELQLKLDELQASDVTVKYREKRLEQEKELLHNQN  194

           241  NTSSPSSEGSLSQRQRSTSTPNVHMVSTTLPVDSRMIEDAIRSHSESASPSALSSSPNNL  300
           195  SWLNTELKTKTDELLALGREKGNEILELKCTLENKKE-----------------------  254

           301  SPTGWSQPKTPVPAQRERAPGSGTQEKNKIRPRGQRDSSYYWEIEASEVMLSTRIGSGSF  360
           255  ------------------------------------------------------------  314

           361  GTVYKGKWHGDVAVKILKVVDPTPEQLQAFRNEVAVLRKTRHVNILLFMGYMTKDNLAIV  420
           315  ------------------------------------------------------------  374

           421  TQWCEGSSLYKHLHVQETKFQMFQLIDIARQTAQGMDYLHAKNIIHRDMKSNNIFLHEGL  480
           375  ------------------------------------------------------------  434

                             *
           481  TVKIGDFGLATVKSRWSGSQQVEQPTGSVLWMAPEVIRMQDNNPFSFQSDVYSYGIVLYE  540
           435  ------------------------------------------------------------  494

           541  LMTGELPYSHINNRDQIIFMVGRGYASPDLSRLYKNCPKAMKRLVADCVKKVKEERPLFP  600
           495  ------------------------------------------------------------  554

           601  QILSSIELLQHSLPKINRSASEPSLHRAAHTEDINACTLTTSPRLPVF  648
           555  ------------------------------------------------  602
```

FIG. 3. Deduced amino acid sequences of normal and activated rat c-*raf* products. The amino acid residue numbers are indicated. The two sequences are aligned to show the identical carboxyl halves. Homologous residues are indicated by dashes. Consensus sequences of the ATP-binding site (20) are indicated by boxes. The putative autophosphorylation site (6) is indicated by an asterisk.

IQ7-2 with this probe, we detected two bands of 6.0 and 4.5 kb besides NIH 3T3-specific bands (lanes b and c). These additional bands were exactly the same as those obtained from Fischer 344 rat DNA (lane a). Thus the upstream sequence was derived from rat genomic DNA.

To test whether the upstream sequence was derived from an actively transcribed gene, RNA blot analysis was performed with the 0.7-kb *Eco*RI-*Hin*dIII fragment of pA5 as a probe (Fig. 5). Two transcripts of 9 to 10 kb and 7 to 8 kb were detected in NIH 3T3 cells (lane a). In the primary transformant, 4.9- and 3.0-kb transcripts were found in addition to those found in NIH 3T3 cells (lane b), and the v-*raf* probe also hybridized with the mRNAs of the same size (18). In contrast to NIH 3T3 cells, the upstream sequence was scarcely transcribed in normal rat liver (lane c). Thus, c-*raf* was rearranged to a gene which is expressed in NIH 3T3 cells but not in normal rat liver. The abundance of 4.9- and 3.0-kb fused mRNA relative to 9- to 10- and 7- to 8-kb mRNAs in the primary transformant was thought to be due in part to the amplification of activated c-*raf* in the transformant. Since we have some transformants in which activated c-*raf* was not amplified (data not shown), we believe that the amplification itself is not responsible for the transformation.

To obtain information about the upstream sequence, we searched for homologous DNA and protein sequences in the GenBank (2) database using the programs of SEQF and SEQFP, IDEAS (21). No significantly homologous sequence was detected in the database, so the nature of the protein coded by the upstream sequence is unknown.

Recombination occurs in the intron between exons 7 and 8 of c-*raf*. As stated above, the rat and human c-*raf* cDNA sequences show high homology. Assuming that the exon-intron organizations in these two species are similar, we located the exon-intron junctions to our sequence according to those of humans (7). We found that the normal and activated rat c-*raf* cDNA sequences start identically at three nucleotides upstream of the junction of exons 7 and 8 of c-*raf* (Fig. 6A; the numbering of exons is according to reference 7). To determine whether the recombination actually occurred in exon 7, we determined the partial nucleotide sequences of the c-*raf* intron between exons 7 and 8 and of its counterpart in the activated form. These sequences are present in the inserts of pNAH and pRAH, which were

FIG. 2. Nucleotide sequences of pR3 and pA5 inserts. The nucleotide numbers of both sequences are indicated. The two sequences are aligned to show the identical 3′ halves. Homologous nucleotides are indicated by dashes. The initiating ATGs and stop codon TAGs of the long open reading frames are boxed. An in-frame stop codon TAG, 195 bp upstream of the initiating ATG of the pA5 insert, is indicated by a wavy underline. The start point of identity of the two sequences is indicated by a heavy arrow. The point at which the homology with v-*raf* starts (6) is indicated by a thin arrow. Two polyadenylation signals of the pA5 insert are shown underlined.

described previously (18). We found that the 3′ halves of these introns have identical sequences but that their 5′-half sequences are quite different (Fig. 6B). As the nucleotide sequence of the normal gene containing the region recombined to c-*raf* has not been determined, the absolute position of recombination is still unknown. However, we concluded that the recombination occurred in the intron between exons 7 and 8 of c-*raf*.

We could determine the splice donor site of the exon just upstream of the recombination point in the activated form. The sequence has good consistency with the reported consensus sequence (28), and the exon(s) upstream of this donor site is derived from the gene recombined to c-*raf*. The last three nucleotides of exon 7 of c-*raf* and of the preceding exon of this donor site were thought to be the same by chance. Identity of the introns of the normal and activated forms starts 141 bp downstream of this donor site.

In the intron of normal rat c-*raf*, we found an inverted repeat, CAGGAT and ATCCTG, separated by one nucleotide near the recombination point (Fig. 6B). These two sequences have motifs which are partially identical with the consensus sequences GAGG and CCTC found in retrovirus recombination sites (see reference 1 for a review). GAGG was also found to be present at the c-*myc* translocation site in murine plasmacytomas (31). These sequences possibly form a secondary structure and may have some relation with the recombination.

Cloned activated rat c-*raf* cDNA has transforming activity. To test whether cloned activated c-*raf* cDNA has a sequence essential for transforming activity, we constructed plasmids to be expressed in NIH 3T3 cells. We replaced the neomycin resistant sequence of pRSVneo (15) with the inserts of pR3 and pA5. In the latter case, the insert was digested with *Pst*I before insertion. This process eliminates the polyadenylation signal and tail from the pA5 insert and ensures use of the splicing and polyadenylation signal intrinsic to pRSVneo. The pR3 insert has no polyadenylation signal, and the whole insert was used. The plasmids constructed in this way were transfected into NIH 3T3 cells with carrier salmon sperm DNA (Table 1). The plasmid expressing the activated form showed significant transforming activity (about 100 to 150 focus-forming units per μg of DNA), whereas up to 10 μg of DNA from the vector alone or a plasmid expressing normal c-*raf* cDNA gave no transformants. Thus we concluded that the pA5 insert has the entire sequence essential for transforming activity. As pR3 and pA5 inserts have the same coding sequences except in the substituted region, the rearrangement was responsible for activation of rat c-*raf*.

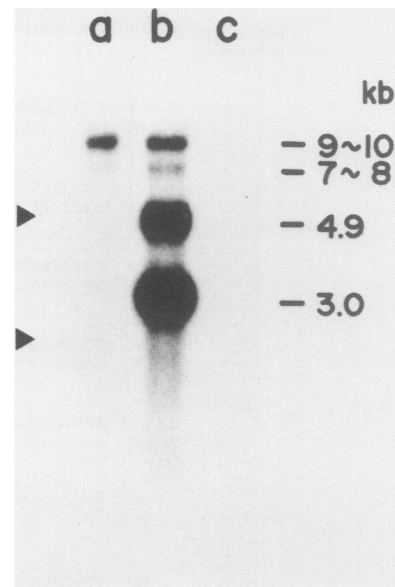

FIG. 5. RNA blot analysis with a 5′-half cDNA probe. Samples of 2 μg of poly(A)$^+$ RNAs were denatured, subjected to electrophoresis in 0.8% denaturing gel, and analyzed by blot hybridization with the 0.7-kb *Eco*RI-*Hin*dIII probe shown in Fig. 4. Lanes: a, NIH 3T3 cells; b, primary transformant IQ7-2; c, normal Fischer 344 rat liver. The positions of 28S and 18S rRNA are indicated by triangles.

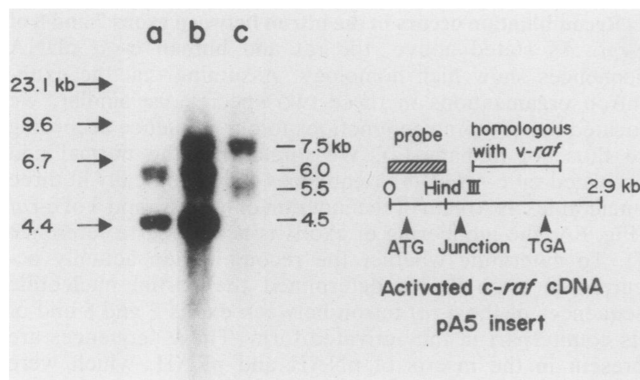

FIG. 4. Southern blot analysis with a 5′-half cDNA probe. The 0.7-kb *Eco*RI-*Hin*dIII probe is indicated in the right panel. Samples of 10 μg of DNA were digested with *Hin*dIII and analyzed by Southern blot hybridization (18). Lanes: a, normal Fischer 344 rat liver; b, primary transformant IQ7-2; c, NIH 3T3 cells. The positions of fragments of λ *c*I857 DNA digested with *Hin*dIII are shown by arrows.

DISCUSSION

Because most oncogenes belonging to the kinase family were identified by retroviral transduction, their activation mechanisms were mainly investigated on retroviral v-*onc* sequences (4). This approach has been fruitful, but it has one disadvantage. During retroviral transduction, many mutations are introduced in v-*onc*; some of these mutations may actually render c-*onc* active, but others may not. Thus, in studies on v-*onc*, it is somewhat difficult to identify the genetic change actually responsible for transforming activity.

Recently some activated oncogenes of the kinase family were found by using an NIH 3T3 cell transfection assay of tumor DNA (1, 10, 13, 19, 35, 38). In these instances, the genetic change is anticipated to be more simple than with v-*onc*. Moreover, studies on the activation mechanism of c-*onc* may provide information about not only the mechanisms of cancer development but also the physiological roles of these oncogene products.

In the present work, we showed that a rearrangement to produce a fused protein is responsible for activation of c-*raf*. Although this activation did not occur in the original cancer tissue in vivo, we think that this is the first demonstration that a rearrangement with another cellular sequence alone can make c-*onc* active.

Many v-*onc*s have truncated 5′ structures and form fused proteins with viral elements. v-*erbB* is a 5′-truncated version of the epidermal growth factor receptor gene (11). v-*fgr* fuses with a cellular DNA sequence which codes cytoskeleton

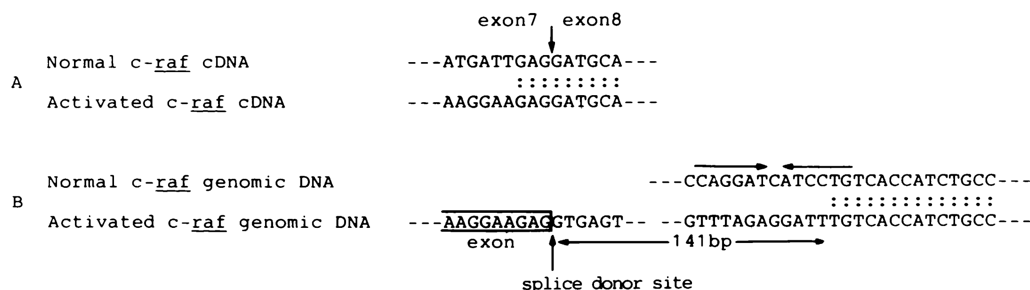

FIG. 6. Regional nucleotide sequences around the recombination point. Identical nucleotides are indicated by dots. (A) cDNAs; (B) genomic DNAs. The junction of exons 7 and 8 described previously (7) is indicated in A. An inverted repeat is indicated by underlining with arrows in B.

protein in its 5′ terminus (29). Recently c-*abl* was found to recombine with *bcr* to form a fused transcript in chronic myelogenous leukemia, although in this case, c-*abl* is not "activated" to have transforming activity (39). More recently, Martin-Zanka et al. identified a new oncogene, *trk*, as a member of the tyrosine kinase family and found that it forms a fused transcript with the tropomyosin gene by rearrangement (25). Some workers have emphasized the role of structural changes in 3′ termini of tyrosine kinase genes. Replacement of the extreme carboxyl terminus of v-*src* seems important for its activation (8). The v-*fms* product has a truncated carboxyl terminus (9). There is thus much indirect evidence that altered 5′-half structures or 3′-terminus truncation are common features of activation of kinase oncogenes.

Thus, qualitative change in the amino termini seems important for acquirement of transforming activity in some kinases. Indeed a regulator element has been suggested to be present in the amino-terminal half of c-*raf* (7). This element may regulate the kinase activity of the c-*raf* protein and inhibit the activation. In this sense, it is noteworthy that the recombination point found in our study was upstream of the ATP-binding domain and the kinase domain. The rearrangement may leave the kinase machinery intact but remove or change the regulatory element. It is also noteworthy that v-*raf* is truncated at a position corresponding to that 133 bp downstream of the recombination point found in our study. Another possibility is that altered amino-terminal structures change the location and substrates of kinases. In our case, there was no evidence for an altered location of the product at the level of the primary structure. The substrates of normal and activated c-*raf* kinases require identification.

The present findings do not exclude the possibility that a quantitative change of the c-*raf* product resulting from the rearrangement is important, whereas a qualitative change is not. The normal c-*raf* transcript has an atypical sequence just before its initiating ATG. According to Kozak (23), the sequence CATCA is not a good candidate for translation initiation by eucaryotic ribosomes. So the c-*raf* protein may not be abundant in normal cells, although the transcript is easily detected by RNA blot analysis. In contrast, the activated form has a sequence that agrees with Kozak's consensus sequence (22). Thus the rearrangement may enhance translation, resulting in production of abundant protein and activation of the gene.

TABLE 1. Transforming activities of normal and activated c-*raf* cDNA in NIH 3T3 cells

DNA	Expt 1[a]		Expt 2[a]	
	DNA (μg)[b]	No. of foci	DNA (μg)[b]	No. of foci
Activated c-*raf* cDNA[c]	1	141	10	419
	0.1	25	1	107
	0.01	4		
Normal c-*raf* cDNA[c]	1	0	10	0
	0.1	0	1	0
	0.01	0		
Vector alone[d]	1	0	10	0
	0.1	0	1	0
	0.01	0		

[a] Results of two independent experiments are shown.
[b] Salmon sperm DNA was added to each plasmid in a total of 60 μg and was transfected into cells in three dishes.
[c] Inserts of pR3 and pA5 were ligated to the *Hin*dIII-*Sma*I digest of pRSVneo (15). Clones oriented correctly to be expressed were transfected.
[d] pRSVneo was digested with *Hin*dIII and *Sma*I to delete most of the sequence of the neomycin resistance gene. The plasmid was religated and used in the transfection assay.

ACKNOWLEDGMENTS

We thank U. R. Rapp for providing cloned v-*raf*. We also thank Kenshi Hayashi for helpful discussion and advice.

This study was supported by a Grant-in-Aid from the Ministry of Health and Welfare for the Comprehensive 10-Year Strategy for Cancer Control, Japan, and grants from Sankyo Foundation Life Science and Yamanouchi Foundation for Research on Metabolic Disease.

LITERATURE CITED

1. **Besmer, P., J. E. Murphy, P. C. George, F. Qiu, P. J. Bergold, L. Lederman, H. W. Snyder, Jr., D. Brodeur, E. E. Zuckerman, and W. D. Hardy.** 1986. A new acute transforming feline retrovirus and relationship of its oncogene v-*kit* with the protein kinase gene family. Nature (London) **320**:415–421.
2. **Bilofsky, H. S., C. Burks, J. W. Fickett, W. B. Goad, F. I. Lewitter, W. P. Rindone, C. D. Swindell, and C.-S. Tung.** 1986. The GenBank[R] genetic sequence databank. Nucleic Acids Res. **14**:1–4.
3. **Bishop, J. M.** 1983. Cellular oncogenes and retroviruses. Annu. Rev. Biochem. **52**:301–354.
4. **Bishop, J. M.** 1985. Viral oncogenes. Cell **42**:23–38.
5. **Blobel, G.** 1980. Intracellular protein topogenesis. Proc. Natl. Acad. Sci. USA **77**:1496–1500.
6. **Bonner, T. I., S. B. Kerby, P. Sutrave, M. A. Gunnell, G. Mark, and U. R. Rapp.** 1985. Structure and biological activity of human homologs of the *raf/mil* oncogene. Mol. Cell. Biol. **5**: 1400–1407.
7. **Bonner, T. I., H. Oppermann, P. Seeburg, S. B. Kerby, M. A. Gunnell, A. C. Young, and U. R. Rapp.** 1986. The complete coding sequence of the human *raf* oncogene and the corresponding structure of the c-*raf*-1 gene. Nucleic Acids Res. **14**:1009–1015.
8. **Cooper, J. A., K. L. Gould, C. A. Cartwright, and T. Hunter.** 1986. Tyr^{527} is phosphorylated in $pp60^{c\text{-}scr}$: implications for regulation. Science **231**:1431–1434.

9. **Coussens, L., C. V. Beveren, D. Smith, E. Chen, R. L. Mitchell, C. M. Isacke, I. M. Verma, and A. Ullrich.** 1986. Structural alteration of viral homologue of receptor proto-oncogene *fms* at carboxyl terminus. Nature (London) **320**:277–280.
10. **Dean, M., M. Park, M. M. Le Beau, T. S. Robins, M. O. Diaz, J. D. Rowley, D. G. Blair, and G. F. Vande Woude.** 1985. The human *met* oncogene is related to the tyrosine kinase oncogenes. Nature (London) **318**:385–388.
11. **Downward, J., Y. Yarden, E. Mayes, G. Scrace, N. Totty, P. Stockwell, A. Ullrich, J. Schlessinger, and M. D. Waterfield.** 1984. Close similarity of epidermal growth factor receptor and v-*erb-B* oncogene protein sequences. Nature (London) **307**:521–527.
12. **Feramisco, J. R., J. E. Smart, K. Burridge, D. M. Helfman, and G. P. Thomas.** 1982. Co-existence of vinculin and a vinculin-like protein of higher molecular weight in smooth muscle. J. Biol. Chem. **257**:11024–11031.
13. **Fukui, M., T. Yamamoto, S. Kawai, K. Maruo, and K. Toyoshima.** 1985. Detection of a *raf*-related and two other transforming DNA sequences in human tumors maintained in nude mice. Proc. Natl. Acad. Sci. USA **82**:5954–5958.
14. **Garber, E. A., J. G. Krueger, H. Hanafusa, and A. R. Goldberg.** 1983. Only membrane-associated RSV *src* proteins have amino-terminally bound lipid. Nature (London) **302**:161–163.
15. **Gorman, C., R. Padmanabhan, and B. H. Howard.** 1983. High efficiency DNA-mediated transformation of primate cells. Science **221**:551–553.
16. **Hunter, T., and J. A. Cooper.** 1985. Protein-tyrosine kinases. Annu. Rev. Biochem. **54**:897–930.
17. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening cDNA libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning, vol. 1. IRL Press, Oxford.
18. **Ishikawa, F., F. Takaku, K. Hayashi, M. Nagao, and T. Sugimura.** 1986. Activation of rat c-*raf* during transfection of hepatocellular carcinoma DNA. Proc. Natl. Acad. Sci. USA **83**:3209–3212.
19. **Ishikawa, F., F. Takaku, M. Ochiai, K. Hayashi, S. Hirohashi, M. Terada, S. Takayama, M. Nagao, and T. Sugimura.** 1985. Activated c-*raf* gene in a rat hepatocellular carcinoma induced by 2-amino-3-methylimidazo[4,5-*f*]quinoline. Biochem. Biophys. Res. Commun. **132**:186–192.
20. **Kamps, M. P., S. S. Taylor, and B. M. Sefton.** 1984. Direct evidence that oncogenic tyrosine kinases and cyclic AMP-dependent protein kinase have homologous ATP-binding sites. Nature (London) **310**:589–592.
21. **Kanehisa, M.** 1984. Computer analysis and structure prediction of nucleic acids and proteins. Nucleic Acids Res. **12**:417–428.
22. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
23. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
25. **Martin-Zanka, D., S. H. Hughes, and M. Barbacid.** 1986. A human oncogene formed by the fusion of truncated tropomyosin and protein tyrosine kinase sequences. Nature (London) **319**:743–748.
26. **Moelling, K., B. Heimann, P. Beimling, U. R. Rapp, and T. Sander.** 1984. Serine- and threonine-specific protein kinase activities of purified gag-mil and gag-raf proteins. Nature (London) **312**:558–561.
27. **Mölders, H., J. Defesche, D. Müller, T. I. Bonner, U. R. Rapp, and R. Müller.** 1985. Integration of transfected LTR sequences into the c-*raf* proto-oncogene: activation by promoter insertion. EMBO J. **4**:693–698.
28. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
29. **Naharro G., K. C. Robbins, and E. P. Reddy.** 1984. Gene product of v-*fgr onc*: hybrid protein containing a portion of actin and a tyrosine-specific protein kinase. Science **223**:63–66.
30. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
31. **Piccoli, S. P., P. G. Caimi, and M. D. Cole.** 1984. A conserved sequence at c-*myc* oncogene chromosomal translocation breakpoints in plasmacytomas. Nature (London) **310**:327–330.
32. **Rapp, U. R., M. D. Goldsborough, G. E. Mark, T. I. Bonner, J. Groffen, F. H. Reynolds, Jr., and J. R. Stephenson.** 1983. Structure and biological activity of v-*raf*, a unique oncogene transduced by a retrovirus. Proc. Natl. Acad. Sci. USA **80**:4218–4222.
33. **Rapp, U. R., F. H. Reynolds, and J. R. Stephenson.** 1983. New mammalian transforming retrovirus: demonstration of a polyprotein gene product. J. Virol. **45**:914–924.
34. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
35. **Schechter, A. L., D. F. Stern, L. Vaidyanathan, S. J. Decker, J. A. Drebin, M. I. Greene, and R. A. Weinberg.** 1984. The *neu* oncogene: an *erb-B*-related gene encoding a 185,000-*M*r tumor antigen. Nature (London) **312**:513–516.
36. **Sefton, B. M., I. S. Trowbridge, and J. A. Cooper.** 1982. The transforming proteins of Rous sarcoma virus, Harvey sarcoma virus and Abelson virus contain tightly bound lipid. Cell **31**:465–474.
37. **Sherr, C. J., C. W. Rettenmier, R. Sacca, M. F. Roussel, A. T. Look, and E. R. Stanley.** 1985. The c-*fms* proto-oncogene product is related to the receptor for the mononuclear phagocyte growth factor, CSF-1. Cell **41**:665–676.
38. **Shimizu, K., Y. Nakatsu, M. Sekiguchi, K. Hokamura, K. Tanaka, M. Terada, and T. Sugimura.** 1985. Molecular cloning of an activated human oncogene, homologous to v-*raf*, from primary stomach cancer. Proc. Natl. Acad. Sci. USA **82**:5641–5645.
39. **Shtivelman, E., B. Lifshitz, R. P. Gale, and E. Canaani.** 1985. Fused transcript of *abl* and *bcr* genes in chronic myelogenous leukemia. Nature (London) **315**:550–554.
40. **Weinberg, R. A.** 1985. The action of oncogenes in the cytoplasm and nucleus. Science **230**:770–776.
41. **Wigler, M., S. Silverstein, L.-S. Lee, A. Pellicer, Y. Cheng, and R. Axel.** 1977. Transfer of purified herpes virus thymidine kinase gene to cultured mouse cells. Cell **11**:223–232.
42. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of M13mp18 and pUC19 vectors. Gene **33**:103–119.