MOLECULAR AND CELLULAR BIOLOGY, Apr. 1987, p. 1378–1385
0270-7306/87/041378-08$02.00/0

Vol. 7, No. 4

ret Transforming Gene Encodes a Fusion Protein Homologous to Tyrosine Kinases

MASAHIDE TAKAHASHI† AND GEOFFREY M. COOPER*

Dana-Farber Cancer Institute and Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115

Received 2 September 1986/Accepted 16 December 1986

The *ret* transforming gene was activated by recombination between two unlinked segments of human DNA, most likely during transfection of NIH 3T3 cells. To further define this transforming gene, we isolated and sequenced *ret* cDNA clones. The nucleotide sequence indicates that the active *ret* transforming gene encodes a fusion protein with a carboxy-terminal domain which is 40 to 50% homologous to members of the tyrosine kinase gene family. This tyrosine kinase domain is preceded by a hydrophobic sequence characteristic of a transmembrane domain. Transcription of the *ret* tyrosine kinase sequence was detected in the SK-N-SH neuroblastoma, HL-60 promyelocytic leukemia, and THP-1 monocytic leukemia cell lines, but not in 25 other human tumor cell lines surveyed. The *ret* tyrosine kinase may thus represent a cell surface receptor which is expressed in a restricted range of human cells.

The *ret* transforming gene was detected by transfection of NIH 3T3 cells with DNA of a human T-cell lymphoma (22). This lymphoma DNA yielded only a single focus in primary transfection assays (corresponding to approximately 0.003 transformants per μg of DNA), whereas transformed NIH cell DNAs induced transformation with high efficiencies (approximately 0.5 transformants per μg of DNA) in secondary and tertiary transfection assays (22). The *ret* gene was cloned from transformed NIH cells by hybridization with human repetitive sequence probes (22). Analysis of the cloned sequence indicated that the active *ret* transforming gene encompassed approximately 34 kilobases (kb) of human DNA and was generated by recombination of two DNA segments (each approximately 17 kb) which were unlinked in DNA of either normal human cells or the T-cell lymphoma which gave rise to the initial focus of transformed NIH cells (22). Both segments of *ret* were cotranscribed in transformed NIH cells, indicating that the transforming gene had been activated by a recombination event which generated a new transcriptional unit (22). The low primary transforming efficiency and the absence of rearranged *ret* sequences in the original lymphoma DNA suggested that *ret* represented a normal human gene whose transforming potential was activated by a DNA rearrangement during transfection (22).

To further characterize the recombination event which resulted in *ret* activation, we cloned and sequenced a *ret* cDNA. Analysis of this cDNA sequence indicated that *ret* encodes a fusion protein, the carboxy-terminal half of which is a new member of the tyrosine kinase family.

MATERIALS AND METHODS

Cell lines. NIH 3T3 cells transformed by *ret* in tertiary transfections were previously described (22). Human tumor cell lines included carcinomas of lung (SK-LC-3, SK-MES-1, SBC-3, and NCI-H69), bladder (EJ), stomach (MKN-1), and colon (HT-29); a hepatoma (HepG2); T-cell leukemias (Molt-4, CCRF HSB-2, CCRF CEM, and JM); Burkitt lymphomas (Raji and CW678); a promyelocytic leukemia (HL-60); a monocytic leukemia (THP-1); a histiocytic lymphoma (U937); neuroblastomas (SK-N-SH, SK-N-MC, and Goto); glioblastomas (U251 MG, MG 178, SK-MG-1, SK-MG-4, and SK-MG-6); osteosarcomas (SAOS-2 and U-2 0S); and a malignant melanoma (SK-MEL-37).

cDNA cloning. Total cellular RNA was isolated by guanidinium-cesium chloride centrifugation (25). Poly(A)$^+$ RNA was isolated by a single passage through oligo(dT)-cellulose and used as a template for cDNA synthesis. The first strand was synthesized as described by Maniatis et al. (8), and the second strand was synthesized by the method of Okayama and Berg (14). Single-stranded ends were filled in with T4 polymerase, and the cDNA was methylated at 37°C for 1 h with *Eco*RI methylase (New England BioLabs, Inc., Beverly, Mass.). *Eco*RI linkers (New England BioLabs) were added by ligation at 14°C overnight, DNA was digested with *Eco*RI, and free linkers were removed by passage through a Sepharose CL-4B column. The purified cDNA was ligated to *Eco*RI arms of λgt10 (Cloning Vector System), and the resulting library was screened by plaque hybridization with genomic fragments of *ret* DNA (22).

DNA sequencing. The *ret* cDNA inserts were subcloned in *Eco*RI sites of pBR322 and sequenced by the procedure of Maxam and Gilbert (11).

Southern blot analysis. Cellular DNA (10 μg) was digested with appropriate restriction endonucleases, electrophoresed in agarose gels (0.7%, wt/vol), and transferred to Gene Screen Plus filters (New England Nuclear Corp., Boston, Mass). Prehybridization, hybridization, and washes were performed under the conditions suggested by the supplier. The *ret* cDNA fragments were purified from low-melting-point agarose gels and ^{32}P-labeled by nick translation for use as probes (22).

Northern blot analysis. Poly(A)$^+$ RNA (5 μg) was electrophoresed in formaldehyde-agarose gels (1%, wt/vol) and transferred to Gene Screen filters (New England Nuclear Corp.) in the presence of 0.025 M Na_2HPO_4-NaH_2PO_4 (pH 6.5). The resulting filters were prehybridized, hybridized, and washed as described previously (22).

RESULTS

Isolation and characterization of *ret* cDNA. The *ret* gene was transcribed into three major poly(A)$^+$ RNAs (approximately 3, 4, and 6 kb) in transformed NIH cells (22). To

* Corresponding author.
† Present address: Laboratory of Experimental Pathology, Aichi Cancer Center Research Institute, Chikusa-ku, Nagoya 464, Japan.

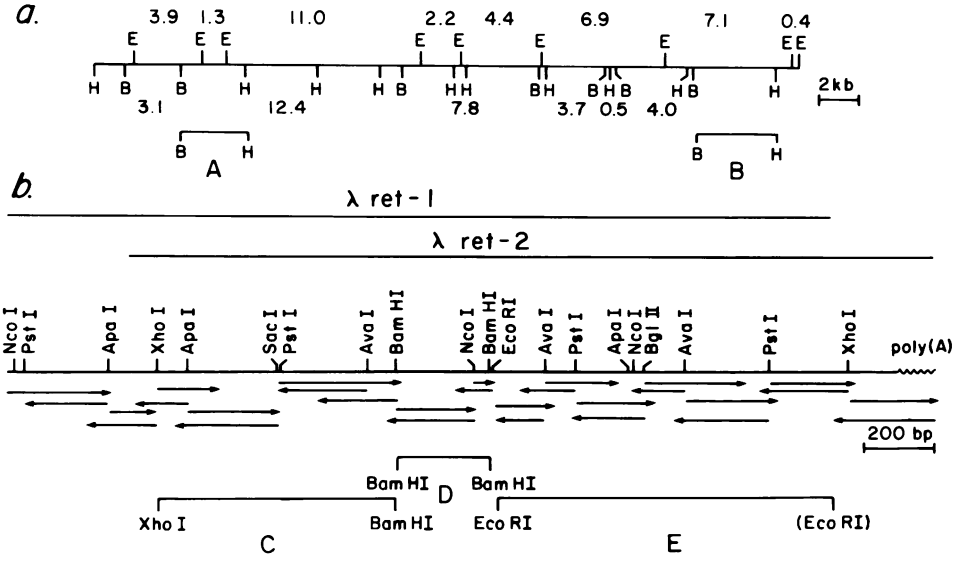

FIG. 1. Restriction map of *ret* genomic and cDNA clones. (a) Restriction map of *ret* genomic DNA (22). The sizes of *Eco*RI (top) and *Bam*HI (bottom) fragments are shown in kilobases. B, *Bam*HI; H, *Hin*dIII; E, *Eco*RI. (b) Poly(A)$^+$ RNA from *ret*-transformed NIH cells (22) was used as the template for oligo(dT)-primed synthesis of double-stranded cDNA which was cloned in λgt10. The recombinant phage library was screened by plaque hybridization with fragments A and B (panel a) as probes. The restriction maps of two overlapping cDNA clones (lambda *ret*-1 and lambda *ret*-2) are shown. The *Eco*RI site in parenthesis indicates the 3′ terminus of the lambda *ret*-1 insert. Arrows below the maps indicate nucleotide sequencing strategy. bp, Base pairs.

isolate *ret* cDNA clones, poly(A)$^+$ RNA of transformed NIH cells was used as the template for oligo(dT)-primed cDNA synthesis, and a cDNA library was constructed in λgt10. The library of 2×10^5 recombinant phage was screened with two genomic fragments (A and B in Fig. 1a) from opposite ends of the *ret* gene. Five plaques which hybridized with both fragments were detected. Each of these clones contained an insert of 2.2 to 2.4 kb in which one internal *Eco*RI site was present. Two of the clones (designated lambda *ret*-1 and lambda *ret*-2) were further characterized. These two clones overlapped and together spanned a total length of 2.6 kb (Fig. 1b). Thus, they included several hundred nucleotides less than the smallest *ret* mRNA.

Three fragments of the cDNA clones (C, D, and E in Fig. 1b) were used as probes in Southern blot analysis to determine which regions of the cDNA represented each of the two human segments which made up the recombinant *ret* gene. Fragment C hybridized to 12.4- and 7.8-kb *Bam*HI fragments in DNA of transformed NIH cells and to 19.0-, 12.4-, and 3.9-kb *Bam*HI fragments in DNA of human cells (Fig. 2A). The 12.4- and 7.8-kb *Bam*HI fragments in NIH transformant DNA represent the left-hand human segment and the recombined region of *ret*, respectively (22). The 19.0- and 12.4-kb *Bam*HI fragments of human DNA correspond to the left-hand segment of *ret*, and the 3.9-kb fragment corresponds to the right-hand segment (22). The 19.0- and 3.9-kb *Bam*HI fragments of human DNA recombined to generate the 7.8-kb *Bam*HI fragment of the active *ret* gene (22). Fragment C of the *ret* cDNA thus includes the left-hand segment of *ret*, the recombination site, and some sequence from the right-hand segment.

Fragment D of *ret* cDNA hybridized to the 7.8-kb *Bam*HI fragment in DNA of transformed NIH cells and to the 3.9-kb *Bam*HI fragment in human DNA (Fig. 2B). Fragment E hybridized to a doublet of approximately 7 kb in *Eco*RI-digested DNAs of transformed NIH cells and human cells, in addition to three fragments in DNA of NIH 3T3 cells (Fig. 2C). Since fragment E also hybridized (data not shown) to both the isolated 6.9- and 7.1-kb *Eco*RI fragments of the genomic *ret* clone (Fig. 1a), these fragments formed the 7.0-kb *Eco*RI doublet detected in the genomic blots. These results indicated that fragments D and E of the *ret* cDNA represented the right-hand segment of genomic *ret*.

Sequence of *ret* cDNA. The inserts of lambda *ret*-1 and lambda *ret*-2 were subcloned in pBR322, and their complete nucleotide sequence was determined by the Maxam-Gilbert method (11) by using the strategy shown in Fig. 1b. The nucleotide and deduced amino acid sequences are shown in Fig. 3.

A single major open reading frame extends from the 5′ end of the cloned sequence (corresponding to the left-hand

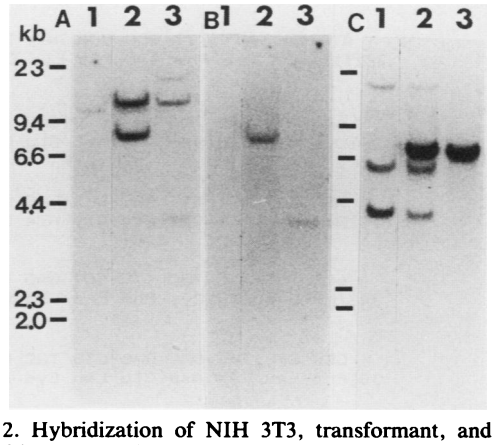

FIG. 2. Hybridization of NIH 3T3, transformant, and human DNAs with cDNA probes. DNAs (10 μg) of NIH 3T3 cells (lanes 1), an NIH tertiary transformant induced by *ret* (lanes 2), and human cells (lanes 3) were digested with *Bam*HI (A and B) or *Eco*RI (C) and analyzed by blot hybridization with fragments of the *ret* cDNA clones as probes (Fig. 1b). (A) Fragment C; (B) fragment D; and (C) fragment E. The sizes of *Hin*dIII-digested lambda DNA markers are indicated.

```
                                               30                                60
   CAG GCC GGC GCC ATG GCC TCC GGG AGT GTG GCC GAG TGC CTG CAG CAG GAG ACC ACC TGC
  1Gln Ala Gly Ala Met Ala Ser Gly Ser Val Ala Glu Cys Leu Gln Gln Glu Thr Thr Cys

                                               90                               120
   CCC GTG TGC CTG CAG TAC TTC GCA GAG CCC ATG ATG CTC GAC TGC GGC CAT AAC ATC TGT
 21Pro Val Cys Leu Gln Tyr Phe Ala Glu Pro Met Met Leu Asp Cys Gly His Asn Ile Cys

                                              150                               180
   TGC GCG TGC CTC GCC CGC TGC TGG GGC ACG GCA GAG ACT AAC GTG TCG TGC CCG CAG TGC
 41Cys Ala Cys Leu Ala Arg Cys Trp Gly Thr Ala Glu Thr Asn Val Ser Cys Pro Gln Cys

                                              210                               240
   CGG GAG ACC TTC CCG CAG AGG CAC ATG CGG CCC AAC CGG CAC CTG GCC AAC GTG ACC CAA
 61Arg Glu Thr Phe Pro Gln Arg His Met Arg Pro Asn Arg His Leu Ala Asn Val Thr Gln

                                              270                               300
   CTG GTA AAG CAG CTG CGC ACC GAG CGG CCG TCG GGG CCC GGC GGC GAG ATG GGC GTG TGC
 81Leu Val Lys Gln Leu Arg Thr Glu Arg Pro Ser Gly Pro Gly Gly Glu Met Gly Val Cys

                                              330                               360
   GAG AAG CAC CGC GAG CCC CTG AAG CTG TAC TGC GAG GAG GAC CAG ATG CCC ATC TGC GTG
101Glu Lys His Arg Glu Pro Leu Lys Leu Tyr Cys Glu Glu Asp Gln Met Pro Ile Cys Val

                                              390                               420
   GTG TGC GAC CGC TCC CGC GAG CAC CGC GGC CAC AGC GTG CTG CCG CTC GAG GAG GCG GTG
121Val Cys Asp Arg Ser Arg Glu His Arg Gly His Ser Val Leu Pro Leu Glu Glu Ala Val

                                              450                               480
   GAG GGC TTC AAG GAG CAA ATC CAG AAC CAG CTC GAC CAT TTA AAA AGA GTG AAA GAT TTA
141Glu Gly Phe Lys Glu Gln Ile Gln Asn Gln Leu Asp His Leu Lys Arg Val Lys Asp Leu

                                              510                               540
   AAG AAG AGA CGT CGG GCC CAG GGG GAA CAG GCA CGA GCT GAA CTC TTG AGC CTA ACC CAG
161Lys Lys Arg Arg Arg Ala Gln Gly Glu Gln Ala Arg Ala Glu Leu Leu Ser Leu Thr Gln

                                              570                               600
   ATG GAG AGG GAG AAG ATT GTT TGG GAG TTT GAG CAG CTG TAT CAC TCC TTA AAG GAG CAT
181Met Glu Arg Glu Lys Ile Val Trp Glu Phe Glu Gln Leu Tyr His Ser Leu Lys Glu His

                                              630                               660
   GAG TAT CGC CTC CTG GCC CGC CTT GAG GAG CTA GAC TTG GCC ATC TAC AAT AGC ATC AAT
201Glu Tyr Arg Leu Leu Ala Arg Leu Glu Glu Leu Asp Leu Ala Ile Tyr Asn Ser Ile Asn

                                              690                               720
   GGT GCC ATC ACC CAG TTC TCT TGC AAC ATC TCC CAC CTC AGC AGC CTG ATC GCT CAG CTA
221Gly Ala Ile Thr Gln Phe Ser Cys Asn Ile Ser His Leu Ser Ser Leu Ile Ala Gln Leu

                                              750                               780
   GAA GAG AAG CAG CAG CAG CCC ACC AGG GAG CTC CTG CAG GAC ATT GGG GAC ACA TTG AGC
241Glu Glu Lys Gln Gln Gln Pro Thr Arg Glu Leu Leu Gln Asp Ile Gly Asp Thr Leu Ser

                                              810                               840
   AGG GCT GAA AGA ATC AGG ATT CCT GAA CCT TGG ATC ACA CCT CCA GAT TTG CAA GAG AAA
261Arg Ala Glu Arg Ile Arg Ile Pro Glu Pro Trp Ile Thr Pro Pro Asp Leu Gln Glu Lys

                                              870                               900
   ATC CAC ATT TTT GCC CCA AAA TGT CTA TTC TTG ACG GAG AGT CTA AAG CAG TTC ACA GAA
281Ile His Ile Phe Ala Pro Lys Cys Leu Phe Leu Thr Glu Ser Leu Lys Gln Phe Thr Glu

                                              930                               960
   AAA ATG CAG TCA GAT ATG GAG AAA ATC CAA GAA TTA AGA GAG GCT CAG TTA TAC TCA GGG
301Lys Met Gln Ser Asp Met Glu Lys Ile Gln Glu Leu Arg Glu Ala Gln Leu Tyr Ser Gly

                                              990                              1020
   GGC AGC ATT GTT GGG GGA CAC GAG CCT GGG GAG CCC CGG GGG ATT AAA GCT GGC TAT GGC
321Gly Ser Ile Val Gly Gly His Glu Pro Gly Glu Pro Arg Gly Ile Lys Ala Gly Tyr Gly

                                             1050                              1080
   ACC TGC AAC TGC TTC CCT GAG GAG GAG AAG TGC TTC TGC GAG CCC GAA GAC ATC CAG GAT
341Thr Cys Asn Cys Phe Pro Glu Glu Glu Lys Cys Phe Cys Glu Pro Glu Asp Ile Gln Asp

                                             1110                              1140
   CCA CTG TGC GAC GAG CTG TGC CGC ACG GTG ATC GCA GCC GCT GTC CTC TTC TCC TTC GTC
361Pro Leu Cys Asp Glu Leu Cys Arg Thr Val Ile Ala Ala Ala Val Leu Phe Ser Phe Val

                                             1170                              1200
   GTC TCG GTG CTG CTG TCT GCC TTC TGC ATC CAC TGC TAC CAC AAG TTT GCC CAC AAG CCA
381Val Ser Val Leu Leu Ser Ala Phe Cys Ile His Cys Tyr His Lys Phe Ala His Lys Pro

                                             1230                              1260
   CCC ATC TCC TCA GCT GAG ATG ACC TTC CGG AGG CCC GCC CAG GCC TTC CCG GTC AGC TAC
401Pro Ile Ser Ser Ala Glu Met Thr Phe Arg Arg Pro Ala Gln Ala Phe Pro Val Ser Tyr
```

```
                                        1290                                      1320
    TCC TCT TCC GGT GCC CGC CGG CCC TCG CTG GAC TCC ATG GAG AAC CAG GTC TCC GTG GAT
421 Ser Ser Ser Gly Ala Arg Arg Pro Ser Leu Asp Ser Met Glu Asn Gln Val Ser Val Asp

                                        1350                                      1380
    GCC TTC AAG ATC CTG GAG GAT CCA AAG TGG GAA TTC CCT CGG AAG AAC TTG GTT CTT GGA
441 Ala Phe Lys Ile Leu Glu Asp Pro Lys Trp Glu Phe Pro Arg Lys Asn Leu Val Leu Gly

                                        1410                                      1440
    AAA ACT CTA GGA GAA GGC GAA TTT GGA AAA GTG GTC AAG GCA ACG GCC TTC CAT CTG AAA
461 Lys Thr Leu Gly Glu Gly Glu Phe Gly Lys Val Val Lys Ala Thr Ala Phe His Leu Lys

                                        1470                                      1500
    GGC AGA GCA GGG TAC ACC ACG GTG GCC GTG AAG ATG CTG AAA GAG AAC GCC TCC CCG AGT
481 Gly Arg Ala Gly Tyr Thr Thr Val Ala Val Lys Met Leu Lys Glu Asn Ala Ser Pro Ser

                                        1530                                      1560
    GAG CTT CGA GAC CTG CTG TCA GAG TTC AAC GTC CTG AAG CAG GTC AAC CAC CCA CAT GTC
501 Glu Leu Arg Asp Leu Leu Ser Glu Phe Asn Val Leu Lys Gln Val Asn His Pro His Val

                                        1590                                      1620
    ATC AAA TTG TAT GGG GCC TGC AGC CAG GAT GGC CCG CTC CTC CTC ATC GTG GAG TAC GCC
521 Ile Lys Leu Tyr Gly Ala Cys Ser Gln Asp Gly Pro Leu Leu Leu Ile Val Glu Tyr Ala

                                        1650                                      1680
    AAA TAC GGC TCC CTG CGG GGC TTC CTC CGC GAG AGC CGC AAA GTG GGG CCT GGC TAC CTG
541 Lys Tyr Gly Ser Leu Arg Gly Phe Leu Arg Glu Ser Arg Lys Val Gly Pro Gly Tyr Leu

                                        1710                                      1740
    GGC AGT GGA GGC AGC CGC AAC TCC AGC TCC CTG GAC CAC CCG GAT GAG CGG GCC CTC ACC
561 Gly Ser Gly Gly Ser Arg Asn Ser Ser Ser Leu Asp His Pro Asp Glu Arg Ala Leu Thr

                                        1770                                      1800
    ATG GGC GAC CTC ATC TCA TTT GCC TGG CAG ATC TCA CAG GGG ATG CAG TAT CTG GCC GAG
581 Met Gly Asp Leu Ile Ser Phe Ala Trp Gln Ile Ser Gln Gly Met Gln Tyr Leu Ala Glu

                                        1830                                      1860
    ATG AAG CTC GTT CAT CGG GAC TTG GCA GCC AGA AAC ATC CTG GTA GCT GAG GGG CGG AAG
601 Met Lys Leu Val His Arg Asp Leu Ala Ala Arg Asn Ile Leu Val Ala Glu Gly Arg Lys

                                        1890                                      1920
    ATG AAG ATT TCG GAT TTC GGC TTG TCC CGA GAT GTT TAT GAA GAG GAT TCC TAC GTG AAG
621 Met Lys Ile Ser Asp Phe Gly Leu Ser Arg Asp Val Tyr Glu Glu Asp Ser Tyr Val Lys

                                        1950                                      1980
    AGG AGC CAG GGT CGG ATT CCA GTT AAA TGG ATG GCA ATT GAA TCC CTT TTT GAT CAT ATC
641 Arg Ser Gln Gly Arg Ile Pro Val Lys Trp Met Ala Ile Glu Ser Leu Phe Asp His Ile

                                        2010                                      2040
    TAC ACC ACG CAA AGT GAT GTA TGG TCT TTT GGT GTC CTG CTG TGG GAG ATC GTG ACC CTA
661 Tyr Thr Thr Gln Ser Asp Val Trp Ser Phe Gly Val Leu Leu Trp Glu Ile Val Thr Leu

                                        2070                                      2100
    GGG GGA AAC CCC TAT CCT GGG ATT CCT CCT GAG CGG CTC TTC AAC CTT CTG AAG ACC GGC
681 Gly Gly Asn Pro Tyr Pro Gly Ile Pro Pro Glu Arg Leu Phe Asn Leu Leu Lys Thr Gly

                                        2130                                      2160
    CAC CGG ATG GAG AGG CCA GAC AAC TGC AGC GAG GAG ATG TAC CGC CTG ATG CTG CAA TGC
701 His Arg Met Glu Arg Pro Asp Asn Cys Ser Glu Glu Met Tyr Arg Leu Met Leu Gln Cys

                                        2190                                      2220
    TGG AAG CAG GAG CCG GAC AAA AGG CCG GTG TTT GCG GAC ATC AGC AAA GAC CTG GAG AAG
721 Trp Lys Gln Glu Pro Asp Lys Arg Pro Val Phe Ala Asp Ile Ser Lys Asp Leu Glu Lys

                                        2250                                      2280
    ATG ATG GTT AAG AGG AGA GAC TAC TTG GAC CTT GCG GCG TCC ACT CCA TCT GAC TCC CTG
741 Met Met Val Lys Arg Arg Asp Tyr Leu Asp Leu Ala Ala Ser Thr Pro Ser Asp Ser Leu

                                        2310                                      2340
    ATT TAT GAC GAC GGC CTC TCA GAG GAG GAG ACA CCG CTG GTG GAC TGT AAT AAT GCC CCC
761 Ile Tyr Asp Asp Gly Leu Ser Glu Glu Glu Thr Pro Leu Val Asp Cys Asn Asn Ala Pro

                                        2370                                      2400
    CTC CCT CGA GCC CTC CCT TCC ACA TGG ATT GAA AAC AAA CTC TAT GGT AGA ATT TCC CAT
781 Leu Pro Arg Ala Leu Pro Ser Thr Trp Ile Glu Asn Lys Leu Tyr Gly Arg Ile Ser His

                                        2430                                      2460
    GCA TTT ACT AGA TTC TAG CAC CGC TGT CCC CTC TGC ACT ATC CTT CCT CTC TGT GAT GCT
801 Ala Phe Thr Arg Phe End

    TTT TAA AAA TGT TTC TGG TCT GAA CA AAA(A)n
```

FIG. 3. Nucleotide sequence of *ret* cDNA. The nucleotide sequence was determined by the method of Maxam and Gilbert (11) according to the strategy indicated in Fig. 1b. Both nucleotide and predicted amino acid sequences are shown. Nucleotides are numbered above the sequence, and amino acids are numbered in the left-hand margin. The putative transmembrane and tyrosine kinase domains are indicated by dashed and solid underlines, respectively.

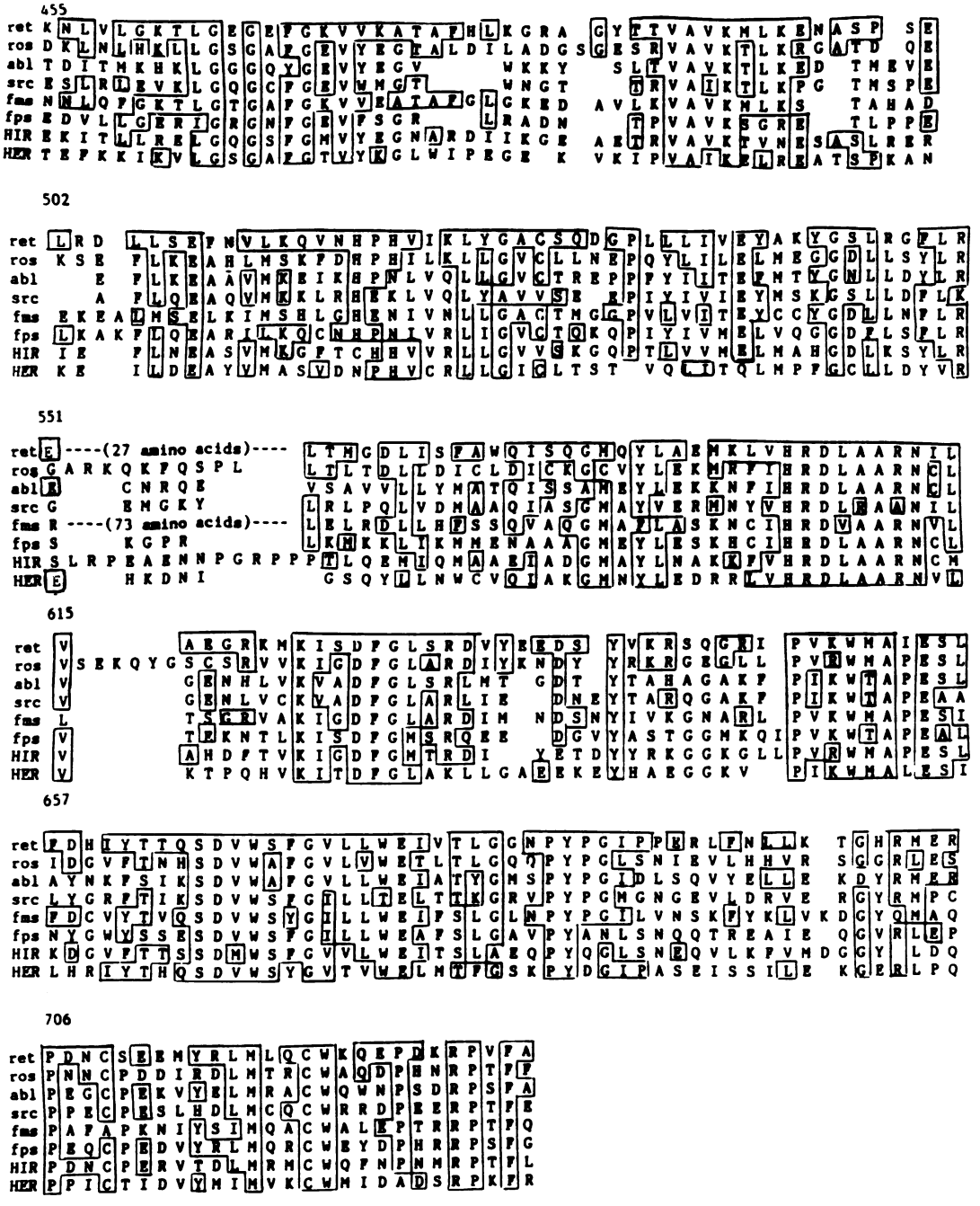

FIG. 4. Homology of *ret* to tyrosine kinases. The homologous region of the predicted *ret* amino acid sequence is aligned with those of the tyrosine kinase family members *ros* (13), *abl* (17), *src* (19), *fms* (5), *fps* (21), human insulin receptor (HIR) (23), and human epidermal growth factor receptor (HER) (24). Numbers refer to the *ret* amino acid sequences. Amino acid identities are enclosed in boxes.

terminus) to the termination codon TAG at nucleotide 2416. This termination codon is followed by 67 nucleotides of 3′ untranslated region and about 100 nucleotides of poly(A). The sequence does not contain a consensus polyadenylation signal (AATAAA), suggesting that the longer species of *ret* mRNA may extend further 3′. Since the total length of the cDNA sequence is less than that of the smallest *ret* mRNA, the *ret* coding sequence may also extend farther in the 5′ direction. The sequence indicates that *ret* has the capacity to encode a protein of at least 800 amino acids. This predicted gene product represents a fusion protein containing recombined coding sequence from both the left-hand (5′) and right-hand (3′) human segments of the *ret* gene, since the site of recombination is between the *Xho*I (nucleotide 406) and *Bam*HI (nucleotide 1077) sites which define the termini of cDNA fragment C.

Homology of *ret* to tyrosine kinase family genes. The *ret* amino acid sequence was compared with those of other known proteins by a computer-assisted homology search with the Beckman Microgenie sequence analysis program. The 5′ portion of the *ret* amino acid sequence did not display significant homology to any known protein. However, a

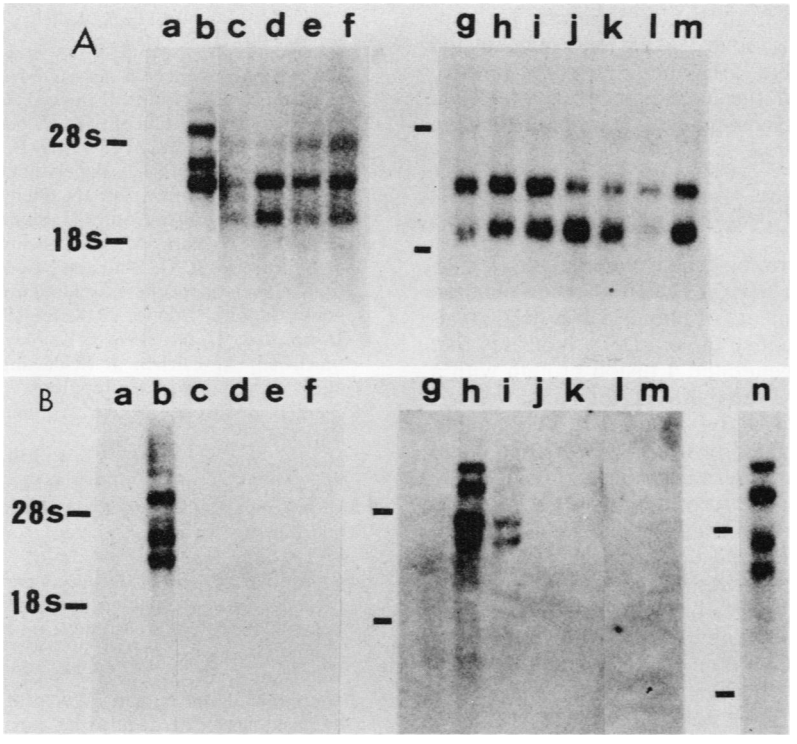

FIG. 5. Analysis of *ret* transcripts with cDNA probes. Poly(A)$^+$ RNAs (5 μg) were analyzed by Northern blot hybridization with *ret* cDNA fragments C (A) and E (B) as probes. The positions of 28S and 18S rRNAs are indicated. RNAs were from NIH 3T3 cells (lane a), NIH *ret* transformant (lane b), EJ bladder carcinoma (lane c), Raji Burkitt lymphoma (lane d), Molt-4 and HSB-2 T-cell leukemias (lanes e and f), U937 histiocytic lymphoma (lane g), THP-1 monocytic leukemia (lane h), HL-60 promyelocytic leukemia (lane i), HepG2 hepatoma (lane j), SK-MES-1 lung carcinoma (lane k), HT-29 colon carcinoma (lane l), MKN-1 stomach carcinoma (lane m), and SK-N-SH neuroblastoma (lane n). Autoradiographic exposures were 16 h for lanes a and b and 2 to 3 days for lanes c to n.

region of the 3′ portion of *ret* from amino acids 455 to 732 (underlined in Fig. 3) derived from the right-hand human segment forming the *ret* gene displayed significant homology to members of the tyrosine kinase gene family.

Analysis of this 278-amino acid domain of *ret* revealed 40 to 50% homology with other tyrosine kinases including *ros* (13), *abl* (17), *src* (19), *fms* (5), *fps* (21), human insulin receptor (23), and human epidermal growth factor receptor (24) (Fig. 4). In addition, it showed similar homology with other recently identified tyrosine kinases including *lsk*$^+$ (9), *met* (3), *neu* (1), *kit* (2), and *trk* (10). The homology of *ret* to these tyrosine kinases includes the lysine thought to be involved in ATP binding (6) (*ret* amino acid 491) and a putative phosphotyrosine acceptor site (16) (*ret* amino acid 638). However, the *ret* sequence is not identical to the sequences of these or any other members of the tyrosine kinase family. Rather, the extent of homology of *ret* to the amino acid sequences of these tyrosine kinases is similar to that between the different members of this family and does not suggest a closer relationship of *ret* to any particular family member.

The predicted *ret* amino acid sequence also includes a stretch of 22 hydrophobic amino acids (369 to 390) which is bounded by basic amino acids and could serve as a membrane-anchoring domain (18). Similar transmembrane domains are found in the other tyrosine kinases which are known cell surface receptors, including the insulin (23) and epidermal growth factor receptors (24) and *fms* (colony-stimulating factor 1 receptor) (5, 20), as well as in *ros* (13), *trk* (10), and *neu* (1). In *ret*, as in each of these other cases, the putative transmembrane domain is located 50 to 70 amino acids upstream of the start of the tyrosine kinase domain.

Transcription of *ret* in human cells. Using fragments of the *ret* genomic clones as probes, we were unable to detect transcription of *ret* sequences in human tumor cell lines (22). However, because of the greater sensitivity provided by cDNA probes, further Northern blot analyses were done to examine the expression of *ret* in human cells. Poly(A)$^+$ RNAs of 28 human cell lines (listed in Materials and Methods) including carcinomas, hemopoietic tumors, nervous system tumors, and sarcomas were examined with *ret* cDNA fragments C and E as probes. These probes hybridized to the previously reported *ret* mRNAs of approximately 3, 4, and 6 kb in transformed NIH cells (Fig. 5A and B, lanes b). In addition, *ret* cDNA fragment C hybridized to less abundant transcripts of approximately 2.4 and 3.4 kb in poly(A)$^+$ RNAs of each of 17 human tumor cell lines tested (representative results in Fig. 5A, lanes c to m). An additional faint band of about 5 kb was also detected in some samples but may represent nonspecific hybridization to 28S rRNA. The human sequence corresponding to the 5′ portion of *ret* was thus transcribed in a wide variety of different human tumor cell lines.

When fragment E of the *ret* cDNA was used as a probe, four transcripts (7.0, 6.0, 4.5, and 3.9 kb) were observed in poly(A)$^+$ RNAs of THP-1 (monocytic leukemia), HL-60 (promyelocytic leukemia), and SK-N-SH (neuroblastoma) cell lines (Fig. 5B, lanes h, i, and n). The level of transcript in HL-60 (lane i) was significantly lower than in THP-1 and

SK-N-SH cell lines (lanes h and n). In contrast, no transcripts were detected in the other 25 cell lines investigated (representative results in Fig. 5B, lanes c to g and j to m). These results indicate that the 3′ sequence of *ret*, which contains the tyrosine kinase domain, is expressed only in restricted human cells.

DISCUSSION

The activated *ret* transforming gene was generated, apparently during transfection of NIH 3T3 cells, by recombination between two unlinked segments of human DNA (22). Analysis of the nucleotide sequence of *ret* cDNA indicates that this recombination event generated a fusion protein with a coding sequence derived from both of the human DNA segments which form the active *ret* gene. The amino-terminal half of *ret* is unrelated to other known amino acid sequences, but the carboxy-terminal half is 40 to 50% homologous to members of the tyrosine kinase gene family, including a variety of retroviral transforming genes, the cell surface receptors for insulin and epidermal growth factor, and the cellular transforming genes *met*, *neu*, and *trk*. Thus, *ret* appears to be a new member of the tyrosine kinase gene family whose transforming potential was activated by fusion of a carboxy-terminal kinase domain with previously unlinked amino-terminal sequences.

Activation of the transforming potential of the *ret* tyrosine kinase by gene fusion is reminiscent of the mode of activation of several other tyrosine kinases, including the epidermal growth factor receptor gene in avian erythroblastosis virus (*erbB*) (4), the *fgr* gene of Gardner-Rasheed feline sarcoma virus (12), the *abl* gene in both Abelson leukemia virus (26) and chronic myelogenous leukemia (7), and the human transforming genes *met* (15) and *trk* (10). In each of these cases, transforming potential is activated by a recombination event which results in deletion of the normal amino-terminal domain of a tyrosine kinase. These observations thus suggest that a variety of tyrosine kinases contain amino-terminal regulatory domains, the loss of which can result in abnormal enzymatic activity leading to cell transformation.

The predicted *ret* amino acid sequence contains a hydrophobic region of 22 amino acids which is bounded by basic amino acids and is located approximately 65 amino acids before the beginning of the tyrosine kinase domain. This hydrophobic region is typical of transmembrane domains (18), and its location in the *ret* sequence is similar to the location of transmembrane domains in those tyrosine kinases (*fms*, epidermal growth factor receptor, and insulin receptor) which are known growth factor receptors (5, 23, 24). This suggests that the normal *ret* tyrosine kinase is a cell surface receptor for a currently unidentified molecule. Since transcription of the *ret* tyrosine kinase domain was detected in only 3 of 28 human tumor cell lines surveyed, it is possible that the *ret* tyrosine kinase is a receptor which normally functions in restricted stages or lineages of differentiation rather than in a wide variety of cell types.

ACKNOWLEDGMENTS

We are grateful to E. Sariban for helpful discussions, to S. O'Keefe for pointing out the *ret* hydrophobic sequence, and to R. Ueda for generously providing human tumor cell lines.

This research was supported by Public Health Service grant CA28946 from the National Institutes of Health and by a faculty research award to G.M.C. from the American Cancer Society.

LITERATURE CITED

1. **Bargmann, C. I., M.-C. Hung, and R. A. Weinberg.** 1986. The *neu* oncogene encodes an epidermal growth factor receptor-related protein. Nature (London) **319**:226–230.
2. **Besmer, P., J. E. Murphy, P. C. George, F. Qui, 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.
3. **Dean, M., M. Park, M. M. LeBeau, T. S. Robins, M. D. 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.
4. **Downward, J., Y. Yarden, E. Meyes, 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.
5. **Hampe, A., M. Gobet, C. J. Sherr, and F. Galibert.** 1984. Nucleotide sequence of the feline retroviral oncogene v-*fms* shows unexpected homology with oncogenes encoding tyrosine specific protein kinases. Proc. Natl. Acad. Sci. USA **81**:85–89.
6. **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.
7. **Konopka, J. B., S. M. Watanabe, and O. N. Witte.** 1984. An alteration of the human c-*abl* protein in K562 leukemia cells unmasks associated tyrosine kinase activity. Cell **37**:1035–1042.
8. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Marth, J. D., R. Peet, E. G. Krebs, and R. M. Perlmutter.** 1985. A lymphocyte-specific protein-tyrosine kinase is rearranged and overexpressed in the murine T cell lymphoma LSTRA. Cell **43**:393–404.
10. **Martin-Zanca, 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.
11. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
12. **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.
13. **Neckameyer, W. S., and L.-H. Wang.** 1985. Nucleotide sequence of avian sarcoma virus UR2 and comparison of its transforming gene with other members of the tyrosine protein kinase oncogene family. J. Virol. **53**:879–884.
14. **Okayama, H., and P. Berg.** 1982. High-efficiency cloning of full-length cDNA. Mol. Cell. Biol. **2**:161–170.
15. **Park, M., M. Dean, C. S. Cooper, M. Schmidt, S. J. O'Brien, D. G. Blair, and G. F. Vande Woude.** 1986. Mechanism of *met* oncogene activation. Cell **45**:895–904.
16. **Patschinsky, T., T. Hunter, F. Esch, J. Cooper, and B. Sefton.** 1982. Analysis of the sequence of amino acids surrounding sites of tyrosine phosphorylation. Proc. Natl. Acad. Sci. USA **79**: 973–977.
17. **Reddy, E. P., M. J. Smith, and A. Srinivasan.** 1983. Nucleotide sequence of Abelson murine leukemia virus genome: structural similarity of its transforming gene product to other *onc* gene products with tyrosine-specific kinase activity. Proc. Natl. Acad. Sci. USA **80**:3623–3627.
18. **Sabatini, D. D., G. Kreibich, T. Morimoto, and M. Adesnik.** 1982. Mechanisms for the incorporation of proteins in membranes and organelles. J. Cell Biol. **92**:1–22.
19. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
20. **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.

21. **Shibuya, M., and H. Hanafusa.** 1982. Nucleotide sequence of Fujinami sarcoma virus: evolutionary relationship of its transforming gene with transforming genes of other sarcoma viruses. Cell **30**:787–795.
22. **Takahashi, M., J. Ritz, and G. M. Cooper.** 1985. Activation of a novel human transformng gene, *ret*, by DNA rearrangement. Cell **42**:581–588.
23. **Ullrich, A., J. R. Bell, E. Y. Chen, R. Herrera, L. M. Petruzzelli, T. J. Dull, A. Gray, L. Coussens, Y.-C. Liao, M. Tsubokawa, A. Mason, P. H. Seeburg, C. Grunfeld, O. M. Rosen, and J. Ramachandran.** 1985. Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature (London) **313**:756–761.
24. **Ullrich, A., L. Coussens, J. S. Hayflick, T. J. Dull, A. Gray, A. W. Tam, J. Lee, Y. Yarden, T. A. Libermann, J. Schlessinger, J. Downward, E. L. V. Mayes, N. Whittle, M. D. Waterfield, and P. H. Seeburg.** 1984. Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells. Nature (London) **309**:418–425.
25. **Ullrich, A., J. Shine, J. Chirgwin, R. Pictet, E. Tisher, W. J. Rutter, and H. M. Goodman.** 1977. Rat insulin genes: construction of plasmids containing the coding sequences. Science **196**:1313–1319.
26. **Wang, J. Y. J., F. Ledley, S. Goff, R. Lee, Y. Groner, and D. Baltimore.** 1984. The mouse c-*abl* locus: molecular cloning and characterization. Cell **36**:349–356.