Four additional members of the *ras* gene superfamily isolated by an oligonucleotide strategy: Molecular cloning of YPT-related cDNAs from a rat brain library

NICOLAS TOUCHOT, PIERRE CHARDIN, AND ARMAND TAVITIAN

Institut National de la Santé et de la Recherche Médicale U-248, Faculté de Médecine Lariboisière Saint-Louis 10, avenue de Verdun 75010 Paris, France

Communicated by Jean-Pierre Changeux, July 27, 1987

ABSTRACT Several oligonucleotide mixtures corresponding to a 6-amino acid sequence that is strictly conserved in all the ras and ras-related proteins (from various organisms) were tested for their ability to hybridize to 11 cloned members of the ras gene superfamily. Among these mixtures, a combination of two sets of partially complementary oligomers were able to hybridize to all the tested sequences. To identify members of the ras superfamily, we screened a rat brain cDNA library with these probes and isolated four genes, denoted rab1, -2, -3, and -4, encoding proteins homologous to the yeast YPT protein. Amino acid homology scores with YPT range from 75% for rabl to 37% for rab4, whereas the homologies with p21 ras and other ras-related proteins are $\approx 30\%$, and homologous residues were clustered in the regions involved in GTP/GDP binding. Another striking similarity shared by the rab and the other ras-related proteins is the conservation of at least one cysteine residue near the carboxyl-terminal end involved in the membrane binding of the ras proteins. rab1 is a mammalian homolog of the yeast YPT gene, and the four rab genes constitute an additional branch of the ras gene superfamily that to our knowledge has not been described in higher eukaryotes.

The c-Ha-ras (1) and c-Ki-ras (2) genes were first characterized as the cellular genes transduced in the Harvey and Kirsten sarcoma viruses. Later, these two genes as well as a third closely related one, N-ras (3), were found to be frequently activated to a transforming potential in mammalian tumors (for a review see ref. 4). Highly homologous genes have been isolated from a wide variety of organisms including Drosophila, Dictyostelium, and yeast (5-8). Two other genes, ral and R-ras, have been identified in mammals (9, 10), and the corresponding proteins share 50% homology with the transforming ras proteins. The ras gene family has also been extended by fortuitous discoveries such as the YPT gene, found as an open reading frame between the tubulin and actin genes of yeast (11), and the rho genes, first identified in an Aplysia cDNA isolated for other purposes and later characterized in human, rat, and yeast genomes (12, 13). Rho and YPT proteins share $\approx 30\%$ homology with the ras proteins.

In mammals, all the known *ras* or *ras*-related genes code for 21- to 24-kDa proteins that share structural and biochemical homologies with the guanine nucleotide-binding regulatory (G) proteins (14) involved in mediation of signal transduction in a variety of receptor/effector systems. ras proteins bind GTP and GDP and exhibit a low GTPase activity, as do the G proteins (15, 16). Four homology boxes corresponding to the GTP binding site are highly conserved in all the ras-related proteins. Among them, a stretch of six residues: Asp-Thr-Ala-Gly-Gln-Glu (in positions 57–62 of Ki-ras) is strictly conserved. A computer search in a protein databank* did not detect any other protein possessing this sequence even among the G proteins or other nucleotide binding proteins; thus, this sequence could be specific for ras-related proteins. According to the genetic code there are 512 distinct nucleotide sequences encoding Asp-Thr-Ala-Gly-Gln-Glu. A probe that could reveal all of them was required to search for additional members of the *ras* superfamily.

We describe here an improved oligonucleotide strategy based on the use of two partially complementary oligonucleotides designed to take advantage of weak G·T base pairings. This low-complexity oligonucleotide mixture hybridized to all the tested *ras* sequences and enabled us to isolate, from a rat brain library, four cDNAs encoding ras-related proteins that share extensive homologies with the yeast YPT protein. Nucleotide[†] and derived amino acid sequences of these four cDNAs are presented.

MATERIALS AND METHODS

Design and Synthesis of the Oligonucleotide Probes. Four 20-base oligodeoxyribonucleotide probes were synthesized corresponding to the sequence Asp-Thr-Ala-Gly-Gln-Glu-(Glu or Asp) (Fig. 1). The oligonucleotide oligo-512 is a mixture of all the possible icosamers that might encode this sequence. The oligonucleotides oligo-61 and oligo-61-COMPL were designed to take advantage of the weak G·T base pairings. Thus, oligo-61 (which hybridizes with the noncoding strand) possesses, in positions corresponding to the third bases of each codon, guanosine instead of guanosine or adenosine and thymidine instead of cytidine or thymidine. The oligo-61-COMPL was synthesized according to the same principle to hybridize with the coding strand. The oligo-In that hybridizes with the coding strand was synthesized by incorporating inosine (described as hybridizing weakly with the four naturally occurring bases) at the highly degenerated positions and a mixture of two deoxynucleotides at the other variable positions. Syntheses were performed manually on a silica gel solid phase, using a modification of the phosphoramidite method developed by Beaucage and Caruthers (17). The 20-mer oligonucleotide mixtures were purified from shorter sequences by electrophoresis on a 20% acrylamide/7 M urea gel and eluted. DNA synthesis kits were purchased from New England Biolabs.

Hybridization Conditions. Each of the oligonucleotide probes was 5'-end-labeled with T4 polynucleotide kinase and $[\gamma^{-32}P]$ ATP and tested on the following large sample of known

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. \$1734 solely to indicate this fact.

Abbreviation: G protein, guanine nucleotide-binding regulatory protein.

^{*}EMBL/GenBank Genetic Sequence Database (1986) EMBL Nucleotide Sequence Data Library (Eur. Mol. Biol. Lab., Heidelberg). [†]These sequences are being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J02998, J02999, J03000, and J03001).

Biochemistry: Touchot et al.

Conserved amino-acids sequence	5 A	7	P	5 T	8 h	r	5 A	9 1	•	6 G	0	y	6 G	1	n	6 G	2 1	u	6 G A	3	• u p
H - ras K- ras Dm - ras Dm - ras Y - RAS1 Y - RAS2 YPT rho 6 rho 12 Apl - rho	000000000000000000000000000000000000000	A AAAAAAAAA	TCTCTTCCCC		000000000000000000000000000000000000000	CAAGTTTGAA	0000000000000	000000000000000000000000000000000000000	CATTTAAGTG	0000000000000	0000000000000	CT A C A G T C G A	0000000000000	~~~~~~	GAAAGAGGA	000000000000000000000000000000000000000	~~~~~~	GGAGAAAGAG	0000000000000	A A A A A A G A A A	GGGGGATCCC
CLIGO 512 20-mer (mix of 512)	G	A	T C	<u>A</u>	c c	A C G T	G	c	T A C G T	G	G	G ACGT	c	<u>A</u>	G G	G	<u>A</u>	G A G	G	A 3' A	Ċ
OLIGO 61 20-mer G/T degenerated (mix of 8)	G	A	т	•	С	G T	G	С	G	G	G	G T	С	•	G	G		G	G	3	
OLIGO 61 COMPL 20-mer comple- mentaty sequence G/T degenerated (mix of 8)	30	г ; Т	G	T	G	G T	С	G	G	C	С	G T	G	т	т	С	т	т	с	5 T	
OLIGO IN 20-mer comple- mentary sequence inosine included (mix of 8)	30	r T	A G	т	G	1	С	G	1	c	С	I	G	т	C T	C	т	C T	c	5 T	

FIG. 1. (Upper) Alignment of the nucleotide sequences encoding Asp-Thr-Ala-Gly-Gln-Glu from the 11 ras or ras-related genes tested in this study. (Lower) Four 20-mer oligodeoxynucleotide mixtures. The amino acid sequence is indicated on the upper line. *, Positions where a single amino acid variation activates p21 ras to a transforming potential. Positions are numbered according to Ki-ras.

ras gene sequences: Ha-ras (18), Ki-ras (19), N-ras (20), Dm-ras1 (15), Y-RAS1 and Y-RAS2 (21), YPT (11), Aplysia rho, rho6, and rho12 (12), and ral (9) (see Fig. 1). An equimolecular quantity of each plasmid, linearized at a unique restriction site, was spotted on nitrocellulose. Filters were prehybridized for 4 hr at 65°C and hybridized for 12 hr at 45°C or 50°C in a solution containing $5 \times SSPE$ (1× SSPE = 3.6 M NaCl/200 mM NaH₂PO₄, pH 7.4/20 mM EDTA, pH 7.4), $5 \times$ Denhardt's solution (0.1% polyvinylpyrrolidone/ 0.1% Ficoll/0.1% bovine serum albumin), and 0.5% Nonidet P-40. Filters were washed with a solution containing 2× SSC (0.3 M NaCl/30 mM sodium citrate, pH 7.0) and 0.1% NaDodSO₄ twice for 30 min at 20°C and once for 30 min at each of the following temperatures: 45°C, 47°C, 50°C, or 52°C. Blots were then autoradiographed.

Screening of the cDNA Library and DNA Sequence Analysis. The λ gt10 rat brain cDNA library was kindly provided by D. Anderson (California Institute of Technology, Pasadena, CA). About 5×10^4 recombinant plaques were screened by plaque hybridization with 300,000 cpm per filter of an equimolar mixture of 5'-end-labeled oligo-61/oligo-61-COMPL. Prehybridization was for 4 hr at 65°C, and hybridization was for 12 hr at 45°C. Blots were washed twice for 30 min at 20°C and once for 30 min at 47°C in the buffers described above. Nucleotide sequences were determined by using phage M13 vectors and the dideoxy chain-termination method (22).

RESULTS

Probes, Assays, and Determination of Hybridization Conditions. The four probes separately plus the equimolecular combination oligo-61/oligo-61-COMPL were tested on 11 ras gene sequences under several hybridization and washing conditions. The results are shown in Fig. 2. Oligo-512 hybridized with every tested sequence, but a background was observed when a nonspecific sequence such as phage λ was



FIG. 2. Hybridization signals obtained with the five mixtures tested on 11 known *ras* sequences. The line T- represents phage λ DNA used as control. Stringency conditions were as follows. Columns: A, hybridization at 45°C and washes at 45°C; B, hybridization at 45°C and washes at 45°C; C, hybridization at 50°C and washes at 50°C; D, hybridization at 50°C and washes at 52°C.

used. Neither oligo-61 nor oligo-61-COMPL used alone hybridized with every tested sequence, but, as expected, all the sequences that gave a low signal with one of these two probes gave a high signal with the other. A mixture of oligo-61/oligo-61-COMPL used under the optimum stringency conditions (Fig. 2) hybridized with every sequence tested. The intensities of the signals were quite comparable to those obtained with the oligo-512; however, nonspecific hybridization with phage λ DNA as control was lower. It was observed that the signal intensities varied with the sequences. Oligo-In was expected to yield more constant intensity signals, and, actually, comparable intensities were observed for all the sequences that hybridized, but some sequences gave very low or even no signals.

To look for additional members of the *ras* superfamily, we decided to use the combination of oligo-61 and oligo-61-COMPL in equimolecular quantities under the stringency conditions described in Fig. 2 for columns B. This mixture successfully hybridized with λ gt10 ral clone and λ EMBL4 Ha-ras clone plaque (data not shown); this combination of probes, therefore, appeared to be appropriate for screening libraries.

Hybridization Study Using a Mixture of Two Partially Complementary Oligonucleotides. From our first results (Fig. 2), it appeared that for most of the tested sequences the signal intensity given by the combination oligo-61/oligo-61-COMPL was higher than the intensity obtained with each oligonucleotide alone. We tested the two probes separately and together on a ral clone, a sequence that mainly hybridized with oligo-61, on a Ki-ras clone, a sequence that mainly hybridized with oligo-61-COMPL, and on a Ha-ras clone, a sequence that hybridized with both. The hybridizations were performed under low-stringency conditions (42°C), and the washings were at increasing temperatures (from 42°C to 60°C). Melting curves were drawn (data not shown), and in each case the curve obtained with the mixture of the two complementary probes could be superimposed on the sum of the curves obtained with oligo-61 alone and oligo-61-COMPL alone. Thus, even though two partially complementary oligonucleotide probes were present in the same mixture during the hybridization process, there was no competition for hybridization with DNA fixed on nitrocellulose.

Screening of the cDNA Library from Rat Brain and Characterization of the Positive Clones. The mixture of the two 20-mer probes oligo-61/oligo-61-COMPL was used to screen a λ gt10 rat brain cDNA library. We screened 50,000 recombinant plaques and isolated 50 positive clones; 14 of these clones gave a high signal and were studied further. For each clone the *Eco*RI fragment carrying the hybridizing sequence was characterized by Southern transfer (23) before cloning in phage M13 vectors. The partial nucleotide sequences of the 14 clones were determined in the region hybridizing with the probes. Six clones contained a nucleotide sequence able to hybridize with the probes, but the corresponding reading frames either were interrupted by a nonsense codon or corresponded to amino acid sequences showing no homology with the ras proteins except in the target sequence. In these cases the homologies with the probes were partial, corresponding to 15 nucleotides at most (but were G+C rich in variable positions). Three clones corresponded to Ha-ras cDNAs and one corresponded to a Ki-ras cDNA. The four others contained an open reading frame. The derived amino acid sequence included the Asp-Thr-Ala-Gly-Gln-Glu sequence and protein sequence homologies with the yeast YPT protein. These were named rab1, -2, -3, and -4 (for ras genes from rat brain). Their nucleotide and deduced amino acid sequences are shown in Fig. 3. Assignment of the initiator AUG codons is based on DNA sequence analysis and comparison with YPT of the protein sequences encoded by the open reading frames. Although the conserved sequence features make these assignments very likely, this has yet to be formally demonstrated. rab1 possesses two ATG codons at its 5' end, both preceded by a eukaryotic ribosome binding site, thus the real ATG initiation site could not be determined. The same structure was also found in a human rab1 cDNA (data not shown). Preliminary results suggest that the rab genes are expressed in brain at higher levels than the rho or ral genes. This might explain why rho or ral cDNAs were not isolated by the screening we used.

Amino Acid Sequence Homologies Among the rab, ras, ral, rho, and YPT Proteins. The amino acid sequences of rab1, -2, -3, -4, Ki-ras, ral, rho, and YPT are aligned in Fig. 4. The percentage of amino acids strictly conserved between YPT and the rab proteins is 75% for rab1, 38% for rab2, 40% for

RDI: ATG TCC AGC ATG AAT CCC GAA TAT GAT TAT TTA TTC AAG TTA CTC CTG ATT GGC GAC TCT Met Ser Net Asn Pro Giu Tyr Asp Tyr Leu Phe Lys Leu Leu Leu iie Giy Asp Ser 9 GGG GTT GGA AAG TCT TGC CTT CTC CTT AGG TTT GCG GAT GAC ACG TAT ACG GAA AGC TAC Giy Val Giy Lys Ser Cys Leu Leu Leu Arg Phe Ala Asp Asp Thr Tyr Thr Giu Ser Tyr 21 ATC AGC ACG ATT GGT GTG GAT TTC AAG ATA CGG ACT ATA GAG CTA GAC GGG AAA ACA ATC 110 Sor Thr 110 Giy Val Asp Pho Lys 110 Arg Thr 110 Giu Lou Asp Giy Lys Thr 110 18 AAG CTT CAG ATA TGG GAC ACA GCA GGC CAG GAA CGG TTT CGA ACA ATC ACC TCC AGT TAT Lys Leu Gin lie Trp Asp Thr Ala Giy Gin Giu Arg Phe Arg Thr lie Thr Ser Ser Tyr 241 TAC AGA GGA GCC CAT GGA ATC ATA GTT GTG TAT GAT GTG ACA GAC CAG GAG TCC TTC AAT Tyr Arg Gly Ala His Gly lie lie Val Val Tyr Asp Val Thr Asp Gin Glu Ser Phe Asn 301 AAC GTG AAA CAG TGG CTG CAG GAG ATC GAT CGC TAC GCC AGC GAA AAT GTC AAC AAG TTG Asn Val Lys Gin Trp Lau Gin Giu ile Aap Arg Tyr Ala Ser Giu Asn Val Asn Lys Leu 391 TTG GTA GGG AAC AAA TGT GAC CTG ACC ACA AAG AAA GTA GTA GAC TAC ACA ACA GCC AAG Lou Val Giy Asn Lys Cys Asp Lou Thr Thr Lys Lys Val Val Asp Tyr Thr Thr Ala Lys 421 GAA TIT GCA GAT TCC CTT GGA ATT CCA TIT TTG GAA ACC AGT GCT AAG AAC GAA AAG AAT Glu Phe Ala Asp Ser Leu Gly lie Pro Phe Leu Glu Thr Ser Ala Lys Ash Glu Lys Ash 481 GTA GAA CAG TCT TTC ATG ACC ATG GCA GCG GAG ATT AAA AAG CGG ATG GGT CCT GGA GCA Val Glu Gin Ser Phe Net Thr Net Ala Ala Glu IIe Lys Lys Arg Net Giy Pro Giy Ala 541 ACA GCT GGA GGT GCG GAG AAG TCC AAT GTT AAA ATC CAG AGC ACT CCA GTC AAG CAG TCA Thr Ala Giy Giy Ala Giu Lys Ser Asn Val Lys IIe Gin Ser Thr Pro Val Lys Gin Ser BI GGT GGA GGC TGC TGC TAA GIY GIY GIY CYS CYS *** rab3: GAC TCT TTT ACT CCA GCC TTT GTC AGC ACT GTG GGC ATA GAC TTC AAG GTC AAA ACC ATC Asp Ser Phe Thr Pro Ala Phe Val Ser Thr Val Gly lie Asp Phe Lys Val Lys Thr lie 8 TAC CGC AAT GAC AAG AGG ATC AAG CTG CAG ATC TGG GAC ACA GCA GGA CAA GAG CGG TAC Tyr Arg Asn Asp Lys Arg lie Lys Leu Gin lie Trp Asp Thr Ala Giy Gin Giu Arg Tyr 21 CGA ACC ATC ACC ACA GCC TAC TAC CGG GGC GCC ATG GGC TTC ATT CTA ATG TAT GAC ATC Arg Thr lie Thr Thr Ala Tyr Tyr Arg Giy Ala Met Giy Phe lie Leu Met Tyr Asp lie 18 ACC AAT GAG GAG TCC TTC AAT GCA GTG CAG GAC TGG TCC ACT CAG ATC AAA ACT TAC TCA Thr Asn Glu Glu Ser Phe Asn Ala Val Gin Asp Trp Ser Thr Gin IIe Lys Thr Tyr Ser 241 TGG GAC AAT GCC CAG GTG CTG CTG GTG GGG AAC AAG TGC GAC ATG GAG GAC GAG CGA GTG Trp Asp Asn Ala Gin Val Leu Leu Val Giy Asn Lys Cys Asp Met Giu Asp Giu Arg Val 201 GTG TCC TCA GAA CGA GGC CGG CAG CTG GCC GAC CAC CTG GGC TTT GAG TTC TTT GAG GCC Val Ser Ser Glu Arg Gly Arg Gln Leu Ala Asp His Leu Gly Phe Glu Phe Phe Glu Ala

381 AGC GCC AAG GAC AAC ATT AAT GTC AAG CAG ACC TTT GAA CGT CTG GTG GAC GTG ATC TGT Ser Ala Lys Asp Asn IIe Asn Val Lys Gin Thr Phe Giu Arg Leu Val Asp Val IIe Cys

421 GAG AAG ATG TCG GAG TCC CTA GAT ACT GCA GAC CTT GCA GTC ACG GGT GCC AAG CAG GGC Giu Lys Mei Ser Giu Ser Leu Asp Thr Ala Asp Pro Ala Val Thr Giy Ala Lys Gin Giy

481 CCA CAG CTC ACC GAC CAG CAG GCA CCA CCT CAT CAG GAT TGC GCC TGC TGA Pro Gin Leu Thr Asp Gin Gin Ala Pro Pro His Gin Asp Cys Ala Cys *** rab3, and 37% for rab4. Four regions corresponding to residues 10-17, 57-63, 113-120, and 143-149 of Ki-ras and located in the GTP binding site are highly conserved in rab, YPT, ral, rho, and ras proteins. Amino acids 64-84 are highly conserved among the rab proteins and YPT and might be considered to be a distinctive characteristic of the rab family. The same region is also conserved among the ral and ras proteins. It is, however, different between these two groups. As in all the ras proteins a cysteine near the carboxyl-terminal end is found in the rab proteins. It should be noted that the last five carboxyl-terminal residues of the rab1 and rab2 proteins, Gly-Gly-Gly-Cys-Cys, are identical to those of YPT, whereas rab3 ends with Asp-Cys-Ala-Cys, and rab4 ends with Glu-Cys-Gly-Cys. These features clearly indicate that the rab genes belong to a third branch of the ras superfamily distinct from the ras protooncogenes and from the rho family.

DISCUSSION

Oligonucleotide probes are powerful tools for the isolation of cDNAs when at least a partial amino acid sequence of a protein is known (24, 25). Each protein poses a specific problem for the choice of a target region and the design of the oligonucleotide mixture. We developed an oligonucleotide strategy to search for additional members of the ras superfamily. Alignment of the ras-related proteins reveals a stretch of six amino acids Asp-Thr-Ala-Gly-Gln-Glu-(Glu or Asp) that appears to represent the best target region for two reasons. (i) It is found only in ras or ras-related proteins (and not in G proteins or in other nucleotide binding proteins). (ii) Half of the Asp-Thr-Ala-Gly-Gln-Glu residues are coded only by two triplets, and this reduces the number of possible coding sequences. Our purpose was to obtain a probe that could hybridize to every nucleotide sequence coding for this peptide. Three strategies were used: very complex oligonucleotide mixtures, G/T-degeneracy mixtures, and probes containing deoxyinosine (26). The use of a very complex

mb2 ATG GCG TAC GCC TAT CTC TTC AAG TAC ATC ATC ATC GGC GAC ACA GCG GTT GGT AAA TCG Nei Ala Tyr Ala Tyr Leu Phe Lys Tyr IIe IIe IIe Giy Aep Thr Giy Val Giy Lys Ser
G TGC TTA TTG CTA CAG TTT ACG GAC AAG AGG GTT CAG CCG GTG CAT GAC CTC ACA ATG GGT Cys Leu Leu Leu Gin Phe Thr Asp Lys Arg Phe Gin Pro Val His Aap Lsu Thr Mei Giy
G TAG GAT TTT GG TC CGA ATG ATA ACC ATT GAT GGG AAA CGG GA ACA CGA TA AAA CTC CAG ATC GAG Val Giu Phe Giy Ala Arg Mei IIe Thr IIe Aap Giy Lys Gin IIe Lys Leu Gin IIe Trp
G GAT CAA GCA GGG GAG GAG TCC TIT GGT TCT ATC ACA AGG TCA TAT TAC CAA GGT GCA GCG Asp Thr Ala Giy Gin Giu Ser Phe Arg Ser IIe Thr Arg Ser Tyr Tyr Arg Giy Ala Ala
21 GGG GCT TA GTG TAT GAT ATT ACA AGG AGA GAC GG TC ACC ATT GAC ACC TG Giy Ala Leu Leu Val Tyr Asp IIe Thr Arg Arg Asp Thr Phe Aan His Lau Thr Thr Trp
21 TTA GAA GAC GCC CGT CAG CAT CC AAT TCC AAC ATG GTC ATC GC ACA CTG GG GA ACT CAG Giu Ala Diu Asp Ala Arg Gin His Ser Aan Mat Val IIe Mat Leu IIe Giy Aan Lys Ser Asp Leu Giu Ser Arg Arg Gin Val Lys Liu Giu Giu Giu Ala Phe Ala Arg GA Leu Giu Asp Ala Arg Gin His Ser Aan Mat Val IIe Mat Leu IIe Giy Aan Lys
231 AGT GAC TTA GAA TCT AGG GAG GAA GTA GAA AAG GAA GAT GAA GCT GAC CTT GC CAT CATC ATG CAG CAG CTT His Giy Leu Giu Ser Arg Arg Giu Val Lys Lys Giu Giu Giy Giu Ala Phe Ala Arg Giu Giu Alas Phe Mat Giu Thr Ser Ala CC AAG AGA GGT GCT TT ATT GGA AAT AAA GA ATT AAC ACA GAA AAT AAA GAA ATT TAT GAA AAA AGG AAA GGG GCT TIT GAC ATT AAT AT Hie Aan Thr Ala Lys Giu Uli Thr Ser Ala Lys Thr Ala Ser Aan Val Giu Giu Ala Phe Ala Arg Giu
241 CAT GGA CTT ACT AAA GAA ATT ATG GAC AAA TC CAA GGG GCT TIT GAC ATT AAT AAT Hie Aan Thr Ala Lys Giu Lie Tyr Giu Lys Hie Gin Giu Giy Val Phe Asp Hie Aan Asn Giu Ala Asn Giy Hie Lys Hie Giy Pro Gin Hie Ala An An Ala Asn GAA CT ACA ASN GGG CATT AT AA Asn Gin Giy Giy Gin Gin Ala Giy Giy Giy Giy Cys Cys ****

#34: GAA TTC GGT CAA AAG ATA ATA ATA GAT GGT GGT AAA TAT GTG AAG TTA CAG ATA TGG GAC GIU Phe GIY GIN LY II & II & AI A ATA GTY GIY GIY LY TY VAI LYS Leu GIN II & TTP ASD GIA CGG GGT GGA CAG GAG GGG GTC AGG TCT AGG AGC AGG TAC TAC AGA GGT GGG GCT GGG Thr AI a GIY GIN GIU ATY Phe Arg Ser Vai Thr Thr Ser Tyr Tyr Arg GIY AI a AI a GIY
© GCA CTC CTC GTC TAT GAC ATC ACC AGC GGA GAA ACC TAC AAT GGC TT ACT AAT TGG TTA AI a Leu Leu Vai Tyr Asp 11 = Thr Ser Arg GIU Thr Tyr Asn AI a Lau Thr Asn Trp Leu
© ACA GAT GCC AGA ATG GTG GCA AGC CAG GGA AACC TAC AAT GGC TT ACT AAT TGG TTA AI a Leu Leu Vai Tyr Asp 11 = Thr Ser Arg GIU Thr Tyr Asn Ai a Lau Thr Asn Trp Leu
© ACA GAT GCC AGA ATG GTG GCA GCC AGG GAG ACTC GTC ATC ATT GCT CT GG GGA AGA CAG AAG Thr Asp Ai a Arg Mei Leu Ai a Ser Gin Asn 11 = Vai 11 = 11 = Leu Cys Giy Asn Lys Lys
© GAC CTG GAT GCC GAC CGG GAA GTC ACC TTC CTT GAA GCC TCC AAG TTC GCA CAA AGG AAT ASp Leu Asp Ai a Atp Arg Giu Vai Thr Phe Leu Giu Ais Ser Arg Phe Aia Gin Giu Asn
© GAG CTC ATG TTC CTG GAA ACC AGT GCA CTG ACT GCG CAG AAC GC GC GAA GAG GAG GTU Leu Mei Phe Leu Giu Thr GIY Giu Asn Vai Giu Giu Ai a Phe Mai Giu Cys Aia Arg Lys Lys
© GAG CTC GGT AGC AGG AAG ATT GAT CA CGT GAC GAG CGC GGA GAG GTG GAC CCC GAG AGG ATG GIN Cys Aia Arg Lys Lys Lys Lu Leu Asp Aia ALC AGT GCA GCC GCC TTG AGA CAG CTG ACC CG AGA AGG ATG GIN Cys Aia Arg Lys Lys Li Leu Ash Lys Li & Giu Ser Giy Giu Leu Asp Pro Giu Arg Wai Gin Cys Aia Arg Lys Lys Lys Ash Lau Lys Lys Giu Su Cit Gy Ser Pro Arg Arg Thr
© GAG CTC CAA GAT GCA CAG GAG CAC CGC CTG AGA CAG CTA CGC CGC AGA CGC GCA ACA GIY Ser Giy 11 = Gin Tyr Giy Asp Aia Aia Lau Arg Gin Lu Arg Ser Pro Arg Arg Thr
© GAG CTC CAA GAT GCA CAG GAG GAT GTG GCC TCC TAG GIN CAA GAT GCT CCC CAA GAG GAT GG GIN Leu Arg Ser Pro Arg Arg Thr
© GAG CTC CAA GAT GGT ACA CAG GAG GTG GCA CTC TAG GIN ALC CAG AGT CCA AGT CCA CAG AGG GAT GGC TCC TAG GIN ALL GUN GIN TH GIN GIN TYR GIY Asp Aia Aia Lau Arg Gin Lu Arg Ser Pro Arg

FIG. 3. Nucleotide and predicted amino acid sequences of the rab1, -2, -3, and -4 open reading frames. Numbers indicate the position of nucleotides starting at the initiation codon (rab1 and rab2) or at the first nucleotide in the open reading frame (rab3 and rab4).

Biochemistry: Touchot et al.

ral K-ras tho YPT rab1 rab2 rab3 rab4		M M M ·	A . A . S	A . A . S	N	K	P.K.	K	G · . N N ·	0 - SP M 	N · · EEA · ·	S • Y Y Y ·		A 1 T 9 9 9 1		HY.FFF.						GGGGGG · ·	S A D N D D	G G S S S T	GGAGG .		GGGGG	****	S . S . S .										E F F G Y Y F F F		EDENEPP.		YYYYHF	EDV11DV.	P P S S L S		K V 0 0 0 0 0 0 0 0 0	AEFGGGG		S N Q D E D E	YYFFFF		K (D (R) R) K)					GGGGGGDG
ral K-ras tho YPT rab1 rab2 rab3 rab4		ID E E K K K K K K	ETOTTORY		OLEKKKK				L L & & & & & & & & & & & & & & & & & &		T T T T T T T T	~ ~ ~ ~ ~ ~ ~ ~	000000000	00000000	EEEEEE	DE D A A S A A	YYFFFYF		AATTSTS		A A A T T T T T	69 D D P S S R T T	• • • • • •		Y Y Y Y Y Y Y Y Y Y										F F F Y Y Y Y	SASDDDDD			ENSDDRNS	MTPOORER	88 EKDEEDEE	S S S S S T S T	FFLFFF												K K H A A S S A	EDFTSNWS	DSCSESDQ	E E P
rai K-ras rho YPT rab1 rab2 rab3 rab4	1: # [] 			P P L N V Q V	FMIKKIVI	1 2 1 1 1 1 1		* * * * * • C	66666666		******	SCKCCSCK	0000000	L L L L L M L	EPRKTEED		K K K R E D	AR.RKRR	Q T D V V E V E	V V E V V V V V	SDSEDKST	12 T T Y K S F	7 EKKDTEEL									NGPKGGGE	V I V M I L F L		Y F Y F F F F F F	VILLMFL		TTCTTT AT	9.5 9 9 9 9 9 9	~ ~ ~ ~ ~ ~ ~ ~	14 K K K L K K K L							VAVASATA	FFFFF					EE - QEEVK		RR KKYCL	AK.QKEEN	RH·SRKKK
ral K-ras rho YPT rab1 rab2 rab3 rab4	18 K K M M I M I N	57 N 5 1 5 1 5 1 5 1 5 1 5 1 5 1 5 1 5 1 5			D M - 0 G G S G	55.N.VLE	KKAL-FDL	EDAN DTD	K - L - I A		GGVTTNPR	KKKQAEAM	K K K K G A V G	KKKK GNTS	RKKEAGGG	KKDEIAI	KK - KK KKQ	S . G S Q Y		A T - V V P P D	KK - NKOOA	a			N N - L	A - R	· · · · · S · S	H . P	- - - - - - - - - - - - - - - - - - -	- - - - - - - - - - - - - - - - - - -	N . T		· · · SSGAA	L T G P P	- TPOPS	· · · · · · · · · · · · · · · · · · ·	к	NOOH A		G G G D E					L	I												

FIG. 4. Alignment of the four rab protein sequences with the human Ki-ras (exon 4B), *Aplysia* rho, and yeast YPT proteins. Dashes indicate gaps introduced for optimal alignment. Sets of identical or conservative residues are indicated by dots. Conservative amino acid substitutions are grouped as follows: cysteine, serine, threonine, proline, alanine, and glycine; asparagine, aspartic acid, glutamic acid, and glutamine; histidine, arginine, and lysine; methionine, isoleucine, leucine, and valine; phenylalanine, tyrosine, and tryptophan. The reference numbering is that of Ki-ras.

mixture corresponding to all the possibilities offered by the genetic code ensures that every nucleotide sequence corresponding to a given protein sequence will be identified; however, a high-complexity probe often leads to nonspecific hybridizations and increases the proportion of irrelevant sequences isolated. A less-complex probe constituted by a mixture of two partially complementary probes containing guanosine, thymidine, or guanosine and thymidine at the third base of each codon yielded the best results in our hands. This approach was tested only on 11 known sequences among the 512 possible ones, although we cannot be sure that all of them could be identified by hybridization. Nonetheless, the most unfavorable conditions were encountered when, for both oligo-61 and oligo-61-COMPL, three out of the six wobble bases hybridized strongly (i.e., G·C or A·T) and three of the six hybridized weakly (i.e., G·T). This was the case for several of the tested sequences, and two of them represented very different situations. (i) For Ha-ras the three low basepairing positions with oligo-61 were placed at a central position, but with oligo-61-COMPL they were located at the two ends. (ii) For Y-RAS1 the three low base-pairing positions were placed at one end with oligo-61 and at the other end with oligo-61-COMPL. Both probes gave positive hybridization signals with these two sequences under the stringency conditions used for screening. Thus, the mixture of oligo-61/oligo-61-COMPL might reveal most of the nucleotide sequences coding for the peptide sequence. Perhaps unknown members of the ras superfamily present some variability in the region corresponding to amino acids 57-63, so that the corresponding clones would not be isolated by the methodology we developed. In fact, 6 clones among the 14 clones we isolated had a few mismatches with the oligonucleotide probes (due to the low-stringency conditions) but none of these clones was ras-related. Thus, a strategy similar to the one we developed for the ras superfamily should be adaptable to search for more members of other gene families.

We isolated four cDNAs coding for the proteins rab1, -2, -3, and -4 that share, respectively, 75%, 38%, 40%, and 37% homology with YPT and around 30% homology with the ras

proteins and the other described ras-related proteins. rab1 and rab2 cDNAs contain full-length coding regions. The rab1 open reading frame encodes a protein of 202 or 205 amino acids (depending on which ATG codon is effective for initiation) with expected molecular weights of 22,500 or 22,800. The predicted rab2 protein is 212 amino acids long with an expected molecular weight of 23,400. Three-dimensional models for p21 ras have been proposed (27, 28). We shall refer to these models for a discussion of the structural homologies between the rab and ras proteins. The homologies are localized in four main regions corresponding to amino acid residues 5–21, 53–62, 112–120, and 141–147 (using the numbering system for Ki-ras) that have been implicated in the binding of GTP/GDP to p21 ras. Two other regions involved in GTP binding, amino acid residues 75-84 (29) and 152-164 (30) are less homologous between the rab and ras proteins, but many of the amino acid changes are conservative. The distances between these six boxes of homology as well as the hydrophobicity curves (data not shown) are almost identical for p21 ras, YPT, and rab proteins. Besides these six regions only a few amino acids are strictly conserved such as Thr-35, Tyr-71, and Cys-192. Thr-35 and Tyr-71 are probably important to maintain structural or functional properties of the ras-related proteins. The role of Cys-192 is discussed below. Amino acids 29-49 that have been shown to be important for the transforming potential of p21 ras (31) are not very conserved between the rab and ras proteins or among the four rab proteins themselves. This region probably corresponds to a large flexible loop interacting with the effector (28, 31). Amino acids 63-75 are highly conserved in the four rab proteins and are very different from the corresponding region of Ki-ras needed for the binding of monoclonal antibody Y13-259 that blocks p21 function (32). No precise role has been assigned to this region, but it might interact with regulatory proteins. These interacting proteins should be different between rab and ras but may be identical or closely related for all the rab proteins.

The most striking differences between all the members of the *ras* superfamily are located in the carboxyl-terminal region from amino acid residue 166 to the end. However, a cysteine residue located four amino acids before the carboxyl-terminal end of Ha-ras and involved in membrane anchoring (33) is found in all the ras proteins. Deletions and point mutations in v-Ha-ras gene (34) suggest that several carboxyl-terminal aliphatic residues in addition to this cysteine might also be important for biological activity. Such a structure with a cysteine followed by aliphatic residues is found in the ras, ral, and rho proteins, but in YPT and the rab proteins the cysteine is the last carboxyl-terminal residue and one other cysteine is found immediately before (YPT, rab1, and rab2) or with one intervening amino acid (rab3 and rab4). rab1 and rab2 possess exactly the same five carboxylterminal residues as YPT. The conservation of this sequence from yeast to humans suggests an important functional role for this region. Several residues upstream from the cysteine, the ras, ral, and rho proteins share a sequence of basic amino acids that is not found in the rab proteins. These variations in the carboxyl-terminal region of the rab and ras proteins might correspond to different sublocalizations in the cellular membrane systems. The precise position of the rab proteins and the nature of the interacting proteins remain to be characterized.

rabl was first identified in a rat cDNA, but we also isolated a human rab1 cDNA and the encoded protein is identical to the rat protein in the sequenced region (data not shown). As for Y-RAS1 and Y-RAS2, which are the yeast counterparts of the classical ras genes, the overall homology between YPT and rab1 proteins is around 70%, and two domains can be discriminated. The first one corresponding to amino acid residues 1-104 shows 95% homology between YPT and rab1, whereas the second one from amino acid residues 104-167 shows around 60% homology. Downstream from amino acid residue 167, rab1 and YPT diverge radically except for the last five carboxyl-terminal amino acids Gly-Gly-Gly-Cys-Cvs. rab2, -3, and -4 share $\approx 40\%$ homology with rab1 and thus are in the same situation compared to rab1 as ral or R-ras compared to Ha-ras, Ki-ras, and N-ras. rab2, -3, and -4 proteins share $\approx 40\%$ homology with YPT, and the presence of their counterparts in yeast remains an open question.

The identification of an increasing number of ras-related proteins raises one question: What specific function might be exerted by each of them? It has been reported (35) that the YPT protein might be involved in microtubule organization and is required for cell viability, but the functional relationships existing between the vertebrate and the yeast rasrelated proteins remain to be defined. In mammals we may speculate that, as with the G proteins, the ras proteins are involved in mediation of signal transduction between membrane receptors and cytoplasmic effectors. The ability to transduce the signal is probably regulated by GTP binding and hydrolysis-essential properties determined by internal highly conserved regions-whereas some external regions are specific in each protein. Thus, the different proteins of the ras superfamily may share common properties that are used in various biological pathways by interacting specifically with various receptors, effectors, or regulatory proteins. We postulate that the presence of a large number of different proteins of the ras superfamily, at least in some cell types such as neural cells, might reflect a role in the integration of many different or even in some cases antagonist signals.

We thank D. Anderson for the gift of his excellent cDNA library and C. Michon for the synthesis of the first oligo-512. We are indebted to Drs. J. Capon, R. Ellis, D. Gallwitz, A. Hall, P. Madaule, M. McCoy, B. Mozer, S. Powers, and E. Taparowsky for the various ras or ras-related clones. N.T. and P.C. are recipients of fellowships from the Ministère de la Recherche et de l'Enseignement Supérieur. This work was supported by the Institut National de la Santé et de la Recherche Médicale.

- 1. De Feo, D., Gonda, M., Young, H., Chang, E., Lowy, D., Scolnick, E. & Ellis, R. (1981) Proc. Natl. Acad. Sci. USA 78, 3328-3332.
- Ellis, R., De Feo, D., Shih, T., Gonda, M., Young, H., 2. Tsuchida, N., Lowy, D. & Scolnick, E. (1981) Nature (London) 292, 506-511.
- Shimizu, K., Goldfarb, M., Perucho, M. & Wigler, M. (1983) 3. Proc. Natl. Acad. Sci. USA 80, 383-387.
- Barbacid, M. (1987) Annu. Rev. Biochem. 56, 779-827.
- 5. Mozer, B., Marlor, R., Parkhurst, S. & Corces, V. (1985) Mol. Cell. Biol. 5, 885-889.
- Reymond, C., Gomer, R., Mehdy, M. & Firtel, R. (1984) Cell 6. 39, 141-148.
- De Feo-Jones, D., Scolnick, E., Koller, E. & Dhar, R. (1983) 7. Nature (London) **306**, 707–709. Fukui, Y. & Kaziro, Y. (1985) *EMBO J.* **4**, 687–691.
- 8.
- Chardin, P. & Tavitian, A. (1986) EMBO J. 5, 2203-2208.
- 10. Lowe, D., Capon, D., Delwart, E., Sakaguchi, A., Naylor, S. & Goeddel, D. (1987) Cell 48, 137-146.
- 11. Gallwitz, D., Donath, C. & Sander, C. (1983) Nature (London) 306, 704-707.
- 12 Madaule, P. & Axel, R. (1985) Cell 41, 31-40.
- Madaule, P., Axel, R. & Myers, A. (1987) Proc. Natl. Acad. Sci. USA 84, 779-783. 13.
- Gilman, A. (1984) Cell 36, 577-579. 14.
- Papageorge, A., Lowy, D. & Scolnick, E. M. (1982) J. Virol. 15. 44. 509-519.
- 16. McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. (1984) Nature (London) 310, 644-649.
- 17. Beaucage, S. & Caruthers, M. (1981) Tetrahedron Lett. 22, 1859-1862.
- 18. Fasano, O., Taparowsky, E., Iddes, J., Wigler, M. & Goldfarb, M. (1983) J. Mol. Appl. Genet. 2, 173-186.
- 19 McCoy, M., Bargmann, C. & Weinberg, R. (1984) Mol. Cell. Biol. 4, 1577-1582.
- 20. Taparowsky, E., Shimizu, K., Goldfarb, M. & Wigler, M. (1983) Cell 34, 581-586.
- Powers, S., Kataoka, T., Fasano, O., Goldfarb, M., Strathern, 21. J., Broach, J. & Wigler, M. (1984) Cell 36, 607-612.
- 22. Messing, J. (1983) Methods Enzymol. 101, 20-78.
- Southern, E. (1975) J. Mol. Biol. 98, 503-517. 23.
- 24. Itoh, H., Kozasa, T., Shigekazu, N., Nakamura, S., Katada, T., Michio, U., Shigenori, I., Ohtsuka, E., Kawasaki, H., Suzuki, K. & Kaziro, Y. (1986) Proc. Natl. Acad. Sci. USA 83, 3776-3780.
- Hanks, S. (1987) Proc. Natl. Acad. Sci. USA 84, 388-392. 25
- Takahashi, Y., Kato, K., Hayashizaki, Y., Wakabayashi, T., Ohtsuka, E., Matsuki, S., Ikehara, M. & Matsubara, K. (1985) 26. Proc. Natl. Acad. Sci. USA 82, 1931-1935.
- Jurnak, F. (1985) Science 230, 32-36. 27
- McCormick, F., Clark, B. F. C., La Cour, T. F. M., Kjeld-gaard, M., Norskov-Lauritsen, C. & Nyborg, J. (1985) Science 28. 230, 78-82.
- 29. Feig, L., Pan, B., Roberts, T. & Cooper, G. (1986) Proc. Natl. Acad. Sci. USA 83, 4607-4611.
- 30. Lacal, J., Anderson, P. & Aaronson, S. (1986) EMBO J. 5, 679-687.
- Willumsen, B. M., Papageorge, A. G., Kung, H. F., Bekesi, E., Robins, F., Johnsen, M., Vass, W. C. & Lowy, D. R. 31. (1986) Mol. Cell. Biol. 6, 2646-2654.
- Sigal, I., Gibbs, J., D'Alonzo, J. & Scolnick, E. (1986) Proc. 32. Natl. Acad. Sci. USA 83, 4725-4729.
- Fujiyama, A. & Tamanoi, F. (1986) Proc. Natl. Acad. Sci. 33. USA 83, 1266-1270.
- 34. Willumsen, B., Norris, K., Papageorge, A., Hubbert, N. & Lowy, D. (1984) EMBO J. 3, 2581-2585.
- Schmitt, A., Wagner, P., Pfaff, E. & Gallwitz, D. (1986) Cell 35. **47,** 401–412.