Biochem. J. (1993) 293, 157–163 (Printed in Great Britain)

Structure of the murine *rab3A* gene: correlation of genomic organization with antibody epitopes

Marion BAUMERT,* Gabriele FISCHER VON MOLLARD,† Reinhard JAHN† and Thomas C. SÜDHOF*‡

*Department of Molecular Genetics and Howard Hughes Medical Institute, University of Texas Southwestern Medical School, Dallas, TX 75235, and †Department of Pharmacology and Howard Hughes Medical Institute, Yale University, New Haven, CT 06510, U.S.A.

Rab3A is a neuronal low-molecular-mass GTP-binding protein that is modified post-translationally by two geranylgeranyl groups and specifically targeted to synaptic vesicles. We have now cloned and characterized the murine gene coding for rab3A. With a size of less than 8 kb including the promoter, the *rab3A* gene is relatively small. It contains five exons, the first of which is non-coding. The organization of the rab3A coding sequence into exons in the gene is different from that of ras proteins, the only other low-molecular-mass GTP-binding proteins with currently characterized gene structures. Nevertheless, the intron placement in the primary structure of rab3A may be indicative of a domain division of the protein, since each coding exon contains one of the four major conserved rab protein sequence motifs. The epitopes of monoclonal and polyclonal antibodies to rab3A were mapped with the hypothesis that antibody epitopes might represent distinct exposed protein domains and correlate with exon structures. Two monoclonal antibodies, named 42.1 and 42.2, were found to recognize epitopes with a different degree of conservation between different rab3 isoforms. These epitopes were mapped to relatively short amino acid sequences corresponding to exons 4 and 5 respectively, whereas a polyclonal antibody recognized a complex epitope that required the presence of intact rab3A. Comparison of the sequence of rab3A with that of ras, whose crystal structure has been determined, revealed that the epitopes for the monoclonal antibodies correspond to regions in ras that are highly exposed. Taken together, these results suggest that exons 4 and 5 at least represent distinct exposed protein domains that also form major natural epitopes in rab3A.

INTRODUCTION

A large number of different low-molecular-mass GTP-binding proteins have been characterized in recent years. Many of them, primarily those belonging to the rab family, are present in membrane-bound and soluble forms, with the membrane-bound forms being localized to specific subcellular compartments. The rab family of low-molecular-mass GTP-binding proteins contains more than 20 members that are often present in several differentially distributed isoforms (Chavrier et al., 1990). The members of this family of low-molecular-mass GTP-binding proteins are thought to play a role in the regulation of intracellular membrane traffic (Balch, 1990; Hall, 1990; Sudhof and Jahn, 1991; Pfeffer, 1992). A model has been proposed for the function of these proteins suggesting that they cycle between membrane-bound and soluble forms and that they are targeted specifically to the intracellular compartments whose traffic they regulate (Bourne, 1988).

Rab3A is a small neuronal GTP-binding protein localized to synaptic vesicles (Fischer von Mollard et al., 1990). In synaptosomes, rab3A dissociates from synaptic vesicles after exocytosis and re-associates with them during recovery, suggesting that its function is related to the exocytosis or endocytotic recycling of synaptic vesicles (Fischer von Mollard et al., 1991). This finding explains the presence of both membrane-bound and soluble forms of rab3A. The primary structure of rab3A does not contain a transmembrane region to account for its tight membrane attachment. Instead, rab3A is isoprenylated at its two C-terminal cysteine residues by geranylgeranyl moieties. This post-translational modification is very hydrophobic, thereby attaching rab3A to the synaptic vesicle membrane (Farnsworth et al., 1991; Johnston et al., 1991; Khosravi-Far et al., 1991). The isoprenylation of rab3A is probably stable and is not removed in soluble rab3A. Since isoprenylated rab3A would be too hydrophobic to be soluble, there must be a mechanism to neutralize or cover the isoprenyl groups to allow the dissociation of rab3A from the synaptic vesicle membrane after exocytosis. Although the details of this process are not understood, it is likely that this is achieved by a protein called GDI (Sasaki et al., 1990), which specifically interacts with membrane-bound rab3A.

At least four isotypes of rab3 are known that appear to be highly homologous to each other (rab3A, rab3B, rab3C and rab3D; Touchot et al., 1987; Matsui et al., 1988; Baldini et al., 1992). Rab3A is primarily expressed in brain (Olofsson et al., 1988; Sano et al., 1989), whereas rab3D appears to be specific for insulin-stimulated transport vesicles for glucose transport in adipose cells (Baldini et al., 1992). Very little is known about the distribution of rab3B and rab3C. In addition, contradictory evidence on the presence of rab3A in non-neuronal and non-endocrine tissues has been presented (e.g. the exocrine pancreas; see Mizoguchi et al., 1989; Matsuda et al., 1992; Schnefel et al., 1992). It seems likely that additional isoforms of rab3 will be identified in future that may play a role in other forms of regulated vesicular traffic. It is evident that the characterization of these isoforms as well as the evaluation of the tissue distribution of rab3s will rely heavily on antibodies with precisely defined epitopes and isoform specificities.

To aid in the further characterization of rab3A, we have now studied the structure of the *rab3A* gene and the immunological characteristics of rab3A. The gene for rab3A was cloned and sequenced in order to elucidate the organization of the coding sequence into exons. In addition, the epitopes recognized by two

‡ To whom correspondence should be addressed.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under accession no. X72966.

monoclonal antibodies and one polyclonal antibody to rab3A were mapped. To date, the *ras* genes are the only genes encoding low-molecular-mass GTP-binding proteins whose intron–exon structure is known (McGrath et al., 1983; Taparowski et al., 1983; McCaffery et al., 1989). It was therefore of interest to determine if the gene structure of *rab3A* corresponds to that of *ras* and if individual exons correspond to antigenic epitopes, thus defining structural domains on the protein surface.

MATERIALS AND METHODS

Cloning, mapping and sequencing of the murine *rab3A* gene

A murine cosmid library (obtained from Stratagene) was screened with uniformly ^{32}P-labelled DNA probes derived from the bovine rab3A cDNA as described (Südhof, 1990). A single strongly hybridizing clone was isolated, and hybridizing DNA fragments were subcloned into pBluescript and analysed by Southern blotting and DNA sequencing. DNA sequencing was performed on M13 subclones of the cosmid clone by the dideoxy chain termination method (Sanger et al., 1977) using fluorescently labelled sequencing primers and an ABI 373A automatic DNA sequencer. Most regions of the sequence were obtained on both strands from multiple clones. The exon–intron structure of the gene was determined by comparison of the genomic sequence with the cDNA sequences from rat, bovine, human and *Drosophila* (Matsui et al., 1988; Touchot et al., 1987; Zahraoui et al., 1988, 1989; Johnston et al., 1991). All sequence analyses were performed using the Beckman Microgenie program; databank searches were performed on release 70 of GenBank.

Construction of plasmids

The cloning of the bovine *rab3A* and *3B* and human *rab3A* genes, and the construction of eukaryotic and prokaryotic expression vectors with these cDNAs, has been described previously (Fischer von Mollard et al., 1990; Matteoli et al., 1991). Two types of rab3A mutants were constructed: vectors for the expression of mutant rab3A proteins in eukaryotic cells utilizing cytomegalovirus promoter-based vectors, and vectors for the expression of mutant rab3A proteins in bacteria using pET vectors (Studier et al., 1991). These two expression systems were chosen in order to utilize both the ability of eukaryotic cells to isoprenylate transfected rab3A constructs (Johnston et al., 1991) and the greater ease of protein expression in bacteria. The different constructs used (see Figure 3) were constructed using standard procedures (Sambrook et al., 1989) and contained only the amino acids indicated in Figure 3, except for constructs C, G and H, which contained an additional leucine residue following amino acid 151 before the termination codon. All residue numbers correspond to those of Matsui et al. (1988).

Transfections, bacterial expression and immunoblot analysis

COS cells were cultured and transfected as described (Mignery et al., 1990). Bacterial expression was performed using pET vectors as described (Studier et al., 1991). Tissue extracts were prepared

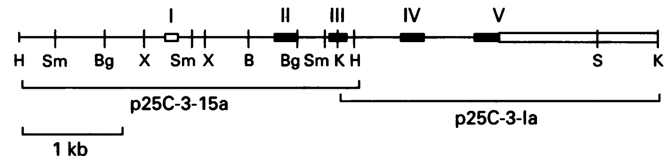

Figure 1 Diagram of the structure of the murine *rab3A* gene and of the cosmid subclones used for sequencing

The diagram of the *rab3A* gene structure (top) demonstrates the location of exons, shown by open boxes for non-coding sequences and filled boxes for coding sequences. The exon number is shown above the respective box. The location of salient restriction sites is marked by vertical lines and the sites are labelled as follows: B, *Bam*HI; Bg, *Bgl*II; H, *Hin*dIII; K, *Kpn*I; S, *Sac*I; Sm, *Sma*I; X, *Xho*I. The names and locations of the two cosmid subclones that were fully sequenced are shown below the gene structure. The scale is indicated by the bar.

from rats by homogenization and quantified using the Bio-Rad protein assay. SDS/PAGE was performed according to Laemmli (1970) and immunoblotting was carried out according to Towbin et al. (1979) using horseradish peroxidase-coupled secondary antibodies. Monoclonal antibodies were raised against recombinant rab3A produced in bacteria as described (Jahn et al., 1985; Matteoli et al., 1991). The polyclonal antibodies raised against rab3A were described previously (Fischer von Mollard et al., 1990; Johnston et al., 1991). Proteolysis of recombinant rab3A was performed by adding trace amounts of trypsin prior to SDS/PAGE.

RESULTS

Structure of the murine *rab3A* gene

A murine genomic cosmid library was screened with a uniformly ^{32}P-labelled DNA probe from the bovine rab3A cDNA. One positive clone was isolated, subcloned into plasmid vectors and analysed by sequencing. The structure of the gene revealed by the mapping and sequencing studies is shown in Figure 1, together with the relative localization of the subclones and the exons. Figure 2 depicts the complete sequence of the gene from the 5′ flanking region, which presumably contains the rab3A promoter, into the 3′ untranslated region.

The localization of the exons in the *rab3A* gene was determined by comparison with the rat, mouse and human cDNA sequences (Zahraoui et al., 1988, 1989; Matsui et al., 1988). This comparison revealed that the gene contains five exons. The first exon is non-coding. The second exon starts with the initiator ATG as the first three nucleotides (Figure 2). Although the exact transcription start site for the *rab3A* gene was not mapped, the 5′ untranslated sequence of the rat rab3A cDNA is found to be completely conserved in murine exon 1, and its beginning would represent a good transcription initiation site in the *rab3A* gene (arrow in Figure 2; Smale et al., 1990). Interestingly, four copies of a six-nucleotide repeat (sequence GTCGCC; shaded in Figure 2 and labelled AA–DD) were found in the 5′ untranslated region of *rab3A*. This repeat is conserved between the mouse and human sequences (Zahraoui et al., 1989). In addition, two copies of an

Figure 2 Nucleotide and translated amino acid sequences of the murine *rab3A* gene

The location and numbers of the exons are shown on the left of the sequence, and the sequences are numbered on the right. Exon sequences, as determined by comparison with the cDNA sequences (Matsui et al., 1988; Zahraoui et al., 1988, 1989), are shown in italic type, and the amino acid sequence is given below the nucleotide sequence in single-letter code (also in italics). Intron donor and acceptor consensus nucleotide sequences are underlined, and the location of the translation stop codon is shown by an asterisk at position 4779. The arrow above nucleotide 1491 marks the 5′ end of the rat cDNA sequence, which may represent the transcription start site. The two classes of direct repeats surrounding the putative transcription start site are marked by shading and labelled A and B or AA, BB, CC and DD.

```
    AAGCTTCCTCCAGCGGCCCCGGATGGGTCGGGGCTCCAGGCTCCTCGCGGGGGTGGACGAGCATCCAGGCGACAAGTTAT     80
    TCTGGTCCATTGGCGCCTCCGCAGCTGGCGGGGAGGTCTATAAACAGGGCCAGGACACGTTCCCAGTTCTGCTTCCTGTG    160
    GGAATGGAAGCGATGGGACCCTGCCCGGCCGCACGGAGAGGAAGAATAGGATAGACCCCAGATTCCCACCTTGACAGGC    240
    AGATTTGCAGGAATGAGTCGTGTGTATCATAGCGGCTTGGAAGTCTGCATTTCTGCGCCCACCTTCCTGAGACACAGGGT    320
    TACCCGGGTCCCAGTGTCCATATCGTAAAGGGCGTTGGGCAGCTACTCTTGAATCTTCTCTGCTTCCTTGGTCTTTCATT    400
    ACAATTTTGGAAGACTTATGGGTCAGGATACTTCTGAGCCCTAAGTGCGCAAGGAAGCAAGGGGTTCTCTGGGGACTCAT    480
    CAGACCCTTTTCTCTTTGCACAGACCGGATTTTCTCCTTTATTTCTTTGAGGCAGGGTTTTATGTGGATCAAGCTGACTT    560
    TAAGACACTATGTAACTGAGGATCACCTTGGACTTCTGCTTCTCCCCTCCTTCCCTTCTCCTCCCAGTGCAGAGAACAGG    640
    AGCCTGGATCAACCCAGGTCTTTGTGATGCAGGTTCCCCACCAAACTGAGAATTATCCTCAGCCCATCAAAAAACAGTTC    720
    TTTCTTTCTTTCTTTTTTTAAAAATTATTTATTTATTTGCTTGTTTATTTTGTTTTTTGAGACAAAGTTTCTCTTCATAG    800
    TCCTGCTGTCCTGGAACTCACGTTGTAGACCAGGGCTGGCCTCGAACTCACAGAGATCTGCCTCCATCTCCCTTCTGAGT    880
    GCTGGGACTAAAAATTCTGTGTGGTCTTTAATGAGTTTTTTGAGCCCTAAAATGTAATCAAGATGTATGCCTGCCAATCC    960
    AACACTTGTGAGGTGCAGGCGGGAGGATCAAGAGTTCAAGATCAGCCTCTGCTACTTGGCAAGTTCGAGACCAGCTTGTT   1040
    ACTTGAGACCATTTCTCAAAACAAAACCTAAACTCCTGAGTGCTTCACGTCTCAAAAGCAAGGACTTCTAACCACAAGTT   1120
    TCCAACCAGAGCTAGGTCCTGGCCTCAGCCTCTCCGCCCTTCAAACACCTAGCAGTCACCATTGGGCCTCTAAAGACGTA   1200
    GGAATCGCTGATTTGATATTCGGTCCGTCGCTCGAGAGCATCAGCCAATGGAAACTTCATGCCTCCCTAGAGCCCCGTCT   1280
    TCCAGTGGCTTACCAGGCCCTGCCCCTCTTCTGCCGGCGCAGGCCAGTGTGGTCCCGCGGCTGTCGATGCAGTCCCTCCC   1360
    CTAACGTCACAGCTTCGTTTGCATAGAAAGCCCCGCCCCTGGCTTTGCAGGACGTCAGCTAGGACTGCAGGGGCCTGAGC   1440
                                                 ↓
    CTCCTCCGCTGCCGCCTGCAGCCCCGGCCGCCGGCTGCATGCCCCGCATCCTCTTCTGGGGCCCGGTCGCCAGCGCAGTC   1520
                       A                B                           AA             BB
I   GCCACGGTCGCCGTCGCCAGCGTTGTCTCAGCTTAGAGAGGGTAAGGTGAGCGCAACCCAGGTCCACACACTGAGGAGCC   1600
           CC     DD
    CGCGTTTCACTGACAGCTGGGGATCAAGGGGAGCGCAACTCGGGTGTAGTGGCTATGGGTCTCGCATTTGGCATTCTTG   1680
    GTCCCGCATCCCGCACCCCTCGCCCTTAACCTGGCACCTGGGGCCCGGGCTCTGCGACAGAGGGGGAGGGGAAAGCCAGG   1760
    CCTTGCGGGAGGCTTGTGCGTGCTCGTGCTCGTGGGGTGAACACGCGTGTGCGCGCCCGCTGGGCATGGAGAGGCTCGAG   1840
    TGACCGCAGGACCGCTGCGTGGCACACTTGTGATACGTTGTGCGTGCATGGTCCGCGTGCGCGGTGCGCGTGGATCTTAG   1920
    TTCAAGTATGTGCTCACACGCGTGTGGAGCCATGCGCGTGTGTGGTGTGTCCTCCCGGATGGCCAGGTACAGTCGTGCAT   2000
    TGGTGCAGGGAGCTGTGCTCACACATGGGCTTGCACACACCATACATGTCTGCTTGTTTCGCCTGTTAGCACCACTCTTG   2080
    CATGCAGGGCAGGTGTTTGGGTCTGCCTGCCCTAGATGCTTCATTGTGCTTGTGTGTGTGTGTGTGTGTATGCGTGCG   2160
    CGCTCTTGTGCTCGCGATTTTGTGTGTGTGAGCCTACGTTCATCCTGTCTTCGTCACGTGATGTGGCTATCTCAGGGCCC   2240
    CAGTTCTTGCAGGAGAAAAATAAGGATGAGCGAGAACTGGGGATCCTCCTGTATTCATGTTCCGGTGTGGGTCTCTGTGA   2320
    GAGACTGATGGAAGGCAGAGGAGGCTCAGTCTTCCCTCAGGGTACAGGGGAATCTGAGGTTCACAGCCTGCTTAAAACC   2400
    TTGGTGCTGTCAAGGTCACAGCTCTGCCTATGTTGGTTGGGGGGGTTGAGCATGTGACATTGGGGGGTATCCTCACCTAC   2480
    CCCTGTGCCGTGCCCACCCTGTGCGTGGTCCTGATGCTCCCCTCAGGCCCTCTGTGCCACGCCTCCCGCAGATGGCTTCC   2560
                                                                             M  A  S      3
    GCCACAGACTCTCGCTATGGGCAGAAGGAGTCCTCAGACCAGAACTTCGACTATATGTTCAAGATCCTGATCATTGGGAA   2640
     A  T  D  S  R  Y  G  Q  K  E  S  S  D  Q  N  F  D  Y  M  F  K  I  L  I  I  G  N     30
II  CAGCAGCGTGGGCAAAACCTCGTTCCTCTTCCGCTACGCAGATGACTCCTTCACTCCAGCCTTTGTCAGCACCGTTGGCA   2720
      S  S  V  G  K  T  S  F  L  F  R  Y  A  D  D  S  F  T  P  A  F  V  S  T  V  G       56
    TAGACTTCAAGGTCAAAACCATCTACCGCAACGACAAGAGGATCAAGCTGCAGATCTGGGTGTGTAGGGCTGGAGGGGAG   2800
    I  D  F  K  V  K  T  I  Y  R  N  D  K  R  I  K  L  Q  I  W                          76
    ATGTCTTCCTGCCAACCCCCACAAAAGGAAGCTGAGGCTAGCCATCGGCTGGCGCTGGGATTTAAGGCCTGCATCTCCAT   2880
    CCCAGTATGAGCTTCCCGTTTTCCTAAGTCCCGAGCATCCCCACTCATCTTCTAGCATAGTGTGGCACAGGGCATTCAGG   2960
    CAGCACCCAGTACATGTTTGTTGAGTGACTAAGTGGCCTTGCAGTGACCCATTTGGGTCACTGTCTGCTTTTCCTGACCC   3040
    ACTGTCCCCGGGCCTCAAAGTGGTTGCTGGACAACCTTGGAGCACGTTTTTCTCTCGGGGGCTGCTGCGAGGTCCCTTCT   3120
    GACCTATCCCTCTTTTCCCCAGGACACAGCAGGGCAAGAGCGGTACCGCACCATCACCACAGCCTATTACCGAGGCGCCA   3200
                          D  T  A  G  Q  E  R  Y  R  T  I  T  T  A  Y  Y  R  G  A         95
III TGGGCTTCATCCTAATGTATGACATCACCAATGAGGAGTCATTTAATGCAGTGCAGGACTGGTGAGCAGCCCACCTACTC   3280
    M  G  F  I  L  M  Y  D  I  T  N  E  E  S  F  N  A  V  Q  D  W                      116
    AGAAAACACAAGGCTGGGATTGGAGGTGACTTAGCTAAGCTTAGCTAAGCTTAGTTGGTAGGGTGCCTGCTTGGCCAGGA   3360
    CAAAGCCTTGGGTTCGTCCCCAGCATGGCCAAAAGACAACCAAGAGTAGGCCTCTGGTGGGACAAGCCATGTGGCTGTTC   3440
    CCAAGGCTTGAGGTTGTGGAAGCACACAAGGTAACCAGTGAGTAGACAAGGAACCTCCCACTGATCTCTACCCAGGTAGC   3520
    CCAGGCAGGAGAACCACGTTAGCTCACTTGTCTGAGGTCAAGTGATTCTCTGGCCAAGCGCTTACCTCAAATGTATGAGA   3600
    ACCTGGATTCCAGCCACAGCACCAAAAATTAACTCCCCCAGGTCTTGACATTTGGTGTAGTGGACACCGGCCATCTTCCC   3680
    AGCTTTGGGAAGATGGAGAGGTCTGAGTGAGATAGGCTGGACTCAGTTACCAAGGGGCAGTGGTCTCTCCTGCAAGGGCT   3760
    AGACACAGGCAGAGACTCGGGGATGGGGCCAGGTCCTCCGGCTTCCTCTTCAGCTCTGGCTGTCTCTGCAGGTCCACTCA   3840
                                                                              S  T  Q   119
IV  GATCAAAACTTACTCGTGGGACAATGCCCAGGTGCTGCTGGTGGGAAACAAGTGTGACATGGAAGATGAGCGAGTGGTGT   3920
      I  K  T  Y  S  W  D  N  A  Q  V  L  L  V  G  N  K  C  D  M  E  D  E  R  V  V      145
    CCTCAGAGCGTGGCCGGCAGCTGGCTGACCACCTGGGTGAGTACTGAGGACTGTGGCCCTTACCAGGCTGGACCAGAGAC   4000
    S  S  E  R  G  R  Q  L  A  D  H  L                                                  157
    CTGGATACCTGATAGCTACACACCCAGTGTCTCACTGATCAGGACATGGGCTTGAATCCAGGGGGTTTGCACTTGAGAAA   4080
    ACTAGGCCTTATCCAGGAGGCTTTACCATGGGTCTTGATGGACATGGAGGTGCTTGCCATGTTGGCTGTCAGCTGTCAGC   4160
    CATGAGCTACCTTCAAGGTACTGGGCCTGTGAAGTTTTCGAACTAGAAAGAGAAATGGGTGGAAGAGGCTAAGCCTGGCT   4240
    TCCCGGAGCAGGCACAACATGCCTGTATGGGAAGGTTGGTATAGGCAGAGTCGTGACAGTGGTGCGTGATCGCCCCTTAG   4320
    AGGAACTGTGGCTGTCACTGCAGGCTGAGCTGACAAGAAACCAGGCCATGGAGGTTGTAGGTCTCTGGAAAGGGACTTGG   4400
    AGCCTGGCTGGGTGGCAGATGGAAGGGATGTCTCACTAGGGTGCTTAAGATTATCCCAAGGGATGGGTAGTAATGGATGG   4480
    CTCACCCAGACCAGTGTGGCCTCCCTGGGAGGGACCCTTCAGTGCGAGGGCCTGTGTGAGTGCGTCTGGAGCCTACAGCT   4560
    CTGTTGCCACCCCTCCATTTCCCCAATATAGGCTTTGAGTTCTTTGAGGCCAGCGCCAAGGACAACATTAATGTCAAGCA   4640
                                   G  F  E  F  F  E  A  S  A  K  D  N  I  N  V  K  Q    174
V   GACGTTTGAACGTCTGGTGGACGTGATCTGTGAGAAGATGTCAGAGTCCCTGGATACTGCAGACCCTGCGGTCACCGGTG   4720
      T  F  E  R  L  V  D  V  I  C  E  K  M  S  E  S  L  D  T  A  D  P  A  V  T  G      200
    CCAAGCAGGGCCCGCAGCTCACCGACCAGCAGGCGCCACCTCATCAGGATTGTGCCTGCTGAGCCACTTCCCTTCCCTGC   4800
    A  K  Q  G  P  Q  L  T  D  Q  Q  A  P  P  H  Q  D  C  A  C  *                       220
    TGCCAGGGGCACGTCCCCCATCCCCGACTACCCTATTTATTATTGTAGCTATTTATTTATTTATTGAGGATGTGCCCCGA   4880
    GCGCACCCCCTTCCCACCCTGTACATAGCTCCACCCAGCTCGGCCGTGGTGTATTGTGGTCACTGCTGCTCCCTCTCCTT   4960
    TACCCCCCACCCCCATTTTGTTTTTGTAAACCATCCCGTCCACAGTTGTCAGTTGTGAAGAGGGGACCAGATGTCACTCT   5040
    GCAGCAGGCTGATGGAAGATGGCGGGGCCTGCCAACCCCGAGGCTGGGGGATCGCCATATGGACCACCTGGTTGACAGAC   5120
    GGCTTCTTGCTTGCCTGGGCCTTCTGCCTTATACTTTGGGATAAATGGGGTGTAGGGGATATATTCACTACTTTGGGATA   5200
    CCCCCAAACCTGTTCTGACGTCGTGAGTGTCAAATGTGTCCCACTCCTCCCTAGAAGTTATGAACACTACACACAGTCAA   5280
    TAAAGCGAATGGACCTTGCTGAGTCTGCTGAGCCAGCCCCTGAACCCACCCACACGCAGTCCTTGACATTAAAGAGAATT   5360
    TAATGAATTTCCGCACCCTGCCCAGCCTTGCATCAAGTTTTCTGAGGCCAAACCAGACCCCACTTCAAGGTTGTGGAGGG   5440
    TGTCCAGGCAGACAGAGGGCTTCATGGTCCACTTGGAAACCGGGAGGGTTTGGGTGGGAGGCGACAGCCTGTGCTGGCGT   5520
    CTGTTCTACATGGGAACAGTCTCTGGTGGACTCAAGGGACTGCTGGTCCAGGGGCAGTCAGCCTGGGTTTTGGACATGGC   5600
    CCAGAGCTTTTTGACCAGCTTCGGTAATCTCCTGTGGGAGTGTATCCATCGGCCAACTCATCTGCCAGGCTTCAGTTTCA   5680
    GTCTGCACAGCCCGGAGTCTGAAGGGGTTCAGGGACCTGTCCGCAGCAGAGGGACAGTCAAACAGGAATAGGCCGGAACC   5760
    CACGCCCTCCCACACTCCCACACCATGGAGTTGAGCTCACCCAGTACTTCAGATAGGACAGGTATGTGTCCCAGCCCAGT   5840
    GTCAGCCCATTGATATAGGTGACTCGGAAATGAGAAGGGATGAAGAGGAAGTTCACTAGCTGGGCAGCTGGCCACACACA   5920
    CCAGTCGGCCTGTGGAAAAGTGGGAGGGCCACTGTGAGAGGGGGGGCGGCTACGAGGGGCGGGGCTTGAAAAGTGGGTGG   6000
    GGGCCAAGGCCTTGGGTGTGGGTCTGGAAGGTAAAGAGGGGGGGTGGAGCCATGTTGAGAGAGCCCACGTGGGCACGCCC   6080
    ACCTTGTAGAAGTCCCAGAATTTGGCCCGCAGCTCCTGGCAGCTCTCCTCTAGTGTCTGACCCTCCAGGCTGCCAAGGCC   6160
    TGGGAAGAAGACCAAGAAGTAGTTATAAAGGCTGAGGTGGCTGCAAGGCAGAAGCAGGGCAGGTCTCAGGCAGAGCAGCA   6240
    GGCAGGTCCAGGAAAATGAGAGGTTGGTGTGAGGGAGCCTGAGTTACTTATGAAAACATAAATGCGTGAGCCTAGATCAC   6320
    GAGCCCCATCGCTGTGTTAACAGGTACC                                                       6328
```

Figure 2 For legend see facing page

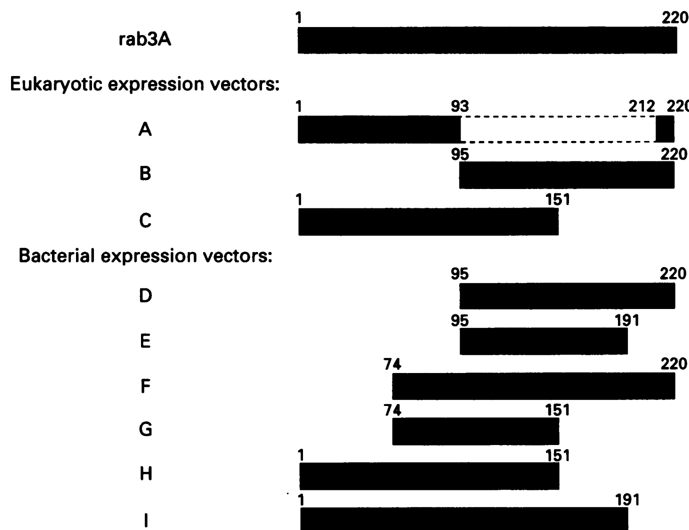

Figure 3 Structures of the rab3A deletion mutants used for epitope mapping

The structure of full-length rab3A is shown diagrammatically (top), with the N-terminus on the left and the C-terminus on the right. Two types of constructs were used: eukaryotic expression vectors analysed by transfection into COS cells (constructs A–C; immunoblots are shown in Figure 4), and bacterial expression vectors for *Escherichia coli* (constructs D–I, immunoblots are shown in Figures 5 and 6).

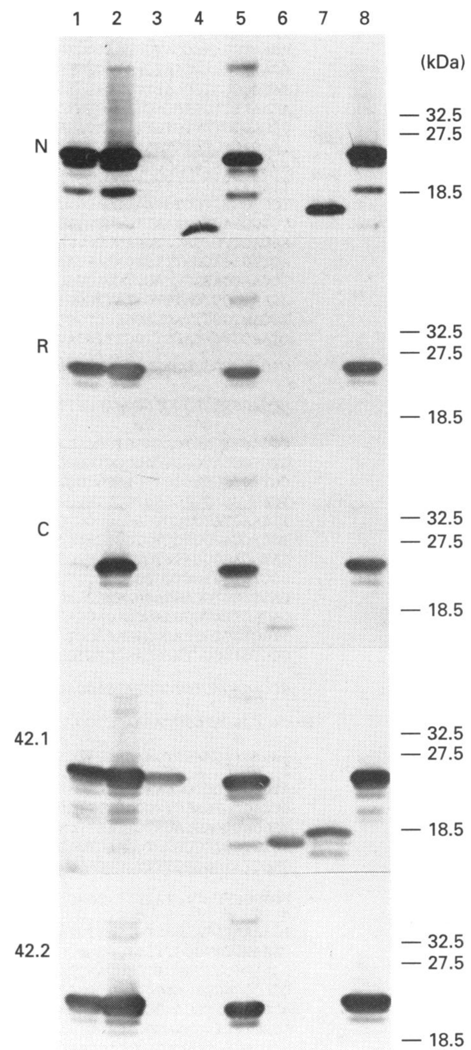

Figure 4 Specificity of anti-rab3A antibodies

Immunoblot analysis of antibody reactivity with proteins from COS cells transfected with different rab3-related expression vectors. Five different antibodies were used as indicated on the left of each blot: an antibody to a peptide from the N-terminus of rab3A (labelled N; antibody T957), an antibody raised against total recombinant rab3A (R; antibody V752), an antibody to a C-terminal peptide (C; antibody V215), and two independent monoclonal antibodies raised against recombinant rab3A (42.1 and 42.2). Lanes were loaded with protein from COS cells transfected with the following constructs (see Figure 3 for a description of the alphabetically named constructs): 1, bovine rab3A; 2, human rab3A; 3, bovine rab3B; 4, construct A; 5, bovine rab3A; 6, construct B; 7, construct C; 8, bovine rab3A. Total protein from transfected COS cells was analysed by SDS/PAGE, followed by immunoblotting using peroxidase-labelled secondary antibodies for detection. Numbers on the right indicate positions of molecular mass markers.

imperfect 18 bp repeat were found surrounding the putative transcription start site (shaded in Figure 2 and labelled A and B).

The intron donor and acceptor sequences of the *rab3A* gene agree well with the consensus sequences for such sites (underlined in Figure 2). The introns are short, with the first intron being the largest (985 bp; see Figure 2). This accounts for the relative shortness of the *rab3A* gene. Databank searches (GenBank release 70) revealed no significant similarity of the sequence to current entries other than rab proteins.

Comparison of the deduced murine amino acid sequence for rab3A with the bovine, human and rat sequences reveals a high degree of conservation, with a single amino acid change between bovine and murine rab3A and two amino acid changes between human and murine rab3A. All amino acid changes are at either the N- or C-terminus. This finding confirms that the present gene represents that for rab3A and attests to the selective evolutionary pressure on this protein which extends as far as *Drosophila* (Johnston et al., 1991).

The crystal structures of ras proteins in several ligand-dependent conformations (Wittinghofer and Pai, 1991), as well as the genomic organizations of several *ras* genes (McGrath et al., 1983; Taparowski et al., 1983; McCaffery et al., 1989), have been determined. Surprisingly, none of the introns in the *ras* genes and the *rab3A* gene are placed at corresponding positions. Since the genomic organization of no other *rab* gene has yet been determined, it is impossible to tell if the intron structure of *rab3A* represents a conserved pattern that is shared by genes for different rab proteins (but not the evolutionarily more distant ras proteins) and that might reflect a division of protein domains. To test if the rab3A exons might represent protein domains, we decided to map the epitopes of monoclonal antibodies in the rab3A structure, on the assumption that epitopes represent exposed protein domains.

Construction and expression of mutant rab3As to map antibody epitopes

A series of eukaryotic expression vectors was constructed that included full-length bovine rab3A, rab3B and human rab3A as well as three deletion mutants of rab3A (Figure 3). These plasmids were transfected into COS cells and the reactivity of the different expressed proteins with different antibodies was assessed by immunoblotting (Figure 4). Initially, eukaryotic expression was chosen for these studies in order to be able to evaluate the ability of different constructs to also become isoprenylated.

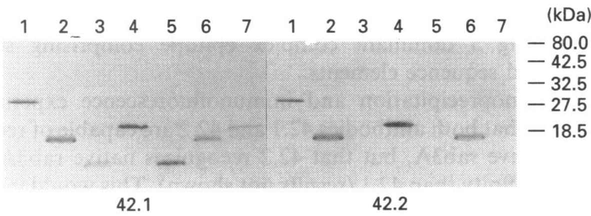

Figure 5 Immunoreactivity of different rab3A constructs expressed in *E. coli* with the monoclonal anti-rab3A antibodies 42.1 and 42.2

Bacteria transformed with the different expression vectors were induced and analysed by SDS/PAGE and immunoblotting. Lanes correspond to the following constructs (see Figure 3 for a description of the constructs): 1, full-length rab3A; 2, D; 3, E; 4, F; 5, G; 6, D; 7, H. Immunoreactive bands were visualized by peroxidase-labelled secondary antibodies. Numbers on the right indicate positions of molecular mass markers.

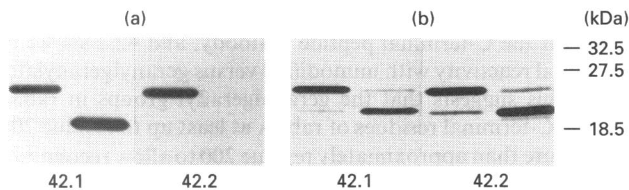

Figure 6 Fine mapping of the epitope for monoclonal antibody 42.2

In (**a**), full-length recombinant rab3A (left lanes of each immunoblot) or rab3A truncated at residue 191 (construct I in Figure 3; right lanes) were analysed by immunoblotting with the two monoclonal antibodies as indicated below each blot. In (**b**), full-length recombinant rab3A (left lanes of each immunoblot) was analysed with the same antibodies and compared with rab3A that had been truncated at its C-terminus by selective proteolysis (right lanes in **b**). Numbers on the right indicate positions of molecular mass markers. Immunoreactive bands were visualized with peroxidase-labelled secondary antibodies.

However, preliminary studies suggested that relatively small deletions abolished isoprenylation of rab3A. Therefore a second series of expression constructs was designed for bacterial vectors (Figure 3) which were also analysed by immunoblotting (Figures 5 and 6).

The epitopes for three antibodies were investigated: two independent monoclonal antibodies designated 42.1 and 42.2 (Matteoli et al., 1991) and one polyclonal antibody against full-length rab3A (R in Figure 4). In addition, two polyclonal anti-peptide antibodies with known epitopes at the N- and C-termini of rab3A (N and C respectively in Figure 4) were used as internal controls for most experiments. The results of these experiments are shown in Figures 4–6, leading to the following conclusions.

1. Monoclonal antibody 42.1 reacts with bovine and human rab3A and rab3B; in addition, it reacts with all rab3A deletion mutants except for construct A. This result suggests that this monoclonal recognizes a conserved rab3 epitope localized between residues 95 and 151. These residues are within the GTP-binding domain of rab3A.

2. The second monoclonal antibody (42.2) recognized human and bovine rab3A but not rab3B; of the constructs, it reacted only with mutant rab3A proteins extending beyond residue 191. However, residues 212–220 only were not sufficient for recognition. These results suggest that antibody 42.2 recognizes an epitope between residues 191 and 212. This epitope can be further narrowed down by two observations. First, rab3A contains a hypersensitive proteolytic site that removes approx. 2.5 kDa from its C-terminus. The proteolysed form of rab3A is still recognized by antibody 42.2, but the form of rab3A truncated at residue 191 is not (Figure 6). Secondly, human and bovine rab3A differ by only three amino acid residues (positions 7, 208 and 212; see Matsui et al., 1988; Zahraoui et al., 1989). The polyclonal antibody against a C-terminal peptide from rab3A recognizes the bovine but not the human rab3A protein, evidently because of the amino acid substitutions at positions 208 and 212 (C in Figure 4). The monoclonal antibody 42.2, however, recognizes both human and bovine rab3A, suggesting that the divergent residues are not part of its epitope. Together, these results suggest that the epitope for 42.2 resides between residues 191 and 207.

3. The polyclonal antibody raised against recombinant rab3A (R in Figure 4) shows a pattern of reactivity that is much more restricted than that of the monoclonal antibodies. It reacts strongly with wild-type rab3A, but not with rab3B or with the mutant rab3As. This suggests that this antibody's dominant epitope is complex and requires multiple separated sequences.

Tissue distribution of rab3A

The epitope mapping of the monoclonal antibodies against rab3A revealed that antibody 42.1 recognizes sequences that are highly conserved between different rab3s, whereas antibody 42.2 reacts with the C-terminal hypervariable domain specific for rab3A. This afforded us an opportunity to probe a wide variety of rat tissues with both antibodies to test if related members of the rab3 family are present in these tissues that react with one antibody but not with the other (Figure 7; note that the lanes

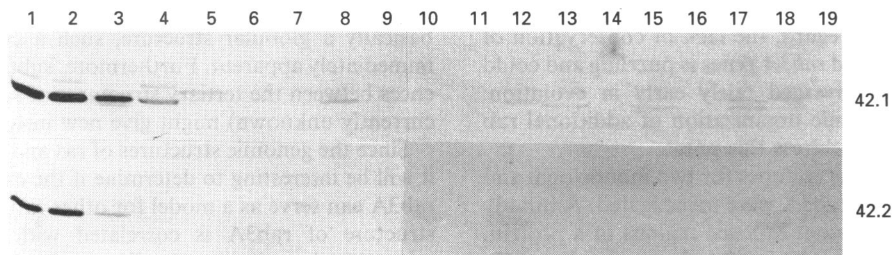

Figure 7 Tissue specificity of rab3A expression

Immunoblots of rat tissue homogenates were probed with the monoclonal anti-rab3A antibodies 42.1 and 42.2. Lanes were loaded with homogenates from the following tissues (5 μg of protein for lane 1; 10 μg for lanes 2 and 3, and 100 μg for lanes 4–19): 1, brain cortex; 2, cerebellum; 3, spinal cord; 4, adrenal gland; 5, salivary gland; 6, intestine; 7, testis; 8, ovary; 9, spleen; 10, lung; 11, liver; 12, heart atrium; 13, heart ventricle; 14, skeletal muscle; 15, kidney; 16, aorta; 17, stomach; 18, adipose tissue; 19, thyroid gland.

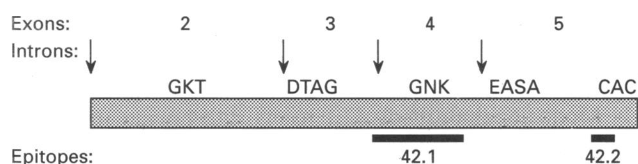

Figure 8 Relative localizations of the GTP-binding consensus sequence motifs, the introns and the epitopes for the two monoclonal antibodies in the primary structure of rab3A

The numbers of the exons are indicated, and the positions of the introns are indicated by vertical arrows. The locations of the four major rab protein consensus sequence motifs GKT, DTAG, GNK and EASA (Chavrier et al., 1990) are shown above the bar representing the rab3A protein, and the geranylgeranylation sequence CAC is indicated at the C-terminus. The locations of the epitopes for the two monoclonal antibodies 42.1 and 42.2 are also shown.

with non-neural tissues contain 10 times as much protein as the lane with protein from brain). Although bands that reacted weakly with one or both antibodies were observed in several extraneural tissues, no major band was found; in particular, none occurred that reacted differently with 42.1 as opposed to 42.2. This result suggests that rab3A is indeed specific for neural and endocrine tissues and that there is no closely related homologue in the tissues tested, e.g. in constitutively secretory tissues. The weakly reacting proteins at higher molecular masses that are observed in some tissues (particularly with antibody 42.1; see Figure 7) are probably not related to rab3A, as they do not react with any other anti-rab3A antibody (results not shown).

DISCUSSION

In the present study, two different approaches were utilized to probe the structure of the rab3A protein. First, the primary structure of the murine gene coding for rab3A was determined by cloning and sequencing. These studies revealed a genomic organization of rab3A into exons that is completely different from that observed for ras proteins, the only other low-molecular-mass GTP-binding proteins for which the gene structure is known (McGrath et al., 1983; Taparowski et al., 1983; McCaffery et al., 1989). Consisting of less than 8 kb, the murine *rab3A* gene is relatively small. Except for the absence of a TATA box, its structure conforms well with the features of eukaryotic genes. The availability of the sequence of a promoter that is highly expressed almost exclusively in neurons (see Figure 7) should be useful for studies on neuron-specific expression.

The *rab3A* gene contains four coding exons, each of which includes one of the four major conserved sequence motifs of low-molecular-mass GTP-binding proteins (summarized in Figure 8). The organization of coding sequences into exons may be informative not only in evolutionary terms but also as an indicator of protein domains. In this regard, the lack of conservation of exon units between the *ras* and *rab3A* genes is puzzling and could indicate that these genes diverged fairly early in evolution. Further studies on the genomic organization of additional rab proteins will be required to address this point.

In our second approach, the epitopes for two monoclonal and one polyclonal antibody on rab3A were investigated. Antibody epitopes are thought to represent exposed regions of a protein. The epitope for monoclonal antibody 42.1 was mapped to a 57-amino-acid sequence in the middle of rab3A that includes a highly conserved GTP-binding motif. The epitope for monoclonal antibody 42.2 was localized to a 17-amino-acid sequence in the C-terminus of rab3A at a position where the sequences of different rab proteins diverge (the so-called hypervariable region). The polyclonal antibody only recognized the full-length protein, suggesting a dominant complex epitope comprising several separated sequence elements.

Immunoprecipitation and immunofluorescence experiments suggest that both antibodies 42.1 and 42.2 are capable of reacting with native rab3A, but that 42.2 recognizes native rab3A with higher affinity than 42.1 (results not shown). This would indicate that the hypervariable domain epitope of 42.2 is more exposed than the consensus sequence motif epitope of 42.1. This result is not surprising considering the fact that the latter is part of the GTP-binding domain. It supports the notion that the hypervariable domain serves as a flexible link between the GTP-binding domain and the C-terminal geranylgeranylation site that attaches rab3A to the membrane.

We have previously shown that an antibody directed to the C-terminal peptide of rab3A (whose epitope comprises amino acid residues 208–220) reacts much better with unmodified rab3A than with geranylgeranylated rab3A (Johnston et al., 1991). The epitope for monoclonal antibody 42.2 immediately precedes the epitope for the C-terminal peptide antibody, and 42.2 shows no preferential reactivity with unmodified versus geranylgeranylated rab3A. This suggests that the geranylgeranyl groups in rab3A cover the C-terminal residues of rab3A at least up to residue 208, but not more than approximately residue 200 to allow recognition of 42.2. Finally, the mapping of the antibody epitopes allowed us to assess the tissue distribution of rab3 proteins, demonstrating that rab3A and close relatives of rab3A are primarily expressed in the nervous system (Figure 7).

The tertiary structure of ras has been extensively studied by X-ray crystallography (reviewed in Wittinghofer and Pai, 1991). Alignment of the rab3A and ras sequences makes it possible to place the rab3A epitopes on to the ras structure. This analysis reveals that the sequences in ras which correspond to the antibody epitopes in rab3A are surface-exposed. This is particularly striking for the epitope for 42.2, which lies at the beginning of the C-terminal hypervariable region. It was actually not possible to determine the structure of this region in the ras crystals, presumably because it is very mobile and extends from the surface of the globular GTP-binding domain. The epitope for 42.1 corresponds to the region from loop 5 to loop 8 in the structure of ras, most of which is also surface-exposed. It is interesting that none of the actual GTP-binding residues of rab3A are included in the antibody epitopes.

It has been argued that the division of proteins into exons in the genome serves to divide proteins into modular elements that can be exchanged during evolution (Gilbert, 1978; Südhof et al., 1985). Placement of the positions of the introns into the tertiary structure of ras (Wittinghofer and Pai, 1991) reveals no obvious correlation between protein domains and exons, except for the fact that each of the GTP-binding motifs is contained in a separated exon. However, since the GTP-binding domain is basically a globular structure, such a correlation may not be immediately apparent. Furthermore, subtle but significant differences between the tertiary structures of ras and rab3A (which is currently unknown) might give new insights.

Since the genomic structures of ras and rab3A are so different, it will be interesting to determine if the exon–intron structure of rab3A can serve as a model for other rab proteins. The genomic structure of rab3A is correlated with the epitopes for the monoclonal antibodies in Figure 8, demonstrating that the epitopes, which presumably represent exposed regions of rab3A, are on separate exons. An X-ray structure of rab3A would be very valuable in order to determine whether the differences in function and genomic organization between rab3A and ras can be correlated with differences in their tertiary structures.

This study was supported by the Perot Family Foundation and by a postdoctoral fellowship grant to M.B. from the Deutsche Forschungsgemeinschaft. We thank I. Leznicki, A. Hopkins and E. Borowicz for excellent technical assistance.

REFERENCES

Balch, W. B. (1990) Trends Biochem. Sci. **15**, 473–477

Baldini, G., Hohl, T., Lin, H. Y. and Lodish, H. F. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 5049–5052

Bourne, H. R. (1988) Cell **53**, 669–671

Chavrier, P., Vingron, M., Sander, C., Simons, K. and Zerial, M. (1990) Mol. Cell. Biol. **10**, 6578–6585

Farnsworth, C. C., Kawata, M., Yoshida, Y., Takai, Y., Gelb, M. H. and Glomset, J. A. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 6196–6200

Fischer von Mollard, G., Mignery, G. A., Baumert, M., Perin, M. S., Hanson, T. J., Burger, P. M., Jahn, R. and Südhof, T. C. (1990) Proc. Natl. Acad. Sci. U.S.A. **89**, 1988–1992

Fischer von Mollard, G., Südhof, T. C. and Jahn, R. (1991) Nature (London) **349**, 79–81

Gilbert, W. (1978) Nature (London) **271**, 501–502

Hall, A. (1990) Science **249**, 635–640

Jahn, R., Schiebler, W., Ouimet, C. and Greengard, P. (1985) Proc. Natl. Acad. Sci. U.S.A. **82**, 4137–4141

Johnston, P. A., Archer, B. T., Robinson, K., Mignery, G. A., Jahn, R. and Südhof, T. C. (1991) Neuron **7**, 101–109

Khosravi-Far, R., Lutz, R. J., Cox, A. D., Conroy, L., Bourne, J. R., Sinensky, M., Balch, W. E., Buss, J. E. and Der, C. J. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 6264–6268

Laemmli, U. K. (1970) Nature (London) **227**, 680–685

Matsuda, K., Sakamoto, C., Nakano, O., Konda, Y., Matozaki, T., Wada, K., Kasuga, M., Mizoguchi, A., Kikuchi, A. and Takai, Y. (1992) Am. J. Physiol. **262**, G69–G73

Matsui, Y., Kikuchi, A., Kondo, J., Hishida, T., Teranishi, Y and Takai, Y. (1988) J. Biol. Chem. **263**, 11071–11074

Matteoli, M., Takei, K., Cameron, R., Hurlbut, P., Johnston, P. A., Südhof, T. C., Jahn, R. and de Camilli, P. (1991) J. Cell Biol. **115**, 625–633

McCaffery, R. E., Coggins, L. W., Doherty, I., Kennedy, I., O'Prey, M., McColl, L. and Campo, M. S. (1989) Oncogene **4**, 1441–1448

McGrath, J. P., Capon, D. J., Smith, D. H., Chen, E. Y., Seeburg, P. H., Goeddel, D. V. and Levinson, A. D. (1983) Nature (London) **304**, 501–506

Mignery, G. A., Newton, C. L., Archer, B. T. and Südhof, T. C. (1990) J. Biol. Chem. **265**, 12679–12685

Mizoguchi, A., Kin, S., Ueda, T. and Takai, Y. (1989) Biochem. Biophys. Res. Commun. **162**, 1438–1445

Olofsson, B., Chardin, P., Touchot, N., Zahraoui, A. and Tavitian, A. (1988) Oncogene **3**, 231–234

Pfeffer, S. R. (1992) Trends Cell Biol. **2**, 41–46

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor

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

Sano, K., Kikuchi, A., Matsui, Y., Teranishi, Y. and Takai, Y. (1989) Biochem. Biophys. Res. Commun. **158**, 377–385

Sasaki, T., Kikuchi, A., Araki, S., Hata, Y., Ismoura, M., Kuroda, S. and Takai, Y. J. (1990) J. Biol. Chem. **265**, 2333–2337

Schnefel, S., Nimmermann, P., Profrock, A., Jahn, R., Aktories, K., Zeuzem, S., Haase, W. and Schulz, I. (1992) Cell. Physiol. Biochem. **2**, 77–89

Smale, S. T., Schmidt, M. C., Berk, A. J. and Baltimore, D. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 4509–4513

Studier, F. W., Rosenberg, A. H., Dunn, J. J. and Dubendorff, J. W. (1991) Methods Enzymol. **185**, 60–89

Südhof, T. C. (1990) J. Biol. Chem. **265**, 1988–1992

Südhof, T. C. and Jahn, R. (1991) Neuron **6**, 665–677

Südhof, T. C., Goldstein, J. L., Brown, M. S. and Russell, D. W. (1985) Science **228**, 815–822

Taparowski, E., Shimizu, K., Goldfarb, M. and Wigler, M. (1983) Cell **34**, 581–586

Touchot, N., Charding, P. and Tavitian, A. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8210–8214

Towbin, H., Staehelin, T. and Gordon, J. (1979) Proc. Natl. Acad. Sci. U.S.A. **76**, 4350–4354

Wittinghofer, A. and Pai, E. F. (1991) Trends Biochem. Sci. **16**, 382–387

Zahraoui, A., Touchot, N., Chardin, P. and Tavitian, A. (1988) Nucleic Acids Res. **16**, 1204

Zahraoui, A., Touchot, N., Chardin, P. and Tavitian, A. (1989) J. Biol. Chem. **264**, 12394–12401

Received 25 November 1992/2 February 1993; accepted 5 February 1993