doi:10.1006/jmbi.2000.4010 available online at http://www.idealibrary.com on IDEAL® *J. Mol. Biol.* (2000) **301**, 1077–1087

JMB

The Mammalian Rab Family of Small GTPases: Definition of Family and Subfamily Sequence Motifs Suggests a Mechanism for Functional Specificity in the Ras Superfamily

José B. Pereira-Leal and Miguel C. Seabra*

Molecular Genetics Section Division of Biomedical Sciences Imperial College School of Medicine, Sir Alexander Fleming Building, London SW7 2AZ, UK

*Corresponding author

The Rab/Ypt/Sec4 family forms the largest branch of the Ras superfamily of GTPases, acting as essential regulators of vesicular transport pathways. We used the large amount of information in the databases to analyse the mammalian Rab family. We defined Rab-conserved sequences that we designate Rab family (RabF) motifs using the conserved PM and G motifs as ''landmarks''. The Rab-specific regions were used to identify new Rab proteins in the databases and suggest rules for nomenclature. Surprisingly, we find that RabF regions cluster in and around switch I and switch II regions, i.e. the regions that change conformation upon GDP or GTP binding. This finding suggests that specificity of Rab-effector interaction cannot be conferred solely through the switch regions as is usually inferred. Instead, we propose a model whereby an effector binds to RabF (switch) regions to discriminate between nucleotide-bound states and simultaneously to other regions that confer specificity to the interaction, possibly Rab subfamily (RabSF) specific regions that we also define here. We discuss structural and functional data that support this model and its general applicability to the Ras superfamily of proteins.

Keywords: Rab proteins; Ras proteins; GTP; sequence motif; vesicular transport

Introduction

The Ras superfamily of GTPases encompasses a large group of proteins that bind GTP and GDP and serve as molecular switches to regulate important cellular processes such as growth, motility and protein trafficking (Barbacid, 1987; Bourne *et al*., 1990, 1991). General guidelines for nomenclature on Ras-like GTPases were set in 1992, following a FASEB Summer Conference (Kahn *et al*., 1992). The proposal that there are five different families within the superfamily (Ras, Rho/Rac, Rab, Arf, Ran) is now widely accepted.

The largest branch of this superfamily is formed by the Rab/Ypt/Sec4 family, proteins that act as essential regulators of vesicular transport pathways (Lazar *et al*., 1997; Novick & Zerial, 1997; Olkkonen & Stenmark, 1997; Schimmoller *et al*., 1998; Chavrier & Goud, 1999; Brennwald, 2000; Rodman & Wandinger-Ness, 2000). Rabs have been traditionally numbered in order of discovery, Rab1, Rab2, Rab3, through to Rab37 at present. The number of Rab-like sequences has been growing steadily over the last decade and we noted that a significant number of Rabs have been deposited in the databases under different names. The problem is that there is no comprehensive definition of what distinguishes a Rab from other small GTPases. Simple criteria such as the presence of a double-cysteine prenylation motif at the C terminus is insufficient, as some *bona fide* Rabs such as Rab8 or Rab13 contain only a single cysteine residue (Casey & Seabra, 1996).

Here, we attempt to identify mammalian Rab-specific regions that serve as diagnostic Rab sequences in order to define a Rab protein on the basis of its primary structure. We found five Rab-specific regions that we termed Rab family (RabF) motifs. Using the RabF motifs, we were able to identify new Rabs from the databases as well as find all the known Rabs, thus validating these

Abbreviations used: RabF, Rab family; RabSF, Rab subfamily; HMM, hidden Markov models.

E-mail address of the corresponding author: m.seabra@ic.ac.uk

0022-2836/00/041077–11 $35.00/0

motifs. We also analysed Rab subfamily sequence motifs (RabSF) and propose a new nomenclature for the family. These new findings suggest general mechanisms by which Rabs and possibly other families of Ras-like GTPases are able to bind such a myriad of regulators and effectors.

Results

Rab-specific sequences (RabF1 to RabF5)

Many previous studies have highlighted conserved regions in all members of the Ras superfamily that are involved in guanine and phosphate/Mg^{2+} binding (Barbacid, 1987; Bourne *et al*., 1990, 1991). These have been referred to as G for guanine (G1-G3) and PM for phosphate/Mg^{2+} (PM1-PM3) (Valencia *et al*., 1991). We first analysed these sequences. As predicted, we found that these GTP-binding regions are not useful to distinguish Rabs because they are extremely conserved between all Ras-like proteins. Also, the variations are not typical of one family. For example, the Rab9 PM1 motif GDGGVGKT is much closer to Rho protein PM1 sequences than Rab. This principle applies to the other PM/G motifs.

Rabs have been shown to be substrates for prenylation by the enzyme Rab geranylgeranyl transferase (Seabra *et al*., 1992). They present prenylation motifs distinct from the motifs found in Ras and Rho, substrates for the C*AAX* prenyltransferases, protein farnesyl transferase or protein geranylgeranyl transferase-I (Casey & Seabra, 1996). The Rab prenylation motifs consist of two C-terminal cysteine residues, found in one of the following combinations: *XXX*CC, *XX*CC*X*, *X*CC*XX*, CC*XXX* or *XX*C*X*C. The presence of the double-cysteine motif in the C terminus is in general a very good diagnostic of a Rab protein. However, some Rabs present a C*XXX* box, where only one cysteine residue is available for prenylation. Therefore, the double-cysteine prenylation motif may confirm that a given small GTPase is a Rab but its absence should not be used to prove otherwise.

To address the existence of Rab-specific sequences, we first aligned all the known mammalian Rabs using the ClustalW 1.80 algorithm. The alignment of the complete mammalian Rab family (gap opening and extension penalties, respectively, 10.00 and 0.20) can be found at http://www.med.ic.ac.uk/db/dbbm/rab_family.html.

Based on this alignment we followed two complementary approaches to describe a ''model Rab'' sequence. The first approach consisted in manually plotting the frequency of the most abundant amino acid for any given position in the alignment (Figure 1). We observed the existence of regions of high amino acid identity, some corresponding to the conserved GTP-binding motifs, and some cor-

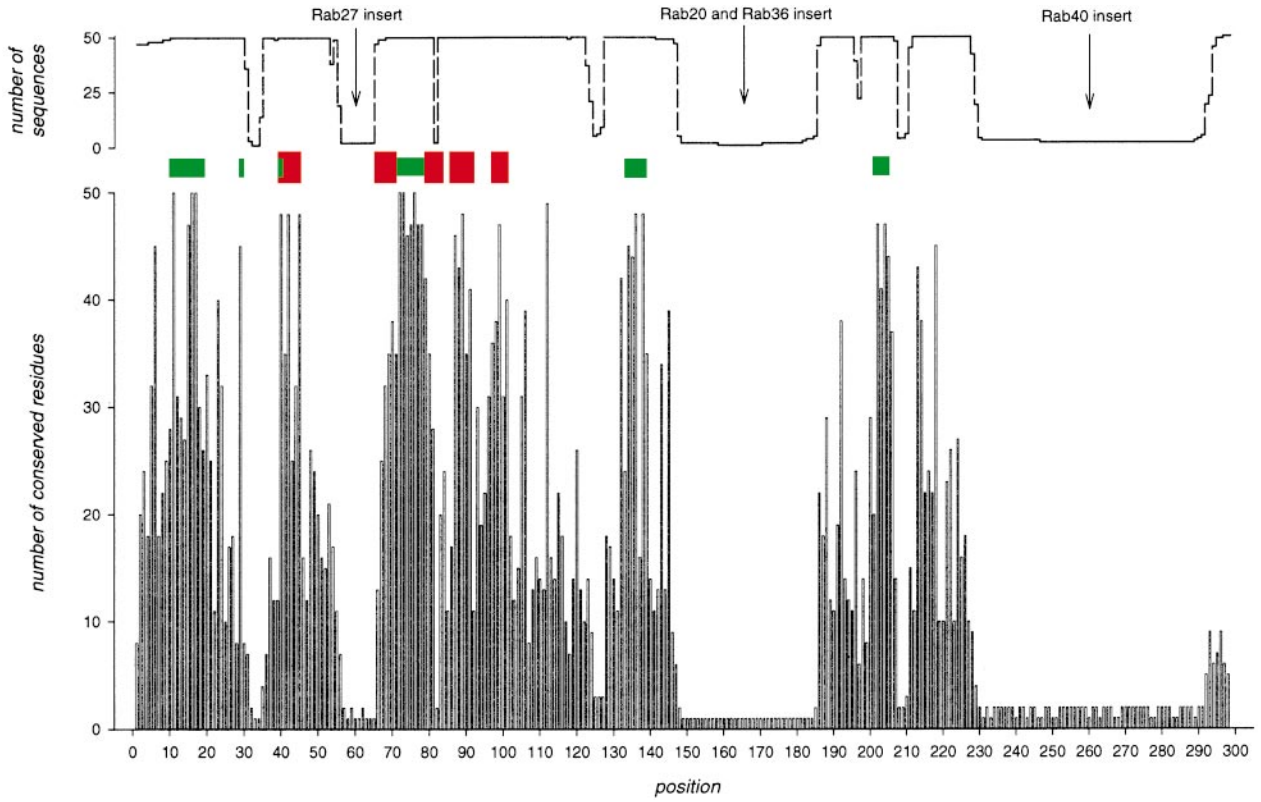

Figure 1. (a) Number of mammalian Rab sequences that align at a given position, highlighting the existence of some insert regions in specific Rabs. (b) Frequency of the most common amino acid in the Rab family alignment. Green and red boxes indicate the conserved GTP binding motifs and the RabF regions, respectively.

responding to other regions. In order to identify the Rab-specific regions, we defined variable length windows around the most promising areas, calculated the average amino acid identity for all the windows and compared it with the identity calculated, for the same window, for the entire Ras superfamily. We found five regions that appeared to be Rab specific (Figure 1, red squares).

The second approach was based on profile hidden Markov models (HMM) (Eddy, 1996). A statistical model of each family consensus was calculated using the software package HMMER 2.0. The aligned model sequences revealed positions conserved throughout the superfamily as well as candidate family-specific regions. Figure 2 summarises the results of the two strategies. The uppercase/lowercase code represents the results of the profile HMM, in which uppercase amino acids were found at $p > 0.5$. This analysis confirmed the existence of five conserved short stretches of residues that seem to be diagnostic for the Rab family that we numbered RabF1 to RabF5 (Figure 2). The RabF motifs include mostly Rab-specific positions but in some cases residues that are also highly conserved in other small GTPase families. These residues are also helpful in that they help rule out specific families, for example the G residue in RabF1 helps rule out Ras and Rho.

RabF1 localises to the so-called effector domain (loop 2 - β2), in the putative switch I region. The prototypical sequence is IGVDF (Figure 2). Olkkonen & Stenmark (1997) suggested that this region was Rab specific, but by itself this sequence is not large or specific enough to serve as the sole criterion to identify a Rab protein. The G position is almost absolutely conserved in Rabs, Arfs and Ran, and represents an insertion relatively to Ras and Rho proteins. The solution structure of Rab3a-GTP revealed that this residue plays an essential role in forcing the main-chain in the putative switch I region to bulge in the direction of the helix α2, promoting a closer interaction with the putative switch II region. This results in a more rigid conformation in the putative switch regions, apparently characteristic of the Rab family (Dumas *et al.*, 1999). While we have been referring to ''switch'' regions in Rabs, it is important to note that there is only one known crystal structure form of Rab to date. In the absence of a Rab-GDP structure, we have no information on the conformational changes between the two nucleotide-bound configurations in Rabs. Whilst it is almost certain that Rab switch regions exist, their precise boundaries as compared to those well defined in Ras are difficult to predict at present (Pai *et al.*, 1989; Tong

Figure 2. Alignment of profile HMM model sequences (manually adjusted to accommodate structural considerations). The uppercase/lower case coding represents the results of the profile HMM method, in which uppercase characters were found at $p > 0.5$. Residues found to be Rab specific are highlighted in red. When a position is conserved in other families, the corresponding position is also highlighted in red. Green characters denote the conserved nucleotide binding (PM/G) motifs.

et al., 1991), and hence it is more prudent to refer to ''putative'' switch regions.

All the other RabF regions cluster in and around the putative switch II region. The RabF2 prototypical sequence is KLQIW (β3) (Figure 2). The W position is not Rab specific, as it is conserved in all small GTPases except Ras. The RabF3 prototypical sequence is RFrsiT (loop 4) (Figure 2). In the Rab3a crystal structure, the hydrophobic residues in this motif and the adjacent hydrophobic T residue (corresponding to α2) reinforce the effect of the bulge in β2, contributing to the higher level of rigidity of the switch I/switch II interface (Dumas *et al.*, 1999). The RabF4 prototypical sequence YYRGA (α2-loop 5) is almost adjacent to RabF3 (Figure 2) followed closely by the RabF5 motif, LVYDIT (β4-loop 6). Therefore, RabF2 to RabF5 cluster between sheets β3 and β4 in a region that includes the putative switch II region.

New Rabs

Based on the criteria described, we found 52 sequences in Genbank that we consider Rabs. The submissions have been numbered in order of discovery, starting with Rab1a through Rab40. The gaps in the numbering are Rab16, which is Rab3d, and Rab31, which is Rab22b.

The newly identified Rabs are shown in Figure 3. Two ''old'' entries, called Rah (accession AAC83182) and Ray (accession AAD25874) are clearly Rabs. Both possess all the RabF motifs, either with the prototypical sequence or in variations that are recognisable as Rab specific, possess the unique double-cysteine prenylation motif and have overall average identities to other Rabs of 23.9 % and 36.3 %, respectively. We propose that they should be renamed Rab34 (rah) and Rab35 (ray) and on our recommendation, the latter has already been renamed Rab35.

Our analysis suggests that a recent entry called Rab36 is indeed a Rab despite being peculiar. Rab36 presents a 125-residue long N-terminal extension (counted from the conserved K residue on position 13 in hRab1a) (Mori *et al.*, 1999). Interestingly, one yeast protein known as Ypt11p or Ybj9 (Garcia-Ranea & Valencia, 1998; Lazar *et al.*, 1997) also possesses a large N-terminal extension. It is not clear whether this protein is the Rab36 ortholog as both proteins share only 10 % identity. No functional information is available on either protein other than the deletion of the yeast gene produces no phenotype.

Two ''Rab-related'' sequences are found in the databases, with accession numbers X99962 (Stankovic *et al.*, 1997) and AAA42000. The first protein we propose to be renamed Rab39. Its closest mammalian relative is hRab2 which is 41 % identical. We propose to rename Rab39 the Rab-like sequence with accession number AAA42000.

Several proteins have been named ''Rab-like''. The human protein Rab7L1 (accession number NP_03920) (Shimizu *et al.*, 1997) is 94 % identical with the protein Rab29 from rat, suggesting that Rab7L1 is the human ortholog of rat Rab29 and should be renamed accordingly. The proteins with

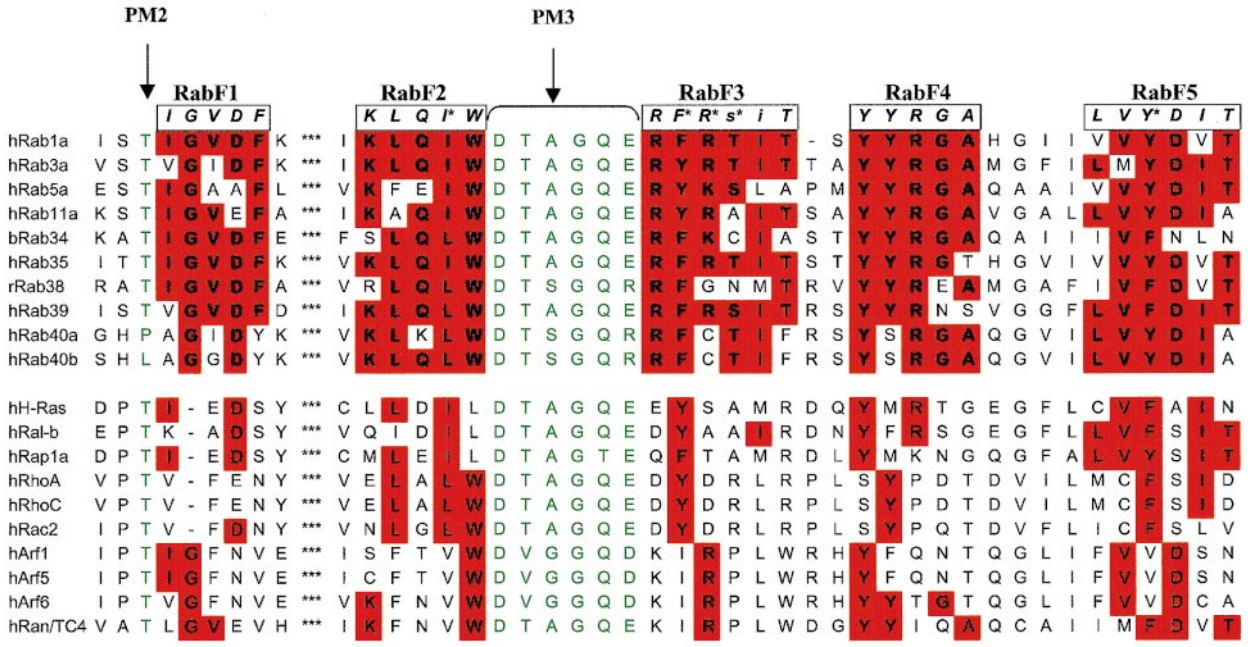

Figure 3. Alignment of RabF regions for novel and selected Rab proteins, and the corresponding regions in other representative small GTPases. Conserved residues are highlighted in red. The prototypical sequence for each of the RabF motifs is indicated on the top of the alignment, in italics. An asterisk indicates that the position is frequently occupied by a conservative substitution and the second most common occurrence is also highlighted in the alignment.

accession numbers NP_09012 and NP_09013 (RabL2B and RabL2A) (Wong *et al.*, 1999) do not present any prenylatable cysteine residues in the C terminus, have low overall identity to the rest of the family (average identity 21 % and 23 %, respectively) and have sequences that do not conform to the RabF motifs described above. They are clearly not Rab proteins but they are also not obvious members of another family. It is possible that RabL2 may represent a new family of small GTPases.

Using Ψ-BLAST of different Rabs, and searching the databases using combinations of the RabF motifs described above, we found two more mammalian sequences that we consider Rabs. The two related proteins CAB09136 and AAA17031 (originally called Rar) are more related to Rabs (27.5 % average identity to the Rab family) than to any other small GTPase family. Both possess recognisable RabF motifs and none of the sequences present structural motifs that suggest they belong to any of the other small GTPase families. We propose that they should be renamed Rab40a and Rab40b, due to their high degree of identity to each other. Both proteins present unusual substitutions at the conserved PM2 position. This residue, typically a threonine is replaced by proline or alanine. The role of the threonine residue has been extensively studied and shown to be involved in coordination of the Mg^{2+} and in stabilisation of the γ-phosphate (Valencia *et al.*, 1991). Its absence in the Rab40 subfamily may imply that these proteins do not cycle between two conformations, but are permanently locked in an inactive state. A precedent for a Rab that appears not to cycle efficiently between two conformations was set by Rab24. Rab24 appears to be locked in the GTP bound conformation due to variations in two PM/G motifs (Erdman *et al.*, 2000).

Rabs subfamilies and subfamily-specific sequences (RabSF)

Phylogenetic analysis of the complete mammalian Rab family using the neighbour joining algorithm (disregarding the gaps in the alignment caused by some Rabs having specific ''insert regions''), revealed several clusters of ''related'' Rabs (Figure 4). Within these clusters there are Rabs that show unusually high homology and are termed isoforms, defining Rab subfamilies. The problem is where to draw the line between isoforms (named Rab1a and Rab1b, for example) and simply ''related'' Rabs. Since isoforms are believed to be functionally related and are thought to interact with the same type of effectors (Lazar *et al.*, 1997; Novick & Zerial, 1997; Olkkonen & Stenmark, 1997; Schimmoller *et al.*, 1998; Chavrier & Goud, 1999; Brennwald, 2000; Rodman & Wandinger-Ness, 2000), it is plausible that this functional conservation reflects conservation in specific regions, as opposed to complete sequence conservation. If such regions exist they are the best criteria to decide if two or more Rabs are isoforms.

Moore and co-workers (Moore *et al.*, 1995) observed that there was high-level amino acid conservation within subfamilies in three regions designated here RabSF2, RabSF3 and RabSF4 and corresponding to α1/loop2, α3/loop7 and α5, respectively (Figure 2). We verified the corresponding regions in all our mammalian Rab sequences and in general we confirmed that these regions show higher identity within the sub-families than the overall sequence (Table 1). This is particularly remarkable with RabSF4, which is located in the so-called hypervariable domain, a region characterised by its sequence divergence: the average identity of this region among Rabs is 14.4 % while subfamilies show an average of 58.4 %.

Recently, the 3D structure of Rab3a in complex with its effector Rabphilin-3A revealed that three regions in Rab3a contribute to form a ''pocket'' that mediates binding to this and possibly other effectors (Ostermeier & Brunger, 1999). These three regions were named RabCDRs (Rab complementary-determining region) and we refer to them as RabCDRI to RabCDRIII, counting from the N terminus of the protein. Interestingly, these regions correspond to subfamily-specific sequences: RabCDRII and RabCDRIII correspond to RabSF3 and

Table 1. Amino acid identity within the different Rab sub-family specific regions (RabSF)

	Complete sequence (%)	RabSF1 (%)	RabSF2 (%)	RabSF3 (%)	RabSF4 (%)
Rab family average identity	**32.7**	**34.2**	**28.6**	**23**	**14.4**
Subfamilies	**78.3**	**92.3**	**90.1**	**87.5**	**58.4**
1	91	100	100	96	84
3	73-82	100	83-100	88-96	46-76
4	82	100	88	76	84
5	81-86	85-199	94-100	88-96	69-76
6	90	100	100	95	53
8	82	100	88	88	30
11	90	100	100	100	84
22	70	85	72	76	23
27	71	100	72	84	76
40	88	100	72	86	54

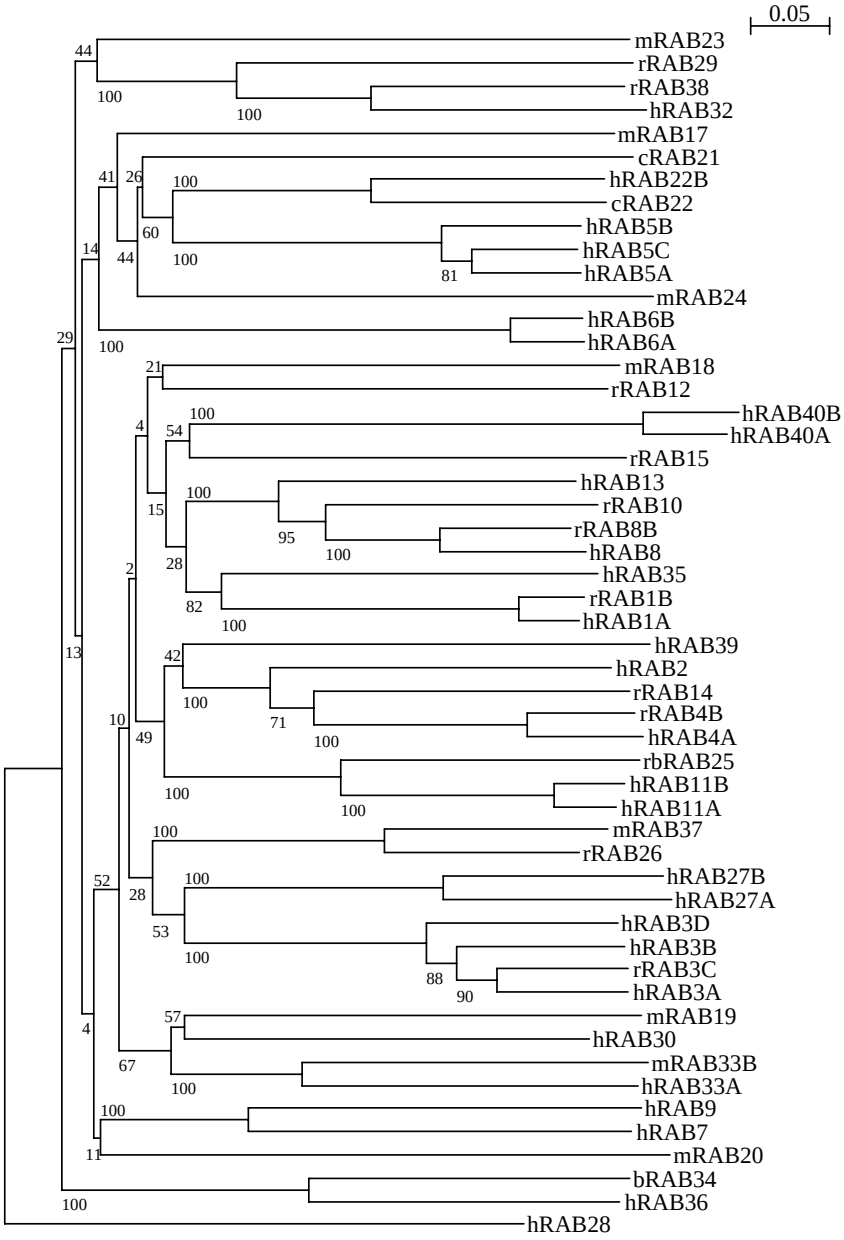

Figure 4. Phylogenetic tree of the 52 members of the mammalian Rab family, calculated using the neighbour-joining method and excluding the gaps. Bootstraping involved 1000 trials and is represented as percentage in the Figure.

RabSF4. RabCDRI includes the N-terminal sequence upstream of PM1 (YXYLFK) previously proposed to be diagnostic of the Rab family (Sanford *et al*., 1995). However, the overall identity within the region is only 34 % and one residue (K13 in hRab1a) is highly conserved throughout the superfamily. We propose instead that this sequence is a good indicator of a Rab subfamily and termed it RabSF1 (Figure 2).

We mapped both the RabF and RabSF regions into the crystal structure of Rab3a (Figure 5). We observed that Rabs present two subfamily-specific surfaces: RabSF1, RabSF3 and RabSF4 form a surface that mediates specific interactions between Rab3a and Rabphilin (Ostermeier & Brunger, 1999). Almost on the opposite side of this surface RabSF2 forms a second subfamily-specific surface near or within the switch I region that could mediate interaction with other effectors. The existence of these two distinct subfamily-specific surfaces suggests that different effectors will bind different RabSF regions. Taken together, this analysis is consistent with the hypothesis that isoforms will interact with the same type of effectors/auxiliary molecules *via* subfamily specific regions.

Criteria in addition to sequence homology at the RabSF regions need to be taken into consideration when defining a Rab subfamily. These include specific variants of PM, G or RabF motifs and small characteristic diversions from the consensus. For example, only Rab27 isoforms present the RabF4 variant FFRDA and have a ten-residue insertion in loop3, supporting the definition of the Rab27 subfamily. These new criteria allowed the definition of one more subfamily, Rab40a and Rab40b (Table 1). The overall sequence identity is 88 %. Both sequences present a high identity within all the RabSF regions as well as all the PM/G and

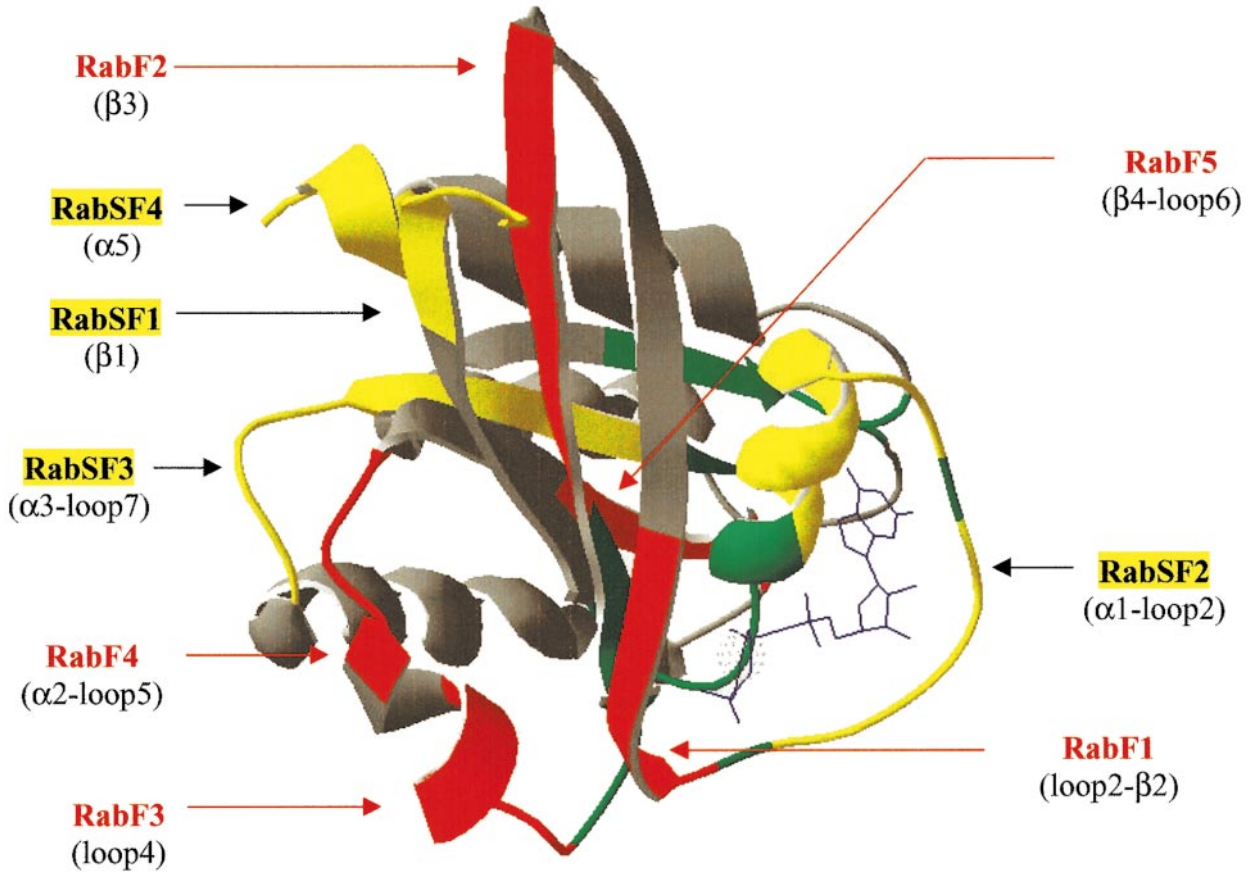

Figure 5. Cartoon representing the Rab3a-GTP 3D structure (PDB code 3RABA). RabF regions are depicted in red, RabSF regions in yellow, and the conserved nucleotide binding (PM/G) motifs in green. The nucleotide and the Mg^{2+} are both represented in blue.

RabF motifs, and both present a long C-terminal insertion.

A number of related Rabs present a problem for classification, in that they are clearly related but should they be considered isoforms? These related Rabs include Rab4a/b and Rab14, Rab8a/b and Rab10, Rab11a/b and Rab25, Rab32 and Rab39, Rab33a and Rab33b (Figure 4). Considering that in all cases the overall identity is lower than 70% and that there is no consistent conservation in more than two out of the four RabSF motifs, we suggest that they should not be considered isoforms (Table 1).

Discussion

Here, we identified five Rab-specific motifs, named RabF1 to RabF5, that in conjunction with the conserved PM/G motifs and a double-cysteine C-terminal prenylation motif allow the definition of a Rab GTPase. As of June 2000, we found 52 sequences in the database that fit our criteria. Unfortunately, it is impossible to use the information in the EST database to predict whether this number is close to the real number of genes in a mammalian genome as only full-length (or near full-length) sequences are useful in this regard. Despite this problem, it is clear that the mammalian Rab family is much larger than the 11-member orthologous gene family in *Saccharomyces cerevisiae* (Garcia-Ranea & Valencia, 1998; Lazar *et al.*, 1997), possibly reflecting the higher intracellular specialisation of mammalian cells. In order to classify a Rab as an isoform, we propose that the sequences must be at least 70% identical and this value should be supported by conservation at the RabSF and RabF motifs as well as unique characteristics. Currently, we recognise ten Rab subfamilies (Rab1, 3, 4, 5, 6, 8, 11, 22, 27 and 40).

When we mapped these regions into the crystal structure of Rab3a, we noted surprisingly that all RabF motifs localise in and around the switch I and switch II regions. The switch regions change conformation significantly upon GTP binding and hydrolysis and therefore mediate interaction with effectors and regulators. If the switch regions are conserved among all Rabs, how are Rabs able to bind specific effectors and regulators, given that these proteins must discriminate between the two nucleotide-bound states of Rabs?

One possibility is that effectors and regulators will bind both RabF regions to discriminate between active/inactive conformations and RabSF regions for specificity. The existence of two distinct

subfamily-specific surfaces further suggests that different effectors/regulators bind to different combinations of RabSF regions. Different lines of evidence support this model. First and foremost, the recent solution structure of the Rab3a-Rabphilin complex (Ostermeier & Brunger, 1999) revealed that the binding surface of this complex involves both switch regions and a Rab subfamily-specific surface composed by the regions we defined herein as RabSF1, RabSF3 and RabSF4. Second, both switch regions and the region corresponding to helix α3/loop 7 (RabSF3) have been shown by site-directed mutagenesis to be involved in interaction with regulators such as GEFs and GAPs (Becker *et al*., 1991; Burstein *et al*., 1992; Brondyk *et al*., 1993; Day *et al*., 1998; McKiernan *et al*., 1993). Third, helix 3/loop 7 (RabSF3), the effector domain (RabF1) and the hypervariable domain (including RabSF4) are key regions in determining both localisation and function of ypt1/sec4 chimeras (Brennwald & Novick, 1993; Dunn *et al*., 1993), Rab2/Rab5/Rab7 chimeras (Chavrier *et al*., 1991) and Rab5/Rab6 chimeras (Stenmark *et al*., 1994). Fourth, the N terminus (RabSF1) and helix 2/loop5 (RabF4) have also been shown to be essential in producing functional Rab5/Rab6 chimeras (Stenmark *et al*., 1994).

Another prediction of this model is that the binding of general regulators such as Rab escort protein (REP) and Rab GDP dissociation inhibitor (RabGDI) is nucleotide sensitive and occurs *via* the newly identified RabF regions. There is some evidence to support this hypothesis (Beranger *et al*., 1994; Burstein *et al*., 1992; Overmeyer *et al*., 1998; Wilson & Maltese, 1993; Wilson *et al*., 1996). It is also likely that binding of general regulators and specific regulators/effectors is mutually exclusive, since both are able to sense the two alternative nucleotide-bound conformations and thus must interact with the switch regions.

We questioned if this could be a general model that applies to other Ras-like GTPases. The answer is generally yes, but not in all cases. The most notorious exception is with Ras proteins. In Ras, the switch regions are also quite conserved (Figure 2) but they appear to be the only contact point in some interactions such as with RalGDS (Vetter *et al*., 1999c) and the Ras binding domain of Raf1 (Nassar *et al*., 1995). The strategy to determine specificity seems to be that distinct Ras effectors bind different subsets of residues within the switch regions. This is particularly well characterised for the switch I region (also called effector loop) (White *et al*., 1995). In other cases, helix α3 (corresponding roughly to RabSF3) is also involved in defining the binding surface, as with the exchange factor SOS (Boriack-Sjodin *et al*., 1998) and p120GAP (Scheffzek *et al*., 1997). Similar conclusions may be drawn from the limited data available for the Arf family, where interaction with effectors and regulators appear in most cases to be restricted to the switch regions and helix α3 (Beraud-Dufour *et al*., 1998; Goldberg, 1998, 1999; Mossessova *et al*., 1998).

The picture is more complex in the Rho family, probably because of the existence of a staggering number of Rho regulators/effectors. The switch regions are involved in binding regulators/effectors in every case we found, but other regions have been shown to be involved. RhoA and Cdc42 use switch I and/or switch II in combination with helix α5 (reminiscent of RabSF4) to bind the effector domain of PKN/PRK1 (Maesaki *et al*., 1999), Wiskott-Aldrich syndrome protein (WASP) (Abdul-Manan *et al*., 1999) and ACK (Mott *et al*., 1999). On the other hand, the binding of Cdc42 to Cdc42GAP appears to involve mainly residues in the switch regions (Nassar *et al*., 1998). As in Ras, different residues in the switch regions appear to confer specificity to the interaction in some cases. Examples for this type of interaction are RhoA/Cdc42 to Lbc/Cdc24 (Li & Zheng, 1997) and RhoA/Cdc42 to PLD (Bae *et al*., 1998). In other cases, it seems that the interaction requires both the switch regions (possibly RhoF regions) and other regions (possibly RhoSF regions) such as helix α3 (Zhong *et al*., 1999), helix α5 (Abdul-Manan *et al*., 1999; Maesaki *et al*., 1999; Mott *et al*., 1999), and the specific Rho insert region (Freeman *et al*., 1996).

As for Ran, structural and biochemical data also implicate the switch regions and the helix α3 in binding effectors, such as binding to RCC1 and NTF2 (Azuma *et al*., 1999; Stewart *et al*., 1998) but other regions also contribute to binding, such as the α4/β6/α5 region with Importin β (Vetter *et al*., 1999a) and the C-terminal helix α6 with RanBD1 (Vetter *et al*., 1999b).

Overall, the available data suggest the interesting possibility that primitive interactions between Ras-like GTPases and effectors/regulators depended upon the switch regions. As the range and complexity of cellular mechanisms controlled by these GTPases grew, so did the interaction surfaces between them and the increasing number of specific effectors/regulators. The future analysis of the Ras, Rho and Arf families as presented here for the Rab family may help clarify this hypothesis.

Materials and Methods

We selected a mammalian representative of each Rab sequence from the databases. Where more than one sequence was available, we chose the following arbitrary priority of species: human, rat, mouse, canine, bovine and rabbit. We performed alignments of the sequences retrieved from Genbank using the CLUSTAL W 1.80 algorithm (Thompson *et al*., 1994) and the multiple sequence alignment with hierarchical clustering algorithm (Corpet, 1988) (the complete sequence alignment data can be viewed at http://www.med.ic.ac.uk/db/dbbm/rab_family.html). The software package HMMER 2.0 (available at http://hmmer.wustl.edu/) was used to calculate profile HMM and generate model sequences for each family. The alignment of model sequences was

manually adjusted to accommodate structural considerations.

The search for new mammalian Rab sequences was performed by Ψ-BLAST (Altschul *et al*., 1997) of known Rab sequences, or by searching the databases with combinations of the Rab family-specific motifs (RabF) defined in this work, using the PatternFind Server at http://www.isrec.isb-sib.ch/software/PATFND_form.html

The final alignments were used to calculate phylogenetic trees using the neighbour-joining method excluding positions with gaps. Bootstraping (Felsenstein, 1985) involved 1000 trials. Trees were calculated using the program CLUSTALX 1.80 (Thompson *et al*., 1997) and plotted using the program Njplot (Perriere & Gouy, 1996). PDB files were retrieved from the Protein Data Bank and the mapping of the different regions identified in this work done using the Swiss-PDBviewer v3.51 program available on http://www.expasy.ch/spdbv/mainpage.htm (Guex & Peitsch, 1996).

Acknowledgements

We thank Alfonso Valencia for critical reading of the manuscript, Alistair Hume and James Bellingham for helpful discussions. This work was supported in part by the Human Frontier of Science Program and the Wellcome Trust. J.P.-L. is a student of the Programa Gulbenkian de Doutoramento em Biologia e Medicina supported by grant PRAXIS XXI BD 9809/96. M.C.S. is a Pew Scholar in the Biomedical Sciences.

References

Abdul-Manan, N., Aghazadeh, B., Liu, G. A., Majumder, A., Ouerfelli, O., Siminovitch, K. A. & Rosen, M. K. (1999). Structure of Cdc42 in complex with the GTPase-binding domain of the ''Wiskott-Aldrich syndrome'' protein. *Nature,* **399**, 379-383.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Azuma, Y., Renault, L., Garcia-Ranea, J. A., Valencia, A., Nishimoto, T. & Wittinghofer, A. (1999). Model of the Ran-RCC1 interaction using biochemical and docking experiments. *J. Mol. Biol.* **289**, 1119-1130.

Bae, C. D., Min, D. S., Fleming, I. N. & Exton, J. H. (1998). Determination of interaction sites on the small G protein RhoA for phospholipase D. *J. Biol. Chem.* **273**, 11596-11604.

Barbacid, M. (1987). Ras genes. *Annu. Rev. Biochem.* **56**, 779-827.

Becker, J., Tan, T. J., Trepte, H. H. & Gallwitz, D. (1991). Mutational analysis of the putative effector domain of the GTP-binding Ypt1 protein in yeast suggests specific regulation by a novel GAP activity. *EMBO J.* **10**, 785-792.

Beranger, F., Cadwallader, K., Porfiri, E., Powers, S., Evans, T., de Gunzburg, J. & Hancock, J. F. (1994). Determination of structural requirements for the interaction of Rab6 with RabGDI and Rab geranylgeranyltransferase. *J. Biol. Chem.* **269**, 13637-13643.

Béraud-Dufour, S., Robineau, S., Chardin, P., Paris, S., Chabre, M., Cherfils, J. & Antonny, B. (1998). A glutamic finger in the guanine nucleotide exchange factor ARNO displaces Mg^{2+} and the β-phosphate to destabilize GDP on ARF1. *EMBO J.* **17**, 3651-3659.

Boriack-Sjodin, P. A., Margarit, S. M., Bar-Sagi, D. & Kuriyan, J. (1998). The structural basis of the activation of Ras by Sos. *Nature,* **394**, 337-343.

Bourne, H. R., Sanders, D. A. & McCormick, F. (1990). The GTPase superfamily: a conserved switch for diverse cell functions. *Nature,* **348**, 125-132.

Bourne, H. R., Sanders, D. A. & McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature,* **349**, 117-127.

Brennwald, P. (2000). Reversal of fortune: do Rab GTPases act on the target membrane? *J. Cell Biol.* **149**, 1-3.

Brennwald, P. & Novick, P. (1993). Interactions of three domains distinguishing the Ras-related GTP-binding proteins Ypt1 and Sec4. *Nature,* **362**, 560-563.

Brondyk, W. H., McKiernan, C. J., Burstein, E. S. & Macara, I. G. (1993). Mutants of Rab3A analogous to oncogenic Ras mutants. Sensitivity to Rab3A-GTPase activating protein and Rab3A-guanine nucleotide releasing factor. *J. Biol. Chem.* **268**, 9410-9415.

Burstein, E. S., Brondyk, W. H. & Macara, I. G. (1992). Amino acid residues in the Ras-like GTPase Rab3A that specify sensitivity to factors that regulate the GTP/GDP cycling of Rab3A. *J. Biol. Chem.* **267**, 22715-22718.

Casey, P. J. & Seabra, M. C. (1996). Protein prenyltransferases. *J. Biol. Chem.* **271**, 5289-5292.

Chavrier, P. & Goud, B. (1999). The role of Arf and Rab GTPases in membrane transport. *Curr. Opin. Cell Biol.* **11**, 466-475.

Chavrier, P., Gorvel, J.-P., Stelzer, E., Simons, K., Gruenberg, J. & Zerial, M. (1991). Hypervariable C-terminal domain of rab proteins acts as a targeting signal. *Nature,* **353**, 769-762.

Corpet, F. (1988). Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881-10890.

Day, G-J., Mosteller, R. D. & Broek, D. (1998). Distinct subclasses of small GTPases interact with guanine nucleotide exchange factors in a similar manner. *Mol. Cell Biol.* **18**, 7444-7454.

Dumas, J. J., Zhu, Z., Connolly, J. L. & Lambright, D. G. (1999). Structural basis of activation and GTP hydrolysis in Rab proteins. *Structure,* **7**, 413-423.

Dunn, B., Stearns, T. & Botstein, D. (1993). Specificity domains distinguish the Ras-related GTPases Ypt1 and Sec 4. *Nature,* **362**, 563-565.

Eddy, S. R. (1996). Hidden Markov models. *Curr. Opin. Struct. Biol.* **6**, 361-365.

Erdman, R. A., Schellenbergen, K. E., Overmeyer, J. H. & Maltese, W. A. (2000). Rab24 is an atypical member of the Rab GTPase family. *J. Biol. Chem.* **275**, 3848-3856.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution,* **39**, 783-791.

Freeman, J. L., Abo, A. & Lambeth, J. D. (1996). Rac ''insert region'' is a novel effector region that is implicated in the activation of NADPH oxidase, but not PAK65. *J. Biol. Chem.* **271**, 19794-19801.

Garcia-Ranea, J. A. & Valencia, A. (1998). Distribution and functional diversification of the ras superfamily in *Saccharomyces cerevisiae*. *FEBS Letters,* **434**, 219-225.

Goldberg, J. (1998). Structural basis for activation of ARF GTPase: mechanisms of guanine nucleotide exchange and GTP-myristoyl switching. *Cell,* **95**, 237-248.

Goldberg, J. (1999). Structural and functional analysis of the ARF1-ARFGAP complex reveals a role for coatomer in GTP hydrolysis. *Cell,* **96**, 893-902.

Guex, N. & Peitsch, M. C. (1996). SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis,* **18**, 2714-2723.

Kahn, R. A., Der, C. J. & Bokoch, G. M. (1992). The Ras superfamily of GTP-binding proteins: guidelines on nomenclature. *FASEB J.* **6**, 2512-2513.

Lazar, T., Gotte, M. & Gallwitz, D. (1997). Vesicular transport: how many Ypt/Rab-GTPases make a eukaryotic cell? *Trends Biochem. Sci.* **22**, 468-472.

Li, R. & Zheng, Y. (1997). Residues of the Rho family GTPases Rho and Cdc42 that specify sensitivity to Dbl-like guanine nucleotide exchange factors. *J. Biol. Chem.* **272**, 4671-4679.

Maesaki, R., Ihara, K., Shimizu, T., Kuroda, S., Kaibuchi, K. & Hakoshima, T. (1999). The structural basis of Rho effector recognition revealed by the crystal structure of human RhoA complexed with the effector domain of PKN/PRK1. *Mol. Cell,* **4**, 793-803.

McKiernan, C. J., Brondyk, W. H. & Macara, I. G. (1993). The Rab3A GTPase interacts with multiple factors through the same effector domain. *J. Biol. Chem.* **268**, 24449-24452.

Moore, I., Schell, J. & Palme, K. (1995). Subclass-specific sequence motifs identified in Rab GTPases. *Trends Biochem. Sci.* **20**, 10-12.

Mori, T., Fukuda, Y., Kuroda, H., Matsumura, T., Ota, S., Sugimoto, T., Nakamura, Y. & Inazawa, J. (1999). Cloning and characterization of a novel Rab-family gene, Rab36, within the region at 22q11. 2 that is homozygously deleted in malignant rhabdoid tumors. *Biochem. Biophys. Res. Commun.* **254**, 594-600.

Mossessova, E., Gulbis, J. M. & Goldberg, J. (1998). Structure of the guanine nucleotide exchange factor Sec7 domain of human Arno and analysis of the interaction with ARF GTPase. *Cell,* **92**, 415-423.

Mott, H. R., Owen, D., Nietlispach, D., Lowe, P. N., Manser, E., Lim, L. & Laue, E. D. (1999). Structure of the small G protein Cdc42 bound to the GTPase-binding domain of ACK. *Nature,* **399**, 384-388.

Nassar, N., Horn, G., Herrmann, C., Scherer, A., McCormick, F. & Wittinghofer, A. (1995). The 2.2 Å crystal structure of the Ras-binding domain of the serine/threonine kinase c-Raf1 in complex with Rap1A and a GTP analogue. *Nature,* **375**, 554-560.

Nassar, N., Hoffman, G. R., Manor, D., Clardy, J. C. & Cerione, R. A. (1998). Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP. *Nature Struct. Biol.* **5**, 1047-1052.

Novick, P. & Zerial, M. (1997). The diversity of Rab proteins in vesicle transport. *Curr. Opin. Cell Biol.* **9**, 496-504.

Olkkonen, V. M. & Stenmark, H. (1997). Role of Rab GTPases in membrane traffic. *Int. Rev. Cytol.* **176**, 1-85.

Ostermeier, C. & Brunger, A. T. (1999). Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3A. *Cell,* **96**, 363-374.

Overmeyer, J. H., Wilson, A. L., Erdman, R. A. & Maltese, W. A. (1998). The putative ''switch 2'' domain of the Ras-related GTPase, Rab1B, plays an essential role in the interaction with Rab escort protein. *Mol. Biol. Cell,* **9**, 223-235.

Pai, E. F., Kabsch, W., Krengel, U., Holmes, K. C., John, J. & Wittinghofer, A. (1989). Structure of the guanine-nucleotide-binding domain of the Ha-ras oncogene product p21 in the triphosphate conformation. *Nature,* **341**, 209-214.

Perriere, G. & Gouy, M. (1996). WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie,* **78**, 364-369.

Rodman, J. S. & Wandinger-Ness, A. (2000). RabGTPases coordinate endocytosis. *J. Cell Sci.* **113**, 183-192.

Sanford, J. C., Pan, Y. & Wessling-Resnick, M. (1995). Properties of Rab5 N-terminal domain dictate prenylation of C-terminal cysteines. *Mol. Biol. Cell,* **6**, 71-85.

Scheffzek, K., Reza, A., Kabsch, W., Wiesmüller, L., Lautwein, A., Schmitz, F. & Wittinghofer, A. (1997). The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic ras mutants. *Science,* **277**, 333-338.

Schimmoller, F., Simon, I. & Pfeffer, S. R. (1998). Rab GTPases, directors of vesicle docking. *J. Biol. Chem.* **273**, 22161-22164.

Seabra, M. C., Goldstein, J. L., Sudhof, T. C. & Brown, M. S. (1992). Rab geranylgeranyl transferase: a multi-subunit enzyme that prenylates GTP-binding proteins terminating in Cys-X-Cys or Cys-Cys. *J. Biol. Chem.* **267**, 14497-14503.

Shimizu, F., Katagiri, T., Suzuki, M., Watanabe, T. K., Okuno, S., Kuga, Y., Nagata, M., Fujiwara, T., Nakamura, Y. & Takahashi, E. (1997). Cloning and chromosome assignment to 1q32 of a human cDNA (RAB7L1) encoding a small GTP-binding protein, a member of the RAS superfamily. *Cytogenet. Cell Genet.* **77**, 261-263.

Stankovic, T., Byrd, P. J., Cooper, P. R., McConville, C. M., Munroe, D. J., Riley, J. H., Watts, G. D., Ambrose, H., McGuire, G., Smith, A. D., Sutcliffe, A., Mills, T. & Taylor, A. M. (1997). Construction of a transcription map around the gene for ataxia telangiectasia: identification of at least four novel genes. *Genomics,* **40**, 267-276.

Stenmark, H., Valencia, A., Martinez, O., Ullrich, O., Goud, B. & Zerial, M. (1994). Distinct structural elements of rab5 define its functional specificity. *EMBO J.* **13**, 575-583.

Stewart, M., Kent, H. M. & McCoy, A. J. (1998). Structural basis for molecular recognition between nuclear transport factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase Ran. *J. Mol. Biol.* **277**, 635-646.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673-4680.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucl. Acids Res.* **25**, 4876-4882.

Tong, L. A., de Vos, A. M., Milburn, M. V. & Kim, S. H. (1991). Crystal structures at 2.2 Å resolution of the catalytic domains of normal ras protein and an oncogenic mutant complexed with GDP. *J. Mol. Biol.* **217**, 503-516.

Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991). The ras protein family: evolutionary tree and role of conserved amino acids. *Biochemistry,* **30**, 4637-4648.

Vetter, I. R., Arndt, A., Kutay, U., Gorlich, D. & Wittinghofer, A. (1999a). Structural view of the Ran-Importin beta interaction at 2.3 Å resolution. *Cell,* **97**, 635-646.

Vetter, I. R., Nowak, C., Nishimoto, T., Kuhlmann, J. & Wittinghofer, A. (1999b). Structure of a Ran-binding domain complexed with Ran bound to a GTP analogue: implications for nuclear transport. *Nature,* **398**, 39-46.

Vetter, I. R., Linnemann, T., Wohlgemuth, S., Geyer, M., Kalbitzer, H. R., Herrmann, C. & Wittinghofer, A. (1999c). Structural and biochemical analysis of Ras-effector signaling *via* RalGDS. *FEBS Letters,* **451**, 175-180.

White, M. A., Nicolette, C., Minden, A., Polverino, A., Van Aeist, L., Marin, M. & Wigler, M. H. (1995). Multiple ras functions can contribute to mammalian cell transformation. *Cell,* **80**, 533-541.

Wilson, A. L. & Maltese, W. A. (1993). Isoprenylation of Rab1B is impaired by mutations in its effector domain. *J. Biol. Chem.* **268**, 14561-14564.

Wilson, A. L., Erdman, R. A. & Maltese, W. A. (1996). Association of Rab1B with GDP-dissociation inhibitor (GDI) is required for recycling but not initial membrane targeting of the Rab protein. *J. Biol. Chem.* **271**, 10932-10940.

Wong, A. C., Shkolny, D., Dorman, A., Willingham, D., Roe, B. A. & McDermid, H. E. (1999). Two novel human RAB genes with near identical sequence each map to a telomere-associated region: the subtelomeric region of 22q13.3 and the ancestral telomere band 2q13. *Genomics,* **59**, 326-334.

Zhong, H., Raman, N., Mickleton-Young, L. A., Atkinson, S. J. & Quilliam, L. A. (1999). Loop 6 of RhoA confers specificity for effector binding, stress fiber formation, and cellular transformation. *J. Biol. Chem.* **274**, 4551-4560.

Edited by M. Yaniv

(Received 9 March 2000; received in revised form 15 June 2000; accepted 3 July 2000)