Conformationally variable Rab protein surface regions mapped by limited proteolysis and homology modelling

Lydia NIKOLOVA^{*1}, Kizhake SOMAN^{†1,2}, Jeffry C. NICHOLS[‡], D. Sundarsingh DANIEL^{*}, Burton F. DICKEY^{*†} and Simon HOFFENBERG^{*†3}

^{*}Department of Medicine, Houston VA Medical Center, 2002 Holcombe Blvd., Houston, TX 77030, U.S.A., [†]Department of Molecular Physiology and Biophysics, Baylor College of Medicine, Baylor Plaza, Houston, TX 77030, U.S.A., and [‡]Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005, U.S.A.

Tryptic proteolysis of the small GTPases Rab4 and Rab5 is a multi-step, nucleotide-dependent process. Using N-terminal peptide sequencing, matrix-assisted laser desorption/ionization–time-of-flight MS and molecular modelling, we identified the three initial sites of proteolysis in Rab5 as Arg-4, Arg-81 and Arg-197. Arg-4 and Arg-81 lie within regions previously implicated in Rab5 endocytic function, and Arg-197 lies in a region involved in membrane targeting. Topologically, Arg-81 lies within the conformationally variable Switch II region shown to be important for protein–protein interactions of other GTPases. Homology modelling studies on Rab5 indicate that the Arg-81 side chain is buried in the Rab5 GTP conformation, but is solvent-accessible in the GDP conformation, explaining the dependence of proteolysis on nucleotides. Peptide mapping of Rab4 was performed to take advantage of additional scissile

bonds within Switch II to determine more precisely the limits of the nucleotide-dependent protease-accessible region. The Rab4 cleavage sites corresponded to Arg-81 and Pro-87 of Rab5, and taken together with the finding that Rab5 was not cleaved at Arg-91 this analysis defines an eight-residue surface-exposed conformationally variable region lying in the centre of Switch II. A sequence comparison of Rab proteins shows these eight residues to have a loosely conserved motif that we term Switch II(v) for its relative variability. C-terminal to Switch II(v) is a highly conserved Rab-specific YYRGA motif that we term Switch II(c) for its constant sequence. N-terminal to Switch II(v) is a sequence-invariant G-domain involved in nucleotide binding and hydrolysis. We propose that the Rab Switch II(v) region imparts specificity to nucleotide-dependent protein–protein interactions.

INTRODUCTION

Rab proteins comprise a large family of Ras-related GTPases that are ubiquitous regulators of intracellular vesicle transport processes. The precise molecular functions of Rab proteins remain elusive but their significance is indicated by the severe transport defects that can result from mutations in Rab genes [1–3]. An important aspect of understanding the function of any protein is a knowledge of the full complement of proteins with which it interacts, and of the physical and biochemical nature of the interactions. Several proteins that interact with many, and possibly all, members of the Rab family have been identified. These include Rab escort protein [4], which associates with nascent Rab proteins and presents them to geranylgeranyl transferase II for C-terminal prenylation [5]; Mss4, which seems to act as a chaperone for several guanine nucleotide-free Rab proteins [6]; and GDP dissociation inhibitor, which extracts Rab proteins from the target membranes of transport reactions for recycling to the donor membranes [7].

In addition to these proteins that interact with multiple Rab family members, proteins specific for individual Rabs have also been identified. Among these are Rabaptin-5, which seems to be an effector of Rab5 [8], and a Rab5 GDP exchange factor [9]. Rab5 has emerged as a relatively well characterized model Rab protein because of a knowledge of these macromolecular interactions, and the extensive biochemical analysis of its GTPase

cycle [10–12] and functional analysis of its role in endocytosis [13,14]. However, only limited data are available on the regions of Rab5 that interact with other proteins. These include the identification of five regions that confer Rab5 function on Rab6/Rab5 chimaeras: the N-terminus, loop 2, helix α 2/loop 5, helix α 3/loop 7, and the C-terminus [15]. The N-terminus has been further implicated in Rab5 function by limited proteolysis and peptide competition [16]. Because the interactions of most proteins with Rab5 are nucleotide-dependent, it can be assumed that Rab5 surface regions whose conformations are nucleotide-dependent participate in these interactions. All regulatory GTPases that have been crystallized with both GDP and GTP analogues bound to them show marked nucleotide-dependent conformational variability in two regions termed Switch I (residues 30–38 in Ras) and Switch II (residues 60–76) [17–20]. Switch I overlaps the ‘effector domain’, initially defined for Ras by point mutations that disrupt its effector functions. Switch II residues also interact with Ras effector proteins, as well as interacting with exchange factors [21].

In the present study we used limited proteolysis coupled with peptide sequencing and MS to identify Rab5 surface regions that are conformationally variable in a nucleotide-dependent fashion. Similar studies were also performed with Rab4 to take advantage of scissile sites distinct from those in Rab5 to refine the boundaries of the proteolysis-susceptible regions. Homology modelling based on empirical structures of Ran and Ras was used

Abbreviation used: AEBSF, 4-(2-aminoethyl)benzenesulphonyl fluoride.

¹ These authors contributed equally to this study.

² Present address: Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch at Galveston, Galveston, TX 77555, U.S.A.

³ To whom correspondence should be addressed at: Tanox, 10301 Stella Link, Houston, TX 77025, U.S.A. (e-mail shoffenberg@tanox.com).

to resolve an ambiguity in the assignment of one proteolytic site, to explain the nucleotide dependence of another site, and to predict the nucleotide-dependent conformations of Rab5 surface regions revealed by proteolysis. Stimulated by results of the proteolytic analysis, we also performed a sequence comparison of the Switch II residues of Rab family proteins. This reveals a somewhat variable motif overlying the region of Switch II accessible to trypsin, and a highly conserved motif overlying the region of Switch II inaccessible to trypsin. The significance of our findings for structure–function analysis of the interactions of Rab proteins with other macromolecules is discussed.

MATERIALS AND METHODS

Protein expression and purification

Human Rab5A protein was expressed in *Escherichia coli* and purified as described [10]. The Rab5 D136N mutant was purified from a urea extract and refolded by dialysis [22]. Recombinant human Rab4 was expressed as a His-tagged protein from the pET15b plasmid (Novagen, Madison, WI, U.S.A.) and purified as follows. A 100 ml *E. coli* cell culture was pelleted, then lysed in 10 ml of buffer [20 mM Tris/HCl (pH 8.0)/2 mM 2-mercaptoethanol/5 mM MgCl₂/1 μ M GDP] and centrifuged at 20000 g for 1 h; the supernatant was passed through a glass fibre filter. The filtrate was supplemented with 0.1% (v/v) CHAPS, 5 mM imidazole and 0.5 M NaCl, then applied to a 2 ml His-Bind column (Novagen), washed with the same buffer supplemented with 60 mM imidazole, and eluted with 100 mM imidazole. The purified protein was concentrated on a Centricon 10 membrane (Amicon, Beverly, MA, U.S.A.), washed twice with storage buffer [20 mM Tris/HCl (pH 8.0)/1 mM EGTA/2 mM 2-mercaptoethanol/5 mM MgCl₂/1 μ M GDP/0.1% (v/v) CHAPS] to remove imidazole, then stored at -80°C .

Proteolysis and protein sequencing

Proteolysis was performed by using trypsin treated with 1-chloro-4-phenyl-3-L-toluene-*p*-sulphonamidobutan-2-one ('TPCK') (Sigma, St. Louis, MO, U.S.A.) as described [10,22]. In brief, proteins were exposed to trypsin for 15 min at 30 $^{\circ}\text{C}$ in proteolysis buffer [20 mM Tris/HCl (pH 8.0)/1 mM dithiothreitol/1 mM EGTA/5 mM MgCl₂/0.1% (v/v) CHAPS/50 nM GDP]. The reaction was stopped by adding 4 $\times$ Laemmli sample buffer [23] and boiling for 5 min, unless binding assays were planned (see below). The polypeptide products were then separated electrophoretically in polyacrylamide gels buffered with Tris/Tricine [24]. For analytical purposes, 2.5 μ g of Rab protein was incubated with trypsin in 50 μ l of proteolysis buffer; polypeptide products were detected by Coomassie Blue G-250 staining of polyacrylamide gels. For peptide sequencing, 5 μ g of Rab protein was proteolysed, resolved by SDS/PAGE, electrotransferred to Immobilon-P membranes (Millipore, Bedford, MA, U.S.A.) and stained with Coomassie Blue G-250. Bands were then cut out and peptides were sequenced in a 370A automatic sequencer (Applied Biosystems, Foster City, CA, U.S.A.). At least ten N-terminal amino acids were determined for each peptide.

Nucleotide binding assay

Rab5 wild-type and D136N proteins (2 μ M) were treated with trypsin in proteolysis buffer for 15 min. Reactions were stopped with the serine protease inhibitor 4-(2-aminoethyl)benzenesulphonyl fluoride (AEBSF) (Calbiochem, La Jolla, CA, U.S.A.) (4 mM) and added to 100 μ l of binding buffer to make the Rab5

protein concentrations 50 nM. The amounts of resulting peptides were estimated by an analysis of gels stained with Coomassie Blue G250 on an Ultrosan XL densitometer (Pharmacia LKB, Piscataway, NJ, U.S.A.). Competitive nucleotide binding was performed as described previously [22].

MS

Rab5 or Rab4 protein (25 μ g) was digested for 15 min with trypsin in a total volume of 500 μ l of proteolysis buffer. The reaction was stopped by the addition of 4 mM AEBSF, transferred to a Microcon 10 concentrator (Amicon) and washed three times with deionized water. Matrix-assisted laser desorption/ionization mass spectra were obtained on a Voyager Elite time-of-flight mass spectrometer (PerSeptive Biosystems, Vestec Products) equipped with a nitrogen laser (wavelength 337 nm, 3 ms pulse). The matrix used was a saturated solution of sinapinic acid in acetonitrile/water/trifluoroacetic acid (70:30:0.1, by vol.). A 0.7 μ l volume of the matrix was placed on the target, 0.5 μ l of sample solution was added and the sample was allowed to dry at room temperature. The positive-ion linear mode was used with an accelerating voltage of 25 kV. Data were acquired with a TDS 520 oscilloscope (Tektronix, Beaverton, OR, U.S.A.); each spectrum was produced by accumulating data from 50–100 laser shots. Mass was calibrated with bovine carbonic anhydrase ($[M+H]^+ = 29022$) and bovine insulin ($[M+H]^+ = 5734.5$) (Sigma) as standards.

Homology modelling

Modelling was performed with QUANTA/CHARMm, Release 4 (Molecular Simulations, San Diego, CA, U.S.A.) running on an Iris workstation (Silicon Graphics, Mountain View, CA, U.S.A.). PDB files with atomic coordinates for Ras were downloaded from the Brookhaven Protein Data Bank [25]: 1q2l for GDP-bound Ras [18] and 5p2l for GTP-bound Ras [17]. Atomic coordinates for Ran were provided by Dr. Klaus Scheffzek [26]. Sequences were aligned in pairs, protein backbones were optimized by energy minimization and hydrogen atoms were added; the models were then refined with the DISCOVER program (Molecular Simulations) as follows. The force field employed was the all-atom version of AMBER, with a distance-dependent dielectric [27,28]. During the calculations of molecular mechanics, no non-bonded cut-offs were in effect and no restraints were applied. Extensive energy minimization was then performed by the successive application of Steepest Descent and Conjugate Gradient minimizers. Minimization of each model was continued until the maximum derivative of energy fell below 1.0 kcal/mol (4.2 kJ/mol). Secondary structures were identified by the method of Kabsch and Sander [29].

RESULTS

Limited digestion of Rab5 with trypsin

Increasing the trypsin concentration in incubations of constant duration led to the progressive digestion of Rab5·GDP (Figure 1a). With 1.25 ng of trypsin the major products were polypeptides of approx. 25 kDa (indistinguishable from the intact protein) and 20 kDa. The 20 kDa polypeptide contaminant present in the absence of trypsin seemed to be an N-terminal Rab5 product of digestion by a bacterial protease, as described previously [10]. With greater amounts of trypsin the 25 kDa polypeptide was completely digested to the 20 kDa polypeptide product. At still higher trypsin concentrations the 20 kDa polypeptide disappeared and two polypeptides of 8 and 12 kDa were observed.

Figure 1 Proteolysis of Rab5 and Rab4 proteins

Rab5 (2.5 μ g) (a) or Rab4 (2.5 μ g) (b) was digested with the indicated amounts of trypsin in a 50 μ l reaction volume that included 50 nM GDP. The products of Rab protein digestion were identified by Coomassie Blue staining of 12% (w/v) polyacrylamide gels, whereas the products of trypsin self-digestion were too small to be detected. The positions of molecular mass standards are shown (in kDa) at the left, and the inferred molecular masses of the major products of proteolysis are shown at the right.

We have shown previously that the 25 kDa polypeptide is protected by GTP, the 20 kDa polypeptide is protected by GTP or GDP, and the 8 and 12 kDa polypeptides are the major proteolytic products in the absence of added nucleotides or in the presence of GMP [10,30]. In keeping with these findings, the Rab5 S34N mutant with impaired nucleotide binding was recently shown to be rapidly digested to 8 and 12 kDa fragments without the transient appearance of the 20 kDa intermediate [31].

To measure the nucleotide-binding properties of the Rab5 proteolytic fragments in solution, we stopped the proteolytic reaction with the water-soluble serine protease inhibitor AEBSF. AEBSF did not influence the binding of [35 S]guanosine 5'-[γ -thio]triphosphate to the intact protein (results not shown). In the presence of 12.5 ng of trypsin, the 20 kDa polypeptide was generated almost stoichiometrically from the full-length protein (Figure 1a). The 20 kDa polypeptide bound guanyl nucleotides with the same affinity and stoichiometry as the intact protein, and its affinity and capacity for the non-cognate nucleotide XTP were also unchanged (Figure 2a). To extend these observations we performed the same set of experiments with the Rab5 D136N mutant, which has reversed affinities for guanyl and xanthyl nucleotides. Cognate and non-cognate nucleotide binding by the 20 kDa Rab5 D136N polypeptide was indistinguishable from that of the full-length protein (Figure 2b). Similar results have been obtained with H-Ras, wherein two C-terminal truncated

proteins (residues 1–182 and 1–166) had unaltered guanyl nucleotide binding [32].

Further digestion of Rab5 wild-type and D136N proteins did alter nucleotide binding. Proteolytic digests containing predominantly 8 and 12 kDa polypeptides displayed a fraction of the original number of binding sites (Figure 2) commensurate with the mass of residual 20 kDa polypeptide measured by laser scanning densitometry of Coomassie-stained polyacrylamide gels (results not shown). These results indicate that the 8 and 12 kDa polypeptides do not generate a nucleotide-binding site, which is not surprising considering that two of the four G-domains that directly interact with bound nucleotides are located in each of the two resulting fragments (see below). Nevertheless the fact that the 8 and 12 kDa polypeptides were present in equimolar concentrations and not rapidly digested to smaller fragments (Figure 1) suggests that, despite losing the capacity to bind guanyl nucleotides, they associate in the presence of Mg^{2+} ions to maintain a protease-protected conformation. In the absence of Mg^{2+} ions these polypeptide fragments are rapidly digested [10].

Identification of the sites of Rab5 proteolysis

Trypsin cleaves peptide bonds C-terminal to arginine and lysine residues. We determined which of the 11 arginine residues and 13 lysine residues of Rab5 are accessible to trypsin by using N-terminal amino acid sequencing and MS. Sequencing the full-length recombinant protein revealed an intact N-terminus except for the initiating formylmethionine residue, yielding a predicted molecular mass in good agreement with that measured by MS (Table 1). Sequencing the 20 kDa fragment gave a mixture of two peptides with the amino acid sequences ASRGATRPNG (66%) and GATRPNGPNT (33%). Thus the 20 kDa fragment is derived from the N-terminus of the protein (Figure 3), with most of the polypeptide having an intact terminus (minus a methionine residue), but some having lost three additional N-terminal amino acids. Mass measurement did not allow us to assign the C-terminal site of proteolysis unambiguously to Arg-195 or Arg-197 (Table 1), but molecular modelling suggested that Arg-197 is the major site (see below). The 8 and 12 kDa fragments had the following N-terminal sequences, respectively: GATRPNGPNT and YHSLAPMYR. Together with mass determination, this identifies the 8 kDa polypeptide as Gly-5–Arg-81 (calculated molecular mass 8400 Da) and the 12 kDa polypeptide as Tyr-82–Arg-197 (12880 Da), assuming that Arg-195 is not a site of digestion. In summary, the major sites of Rab5 proteolysis are Arg-4, Arg-81 and Arg-197. Of these, Arg-197 is the most susceptible to proteolysis, and Arg-81 the least so.

Refinement of the trypsin-accessible boundaries of the Rab Switch II region

The Arg-81 site of Rab5 proteolysis lies within a region corresponding to the conformationally variable Switch II of other GTPases [17–19,33]. To determine more precisely the extent of the Rab Switch II region with nucleotide-dependent trypsin accessibility, we performed proteolysis experiments with Rab4 similar to those performed with Rab5. Additional arginine residues in Rab4 that can be probed with trypsin lie between the amino acid positions corresponding to Rab5 Arg-81, which is cleaved, and Arg-91, which remains intact (Figure 4). The Rab4 protein was His-tagged and included a thrombin cleavage site, but these residues are excluded from the reported data. Digestion of Rab4 gave essentially the same pattern as Rab5 except that

Figure 2 Nucleotide binding to intact and trypsin-treated Rab5

Binding of [³⁵S]guanosine 5′-[γ-thio]triphosphate to intact and proteolysed Rab5 wild-type (**a**), or of [³H]XDP to intact and proteolysed Rab5 D136N mutant (**b**), were competed with varying concentrations of unlabelled GTP and XTP. Symbols: ●,■,▲, GTP; ○,□,△, XTP; ●,○ no trypsin; ■,□ predominantly 20 kDa peptide (12.5 ng of trypsin); ▲,△ predominantly 8 and 12 kDa peptides (125 ng of trypsin). Insets represent normalized binding curves of the same data.

Table 1 Rab5 fragments generated by tryptic proteolysis

Masses of proteolytic fragments determined by matrix-assisted laser desorption ionization–time-of-flight MS are compared with candidate fragments on the basis of a calculation of their molecular masses. Molecular masses by MS are means ± S.E.M.

Approx. molecular mass by SDS/PAGE (kDa)	Molecular mass by MS (Da)	Candidate fragments	
		Designation	Molecular mass (Da)
25	23550 ± 350	Rab5 _{1–215}	23 700
		Rab5 _{2–215}	23 500
20	21500 ± 300	Rab5 _{2–197}	21 590
		Rab5 _{2–195}	21 370
		Rab5 _{5–197}	21 270
		Rab5 _{82–195}	12 670
12	12800 ± 120	Rab5 _{82–197}	12 880
		Rab5 _{2–81}	8720
8	8600 ± 100	Rab5 _{5–81}	8400

Rab5

1 MASRGATRPN GPNTGNKICO FKLVLIGESA VCKSSLVLR VKGFHEFOE

Rab4

1 MSETYDFL FKFLVIGNAG TGKSCLLHOF IEKKFKDDSN

51 STIGAAFLTO TVCLDDTTVK FEIWDTAGOE RYHSLAPMY RGAQAAIVVY DITNEESFAR

39 HTIGVEFGSK IINVGKQYVK LOIWDTAGOE RFRSVTRSTY RGAAGALVY DITSRETYNA

111 AKNWVKELQR QASPNIVIAL SGNKADLANK RAVDFQEAQS YADDNSLLEF ETSAKTSMNV

99 LTNWLTARM LASONIVIIL CGNKKBLDAD REVTFLEASR FAQENELMFL ETSALTGEDV

171 NEIFMAIAKK LPKNEPONPG ANSARGRGVD LTEPTQPTRN QCCSN

159 EEAQVQCARK ILNKIESGEL DPERMGSGIQ YGDAALRQLR SPRRTQAPNAQECGC

Figure 3 Amino acid sequence alignment of Rab5 and Rab4

N-terminal sequences determined by automated Edman degradation are shown in outline font, and consensus G-domains are shadowed. The amino acid sequences of the 8 kDa fragments inferred from sequence analysis and MS are singly underlined, and the inferred amino acid sequences of the 12 kDa fragments are doubly underlined.

		— G — — V — C —	
Rab5	74	WDTAGQER <u>RY</u> HSLAPMYRGA	93
Rab4	62	WDTAGQER <u>R</u> FRSVTRSYRGA	81
Rab3	76	WDTAGQER <u>RY</u> RTITTAYRGA	95
Rab2	58	WDTAGQES <u>S</u> FRSITRSYYRGA	79
H-Ras	56	LDTAGQEEYSAMRDQYMRGTG	75

Figure 4 Sequence alignment of Rab Switch II

The G-domain involved in nucleotide binding is shown in outline font, the sequence-variable and trypsin-accessible region of Switch II is shown in bold font, and the sequence-constant trypsin-inaccessible region of Switch II is shown in outline font. Rab5 Arg-81 and corresponding residues in other Rab proteins are doubly underlined. The human H-Ras Switch II region is presented for comparison.

progression from a 20 kDa polypeptide to 8 and 12 kDa polypeptides occurred at lower concentrations of trypsin (Figure 1b), possibly because Rab4 has multiple trypsin-sensitive sites in the Switch II region, whereas Rab5 has only one (see below). The major digestion products of Rab4 (calculated molecular mass

23 900 Da) were a 20 kDa polypeptide (measured molecular mass 20 600 Da), an 8 kDa polypeptide (N-terminal sequence MSETYDF, measured molecular mass 7800 Da) and a 12 kDa polypeptide (N-terminal sequence SYRGAAGALL, measured molecular mass 12 000 Da) (Figure 3). These results indicate that the sites of proteolysis are Arg-69, Arg-75 and Arg-182, and the three polypeptides are Met-1–Arg-182, Met-1–Arg-69 and Ser-76–Arg-182 (Figure 3). If the results from the Rab4 and Rab5 digestions are integrated, the limits of the nucleotide-dependent, trypsin-accessible surface of the Rab Switch II region are Arg-81 to Met-88 for Rab5 and Arg-69 to Ser-76 for Rab4 (Figure 4).

Protein homology modelling

To determine the nucleotide-dependent conformations of Rab5 corresponding to those revealed by proteolysis, we performed homology modelling (Figure 5 and Table 2). For comparison of GTP-bound and GDP-bound conformations of the Rab5 Switch II region, we used H-Ras structures. However, the position of Arg-197 could not be modelled with Ras as a template because the crystallized Ras protein was truncated at its C-terminus. Therefore the Ran structure, whose coordinates are available only for the GDP-bound form [26], was used as a template to build a third model. We refer to our three models as Rab5_{Ras}·GDP, Rab5_{Ras}·GTP, and Rab5_{Ran}·GDP, where the subscripts denote 'Ras-based' or 'Ran-based'. The models are shown in Figure 5, and a detailed analysis of secondary structural differences between them is presented in Table 2. The overall architecture of the protein, a six-stranded β -sheet surrounded by six α -helices, is conserved in all three models, although the

Figure 5 Secondary structure renderings for Rab5 homology models

(a) Ras·GDP-based model, residues 21–174; (b) Ras·GTP-based model, residues 18–181; (c) Ran·GDP-based model, residues 17–213. The N-termini (small grey spheres) and C-termini (C) are marked. Secondary structure rendering is based on the Kabsch and Sander [29] algorithm: helices (yellow cylinders), β -strands (grey ribbons), and turns and coil (orange tubes). Three key arginine residues are shown in detail: residues 81 (white), 195 (yellow) and 197 (magenta).

Table 2 Secondary structures in Rab5 models

Secondary structures are identified in accordance with Kabsch and Sander [23].

	Ras-based model		Ran-based model	Stenmark et al. [15] model
	GDP-bound	GTP-bound	GDP-bound	GTP-bound
β -sheet1	25–26	23–24	22–26	19–27
Turn	—	—	29–32	29–30
Helix1	33–41	33–41	33–39	33–42
Turn	—	—	41–44*	—
Turn	—	—	45–46	—
Strand†	—	—	48–50	—
Turn	51–52	—	51–54	—
β -sheet2	55–58	54–61	55–57	55–63
Turn	64–66*	64–66*	65–66	64–66
β -sheet3	72–75	69–76	72–74	67–75
Turn	—	80–81	80–82	—
Helix2	87–90	83–90	—	85–91
Turn	91–92	91–92	87–92	—
β -sheet4	96–101	95–101	93–99	95–101
Turn	102–103	102–103	102–103	102–103
Helix3a	105–111	105–111	105–111	105–111
Helix3b	113–121	113–119	113–116	113–120
Turn	—	120–122	117–122	—
Turn	124–125	—	124–125	123–124
β -sheet5	128–133	128–133	127–131	128–133
Turn	—	135–136	135–136	135–136
Turn	—	139–140	—	—
Turn	145–146	—	145–147	—
Helix4	147–150	145–154	151–154	145–154
Turn	151–156	155–156	155–156	155–156
β-sheet6	159–162	159–162	158–160	158–162
Turn	164–166*	164–166	164–169	164–171
Helix5	170–174	170–181	170–180	172–180
Turn	—	—	181–183	181–183
Turn	—	—	185–186	—
Helix6	—	—	205–210	—
Turn	—	—	211–212	—
Total residues	21–174	18–181	17–213	18–184

* A very short stretch of helical conformation.

† Strand, not part of a β -sheet.

Table 3 Root-mean-square deviations among Rab5 models

Root-mean-square deviations of the Rab5 models described in Table 2 were calculated by using the best superpositions of α -carbons 21–174.

Model	Deviation (Å)		
	Rab5 _{Ras} ·GTP	Rab5·GTP	Rab5 _{Ran} ·GDP
Rab5 _{Ras} ·GDP	1.988	2.101	4.340
Rab5 _{Ras} ·GTP	—	1.670	4.600
Rab5·GTP [15]	—	—	4.530

precise lengths of helices and strands vary slightly. The sixth helix is seen only in the Ran-based model because the corresponding residues are absent from the other two. Table 3 summarizes the overall structural differences between our three models and an earlier Rab5 model built by Stenmark et al. [15], which is based on Ras·GTP. It is worth noting that the last model has a Gly residue at the important position 81 that has since been corrected to Arg-81 [10], as in our models.

Arg-81 in GDP-bound and GTP-bound Rab5

Nucleotide-dependent solvent accessibility and interactions within chains (Figures 5 and 6) provide an explanation for the smaller susceptibility of Arg-81 to digestion in the presence of GTP than in the presence of GDP [10,30]. The side chain of Arg-81 juts outwards into the solvent in Rab5_{Ras}·GDP and Rab5_{Ran}·GDP, whereas it is folded against the surface of the protein in Rab5_{Ras}·GTP (Figure 5). In GTP-bound Rab5 (Figure 6b), the Arg-81 side chain forms two strong hydrogen bonds with the backbone carbonyl groups of Thr-52 and Ile-53. These interactions seem to stabilize the Arg-81 side chain in that conformation, protected from proteolysis by the surrounding residues 52–54. In contrast, in GDP-bound Rab5 (Figure 6a) these residues are too far from the Arg-81 side chain for stabilizing interactions to occur.

Arg-195 and Arg-197 of the Ran-based Rab5 model

We analysed the Rab5 Arg-195–Arg-197 region and its environment in a similar fashion by using the Ran-based model to determine the likely site of tryptic digestion because MS analysis did not unambiguously identify the site. The Arg-195 side chain forms four hydrogen bonds with the side chains of Asn-192 and

Figure 6 The neighbourhood and interactions of Arg-81 in Rab5 models

Residues around Arg-81 and other residues that might interact with it are shown. The atomic colouring scheme is as follows: C, black; N, blue; O, orange; H, grey; S, yellow. Broken lines connecting O and H atoms represent hydrogen bonds. (a) Ras·GDP-based model; (b) Ras·GTP-based model.

Figure 7 The environment and interactions of residues 195–197 in the Ran-based Rab5 model (Rab5_{Ran}·GDP)

Broken lines mark hydrogen bonds, and the two arrows point to the peptide bonds cleavable by trypsin (see the text for discussion). The colouring scheme is as described in the legend to Figure 6.

Asn-171, effectively burying this amino acid from the solvent (Figure 7). In contrast, Arg-197 is exposed to the solvent, indicating its susceptibility to proteolysis. Because the con-

formation of side chains is generally more subject to error than the protein backbone in homology modelling, especially when the residue's identity is not conserved between the reference protein and the model, there is some concern that our predicted side chain orientations could be modelling artifacts. However, for the following reasons we believe that the orientations of Arg-81, Arg-195 and Arg-197 in our models are reasonable. First, in the homology modelling process, the side chains were oriented in accordance with the Karplus library built from side chain conformations observed in globular proteins [34]; during subsequent energy minimization of the model, the three Arg side chains did not move significantly, indicating that they were already at or near stable local energy minima. Secondly, the side chains of Arg-81 in Rab5_{Ras}·GTP and Arg-195 in Rab5_{Ran}·GDP, which are buried, have extensive intramolecular hydrogen-bonding interactions that compensate for the destabilizing free energy of burying polar residues from solvent water. Thirdly, on comparing the reference and model proteins, each Arg side chain (Arg-81 in Rab5_{Ras}·GDP and in Rab5_{Ras}·GTP; Arg-195 and Arg-197 in Rab5_{Ran}·GDP) is in the same exposed/buried state as its counterpart in the reference proteins (Glu-63 in Ras, and Leu-182 and Pro-184 in Ran).

DISCUSSION

The sites of nucleotide-dependent tryptic digestion that we have identified here correlate with sites of protein–protein interaction previously described for Rab5 and other small GTPases. The first site of Rab5 proteolysis, Arg-197, lies within an RXRGVDLXE motif shared by all Rab5 isoforms. There is evidence that this region is important for targeting Rab proteins to membranes [35]. The fact that this site is the most susceptible to trypsin cleavage but is protected in the GTP-bound state correlates with the observation that the Rab5 C-terminus can

Figure 8 Stereo diagram of superimposed models of Rab5_{Ras} · GDP (blue) and Rab5_{Ras} · GTP (red), showing residues 77–91

The side chain represented in atomic detail is Arg-81, coloured grey in Rab5_{Ras} · GDP and orange in Rab5_{Ras} · GTP. Note the difference between its orientations in the two models.

Table 4 Frequencies of amino acids in the Switch II(v) motif of Rab proteins

The frequencies of amino acids in 23 Rab protein sequences published by Chardin [42] are expressed as percentages. The Rab5 sequence has been corrected as described [10].

Amino acid at numbered position (%)								
Amino acid position ...	1	2	3	4	5	6	7	8
	Arg (96)	Phe (70)	Arg (87)	Thr (52)	Ile (61)	Thr (78)	Thr (39)	Ala (52)
	Ser (4)	Tyr (30)	His (9)	Ser (26)	Leu (22)	Ile (9)	Ser (35)	Ser (39)
			Gln (4)	Ala (17)	Val (13)	Ala (4)	Pro (13)	Pro (4)
				Trp (4)	Met (4)	Gly (4)	Arg (9)	Met (4)
						Arg (4)	Val (4)	

become prenylated in the GDP-bound but not in the GTP-bound conformation [36]. We speculate that the C-terminus is buried and inaccessible to Rab geranylgeranyl transferase in the GTP-bound conformation. The initial site of Rab4 proteolysis is also located in the C-terminus, three residues upstream from the corresponding site in Rab5.

The region surrounding the second site of Rab5 digestion, Arg-4, has been previously implicated in Rab5 function. Endosome fusion is inhibited *in vitro* either by tryptic cleavage of the Rab5 N-terminus or by competition with an N-terminal peptide that requires the four N-terminal amino acids [16]. However, truncation of the four N-terminal amino acids does not result in any loss of endocytic activity *in vivo*, although truncation of 19 amino acids results in partial loss [37]. Replacement of the 20-residue Rab5 N-terminus with Rab6 in a chimaeric protein leads to loss of Rab5 function *in vivo*, whereas the addition of the Rab5 N-terminus to a chimaera that inhibits endocytosis converts it into one that stimulates [15].

The third site of Rab5 proteolysis, Arg-81, lies within the Switch II region. Switch II is the only region in which superimposed α -carbon traces of the Rab5_{Ras} · GDP and Rab5_{Ras} · GTP models diverge substantially. The root-mean-square deviation for the eight trypsin-accessible residues of the Switch II region (residues 81–88) was 3.32 Å, compared with 1.52 Å for the entire protein. Figure 8 illustrates the magnitude of the conformational change predicted to take place in this region on GTP hydrolysis, including a loss of helical structure in residues 83–87.

The Switch II regions of several GTPases have been shown to participate in protein–protein interactions. For example, Ras proteins require specific amino acids within the Switch II region for interaction with Raf-1, B-Raf, phosphatidylinositol 3-kinase and guanine nucleotide exchange factors [21]. Similarly, the Switch II region of the α subunit of transducin in the GDP-bound form is accessible to proteases [38,39] and interacts with the $\beta\gamma$ subunits of the G_i trimer [40]. Direct evidence for the importance of the Switch II region in Rab proteins comes from

an endosome fusion assay *in vitro* in which an R81A mutation in Rab5 decreases its activity moderately and, when introduced as a second mutation, completely prevents prenylation and the inhibitory effect of Rab5 N133I [41].

In comparing the primary structure of the Switch II regions of different Rab proteins [42], a loosely conserved motif [Arg-Phe/Tyr (aromatic)-Arg (basic)-Thr/Ser-Ile/Leu (aliphatic)-Thr-Thr/Ser-Ala/Ser] was found in those amino acids (residues 81–88 in Rab5) having nucleotide-dependent trypsin susceptibility and a high degree of conformational mobility by modelling (Table 4 and Figure 5). Only the aromatic amino acid in position 2 is conserved throughout the Ras GTPase superfamily. We term this Rab motif 'Switch II(v)' for its relative sequence variability. C-terminal to Switch II(v) is a region whose primary structure is highly conserved in Rab proteins but different from other GTPases (Figure 4). We term this 'Switch II(c)' for its constant YYRGA motif. Only the arginine residue in position 3 is conserved throughout the Ras superfamily. N-terminal to Switch II(v) is an invariant G-domain (WDTGAQE) involved in nucleotide binding and hydrolysis (Figure 4). The tryptophan at the beginning of this motif is found only in Rab and Rho GTPases, whereas the other residues are found in all regulatory GTPases. In summary, the Rab Switch II(v) region is flanked on both sides by sequence-invariant regions. We propose that the variable residues of the Rab Switch II(v) region have a role in determining the specificity of protein-selective interactions, such as those with certain effectors, GTPase accelerating proteins and guanine nucleotide dissociation stimulators [8,43,44]. The highly conserved residues of the Switch II(c) region might mediate family-wide interactions, such as those with Rab escort protein, GDP dissociation inhibitor and prenyltransferase [4,5,45], or they might act primarily to increase the affinity of protein-specific interactions mediated by Switch II(v). In support of a role for Switch II(v) in determining the specificity of macro-molecular interactions, Rab proteins that interact with Mss4 have nearly identical Switch II(v) regions, whereas those that do not interact with Mss4 have distinct Switch II(v) regions [46].

We thank Dr. William Seifert and Annie Ballatore (Analytical Chemistry Center, University of Texas Health Science Center—Houston, Houston, TX, U.S.A.) for the MS work, Dr. Richard Cook (Protein Chemistry Core Laboratory, Baylor College of Medicine, Houston, TX, U.S.A.) for the amino acid sequencing, Dr. Keith Morrison (Center for Experimental Therapeutics, Baylor College of Medicine, Houston, TX, U.S.A.) for the plasmid carrying His-tagged Rab4, and Julie Pan and Dr. Marianne Wessling-Resnick (Harvard School of Public Health, Cambridge, MA, U.S.A.) for the plasmid carrying Rab5 D136N. This work was supported by a Grant-In-Aid from the American Heart Association, Texas Affiliate (to S.H.), and United States Public Health Service Award HL43161 and a Veterans Administration Merit Review Award (to B. F. D.).

REFERENCES

- Walworth, N. C., Goud, B., Kabacnel, A. K. and Novick, P. J. (1989) *EMBO J.* **8**, 1685–1693
- Tisdale, E. J., Bourne, J. R., Khosravi-Far, R., Der, C. J. and Balch, W. E. (1992) *J. Cell Biol.* **119**, 749–761
- Flynn, G. C., Pohl, J., Flocco, M. T. and Rothman, J. E. (1991) *Nature (London)* **353**, 726–730
- Andres, D. A., Seabra, M. C., Brown, M. S., Armstrong, S. A., Smeland, T. E., Cremers, F. P. M. and Goldstein, J. L. (1993) *Cell* **73**, 1091–1099
- Seabra, M. C., Goldstein, J. L., Südhof, T. C. and Brown, M. S. (1992) *J. Biol. Chem.* **267**, 14497–14503
- Nuoffer, C., Wu, S. K., Dascher, C. and Balch, W. E. (1997) *Mol. Biol. Cell* **8**, 1305–1316
- Soldati, T., Riederer, M. A. and Pfeffer, S. R. (1993) *Mol. Biol. Cell* **4**, 425–434
- Stenmark, H., Vitale, G., Ullrich, O. and Zerial, M. (1995) *Cell* **83**, 423–432
- Horiuchi, H., Lippé, R., McBride, H. M., Rubino, M., Woodman, P., Stenmark, H., Rybin, V., Wilm, M., Ashman, K., Mann, M. and Zerial, M. (1997) *Cell* **90**, 1149–1159
- Hoffenberg, S., Sanford, J. C., Liu, S., Daniel, D. S., Tuvin, M., Knoll, B. J., Wessling-Resnick, M. and Dickey, B. F. (1995) *J. Biol. Chem.* **10**, 5048–5056
- Simon, I., Zerial, M. and Goody, R. S. (1996) *J. Biol. Chem.* **271**, 20470–20478
- Rybin, V., Ullrich, O., Rubino, M., Alexandrov, K., Simon, I., Seabra, M. C., Goody, R. and Zerial, M. (1996) *Nature (London)* **383**, 266–269
- Bucci, C., Parton, R. G., Mather, I. H., Stunnenberg, H., Simons, K., Hoflack, B. and Zerial, M. (1992) *Cell* **70**, 715–728
- Gorvel, J.-P., Chavrier, P., Zerial, M. and Gruenberg, J. (1991) *Cell* **64**, 915–925
- Stenmark, H., Valencia, A., Martinez, O., Ullrich, O., Goud, B. and Zerial, M. (1994) *EMBO J.* **13**, 575–583
- Steele-Mortimer, O., Clague, M. J., Huber, L. A., Chavrier, P., Gruenberg, J. and Gorvel, J.-P. (1994) *EMBO J.* **13**, 34–41
- Pai, E. F., Krenzel, U., Petsko, G. A., Goody, R. S., Kabsch, W. and Wittinghofer, A. (1990) *EMBO J.* **9**, 2351–2359
- Tong, L., Devos, A. M., Milburn, M. V. and Kim, S.-H. (1991) *J. Mol. Biol.* **217**, 503–516
- Stouten, P. F. W., Sander, C., Wittinghofer, A. and Valencia, A. (1993) *FEBS Lett.* **320**, 1–6
- Kjeldgaard, M., Nyborg, J. and Clark, B. F. C. (1996) *FASEB J.* **10**, 1347–1368
- Moodie, S. A., Paris, M., Villafranca, E., Kirschmeier, P., Willumsen, B. M. and Wolfman, A. (1995) *Oncogene* **11**, 447–454
- Hoffenberg, S., Nikolova, L., Pan, J. Y., Daniel, D. S., Wessling-Resnick, M., Knoll, B. J. and Dickey, B. F. (1995) *Biochem. Biophys. Res. Commun.* **215**, 241–249
- Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685
- Schagger, H. and von Jagow, G. (1987) *Anal. Biochem.* **166**, 368–379
- Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, Jr., E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. and Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542
- Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W. and Wittinghofer, A. (1995) *Nature (London)* **374**, 378–381
- Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, Jr., S. and Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784
- Weiner, S. J., Kollman, P. A., Nguyen, D. T. and Case, D. A. (1986) *J. Comp. Chem.* **7**, 230–252
- Kabsch, W. and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- Hoffenberg, S., Shannon, T. M., Noonan, T. P., Liu, S., Daniel, D. S., Fishman, J. B., Rubins, J. B., Misra, H. K., Wright, G. E. and Dickey, B. F. (1996) *Mol. Pharmacol.* **49**, 156–164
- Pan, J. Y., Sanford, J. C. and Wessling-Resnick, M. (1996) *J. Biol. Chem.* **271**, 1322–1328
- John, J., Schlichting, I., Schiltz, E., Rusch, P. and Wittinghofer, A. (1989) *J. Biol. Chem.* **264**, 13086–13092
- Song, J. S., Gomez, J., Stancato, L. F. and Rivera, J. (1996) *J. Biol. Chem.* **271**, 26962–26970
- Dunbrack, Jr., R. L. and Karplus, M. (1993) *J. Mol. Biol.* **230**, 543–574
- Bucci, C., Lütcke, A., Steele-Mortimer, O., Olkonen, V. M., Dupree, P., Chiariello, M., Bruni, C. B., Simons, K. and Zerial, M. (1995) *FEBS Lett.* **366**, 65–71
- Sanford, J. C., Pan, Y. and Wessling-Resnick, M. (1993) *J. Biol. Chem.* **268**, 23773–23776
- Li, G. and Stahl, P. D. (1993) *J. Biol. Chem.* **268**, 24475–24480
- Lambright, D. G., Noel, J. P., Hamm, H. E. and Sigler, P. B. (1994) *Nature (London)* **369**, 621–628
- Fung, B. K.-K. and Nash, C. R. (1983) *J. Biol. Chem.* **258**, 10503–10510
- Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E. and Sigler, P. B. (1996) *Nature (London)* **379**, 311–319
- Li, G., Barbieri, M. A., Colombo, M. I. and Stahl, P. D. (1994) *J. Biol. Chem.* **269**, 14631–14635
- Chardin, P. (1993) in *GTPases in Biology* (Dickey, B. F. and Birnbaumer, L., eds.), pp. 159–176, Springer-Verlag, Berlin
- Fukui, K., Sasaki, T., Imazumi, K., Matsuura, Y., Nakanishi, H. and Takai, Y. (1997) *J. Biol. Chem.* **272**, 4655–4658
- Wada, M., Nakanishi, H., Satoh, A., Hirano, H., Obaishi, H., Matsuura, Y. and Takai, Y. (1997) *J. Biol. Chem.* **272**, 3875–3878
- Pfeffer, S. R., Dirac-Svejstrup, A. B. and Soldati, T. (1995) *J. Biol. Chem.* **270**, 17057–17059
- Burton, J. L., Slepnev, V. and De Camilli, P. V. (1997) *J. Biol. Chem.* **272**, 3663–3668