Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Refinement of the structure of human Rab5a GTPase domain at 1.05 Å resolution

Simon Terzyan,[a]* Guangyu Zhu,[a] Guangpu Li[b] and Xuejun C. Zhang[a]

[a]Crystallography Research Program of Oklahoma Medical Research Foundation, 825 NE 13th Street, Oklahoma City, OK 73104, USA, and [b]Department of Biochemistry and Molecular Biology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Street, Oklahoma City, OK 73104, USA

Correspondence e-mail: terzyans@omrf.ouhsc.edu

Received 25 August 2003
Accepted 30 September 2003

PDB Reference: Rab5a GTPase domain, 1r2q, r1r2qsf.

Rab5 is a GTPase that regulates early endosome fusion. Its GTPase domain crystal structure is reported here at 1.05 Å resolution in complex with a GTP-analog molecule. It provides the highest resolution three-dimensional model so far obtained for proteins from the Ras-like GTPase family. This study allows extension of structural examination of the GTPase machinery as well as of high-resolution protein structures in general. For example, a buried water-molecule network was observed underneath the switch regions, which is consistent with the functional roles of these regions in the molecular-switching process. Furthermore, residues of multiple conformation and clustered distribution of anisotropic thermal motions of the protein molecule may have general implications for the function of Ras-like GTPases.

1. Introduction

Rab5, a member of the Rab protein family, is an essential regulator of early endosome fusion (Li, 1996). It belongs to the Ras-like small GTPase superfamily (Schweins & Wittinghofer, 1994; Vetter & Wittinghofer, 2001) and cycles between GTP-bound (active) and GDP-bound (inactive) forms. Its intrinsic GTPase activity is relatively high among known Rab proteins (Simon *et al.*, 1996) and can be further stimulated by Rab5-specific GTPase-activating proteins (Lanzetti *et al.*, 2000; Liu & Li, 1998; Xiao *et al.*, 1997). In humans, Rab5 has been identified in three isoforms (Rab5a–c; Bucci *et al.*, 1995; Pereira-Leal & Seabra, 2000). Rab5 has been studied extensively for its roles in early endosome fusion (Bucci *et al.*, 1992; Gorvel *et al.*, 1991; Li & Stahl, 1993) and as a model system of GTPase biochemistry (Li & Liang, 2001; Zhu *et al.*, 2003). Although Ras-like small GTPases have been widely recognized as an extremely important family of proteins that regulate many aspects of cellular processes (Bourne *et al.*, 1991; Vetter & Wittinghofer, 2001), only the p21/H-Ras crystal structure has so far been reported at near-atomic resolution (1.26 Å; PDB code 1ctq). Reported here is the crystal structure of Rab5a GTPase domain complexed with a GTP analog refined at 1.05 Å resolution using X-ray diffraction data collected at a synchrotron source.

2. Experimental procedures

2.1. Expression and purification of recombinant Rab5a protein from *Escherichia coli*

Human Rab5a was subcloned into the bacterial expression vector pET11 from a previously described pGEX-3X construct containing the Rab5a cDNA (Liang *et al.*, 2000). After comparison with the canonical folding of small GTPases,

we deleted the N- and C-terminal 'hypervariable' regions of Rab5a in order to promote crystallization. As a result, the recombinant protein contained a starting-codon-derived methionine residue and residues 15–184 of the 215-residue native Rab5a. The recombinant protein was expressed in the BL21 strain of *E. coli* and purified as a soluble protein from the cytoplasm as described previously (Zhu *et al.*, 2003).

2.2. Crystallization of Rab5a

Rab5a recombinant protein was crystallized using the hanging-drop vapor-diffusion method using previously reported conditions (Merithew *et al.*, 2001). Protein at $\sim$15 mg ml^{-1} concentration in a buffer containing 20 m*M* Tris–HCl pH 8.0 and 5 m*M* guanosine-5′-(β,γ)-imidotriphosphate (GppNHp) was mixed with an equal volume of reservoir solution containing 10%(*w*/*v*) polyethylene glycol 6000, 50 m*M* 2-(*N*-morpholino)ethanesulfonic acid pH 6.0, 0.2 *M* NaCl, 1 m*M* $MgCl_2$ and 0.1%(*v*/*v*) β-mercaptoethanol. Crystals of size 0.2 × 0.2 × 0.3 mm grew over one week. Reservoir solution supplemented with 30%(*v*/*v*) glycerol was used as a cryoprotection solution to soak the crystal prior to cooling in a 100 K nitrogen-gas stream for data collection.

2.3. Data collection

A diffraction data set from the Rab5a crystal was collected to 1.00 Å resolution at the Cornell High Energy Synchrotron Source (CHESS) F1 station (Cornell University, Ithaca, NY, USA) using a radiation wavelength of 0.916 Å and a Quantum 4 CCD detector system. In order to cover the whole range of intensities, data were collected in two sweeps differing in crystal-to-detector distance (70 and 200 mm, with corresponding highest possible resolutions of 1.0 and 2.0 Å) and exposure time (50 s for high-resolution and 10 s for low-resolution images). Data were processed with the *HKL* program suite (Otwinowski & Minor, 1997). The highest resolution of 1.05 Å was chosen based on the criterion that in the highest shell more than 50% reflections satisfied $I/\sigma(I) > 1$ and the average $I/\sigma(I)$ was greater than 2.0 for all reflections in the shell. The isotropic crystal mosaicity parameter was refined to be 0.36°.

2.4. Structural refinement

The crystal belongs to space group $P2_12_12_1$ with one Rab5 molecule per asymmetric unit. Initially, several cycles of Rab5a refinement were carried out with the program *CNS* (Brünger *et al.*, 1998) using the previously reported 1.5 Å isomorphous crystal structure of human Rab5a (PDB code 1n6h; Zhu *et al.*, 2003) with removed substrate, cofactor and water molecules as the initial model. One cycle of rigid-body refinement at 50–2.0 Å resolution was performed, followed by ten cycles of geometry-restrained positional and temperature-factor refinement, during which the resolution was extended to its highest limit. In the next step, the GppNHp molecule and an Mg^{2+} ion were added to the model using the interactive graphics program *TURBO-FRODO* (Roussel & Cambillau, 1989) and guided by $2|F_o| - |F_c|$ and $|F_o| - |F_c|$ difference maps.

Table 1
Crystallographic data-collection and refinement statistics.

Valuess in parentheses are for the highest resolution shell.

Data statistics	
Space group	$P2_12_12_1$
Unit-cell parameters (Å)	
a	35.73
b	64.57
c	66.18
Unit-cell volume (Å^3)	152677
Solvent content (%)	39
Mosaicity (°)	0.36
Resolution (Å)	50–1.05 (1.07–1.05)
R_{merge} (%)	6.5 (52)
No. of measured reflections	332387
No. of independent reflections	68700 (3050)
Completeness (%)	95 [79; 52 for $I > \sigma(I)$]
Redundancy	4.8
$\langle I\rangle/\langle\sigma(I)\rangle$	18.7 (2.2)
Refinement statistics	
R_{work} (%)/No. of reflections	12.39/66637
R_{work} (%)/No. of reflections [$F > 4\sigma(F)$]	11.68/56721
R_{free} (%)/No. of reflections	16.88/2063
R_{free} (%)/No. of reflections [$F > 4\sigma(F)$]	15.92/1740
R for all data (%)/No. of reflections	12.47/68700
R for all $F > 4\sigma(F)$ (%)/No. of reflections	11.76/58461
No. of non-H atoms	
No. protein, GppNHp and Mg^{2+} atoms	1372
No. solvent atoms	205
R.m.s. deviation from ideal values†	
Bond length (Å)	0.015
Bond angle (°)	2.366
Average B_{iso} (Å^2)	
Protein	20.4 (Wilson *B* factor 10.2)
Protein main chain	16.2
Protein side chain	23.7
GppNHp	12.4
Solvent atoms	35.3
Average anisotropy for all protein atoms	0.51
Ratio of data to anisotropic parameters	4.5

† Calculated with *PROCHECK* (Laskowski *et al.*, 1993).

Two cycles of automatic solvent divining located 168 bound water molecules, which were incorporated into the model. Several further cycles of positional and *B*-factor refinement with anisotropic overall temperature-factor and bulk-solvent correction in *CNS* resulted in an R_{work} of 0.235 (64 991 reflections with $|F_o| > 0$) and an R_{free} of 0.245 (for 2054 reflections). This model was further refined using *SHELX*97 (Sheldrick & Schneider, 1997). Refinement was carried out against all unique reflections [with the $F_o^2 > -3\sigma(F_o^2)$ criteria in *SHELX*97], except those used in the validation set, which was kept the same for both the *CNS* and the *SHELX*97 refinements. After 20 cycles of conjugate-gradient least-squares refinement with isotropic thermal parameters, which reduced R_{work} to 0.21, the model was thoroughly examined for possible errors using $2m|F_o| - D|F_c|$ (1.5σ) and $|F_o| - |F_c|$ ($\pm$3σ) difference maps. Full anisotropic refinement introduced at this point resulted in an R_{work} of 0.156 and an R_{free} of 0.199. The following iterative cycles of refinement and model building revealed an additional 37 water and five glycerol molecules. Water molecules were refined with the ISOR restraint in order to reduce their parameter freedom. In addition, 25 residues were modeled in multiple conformations. The occupancies of

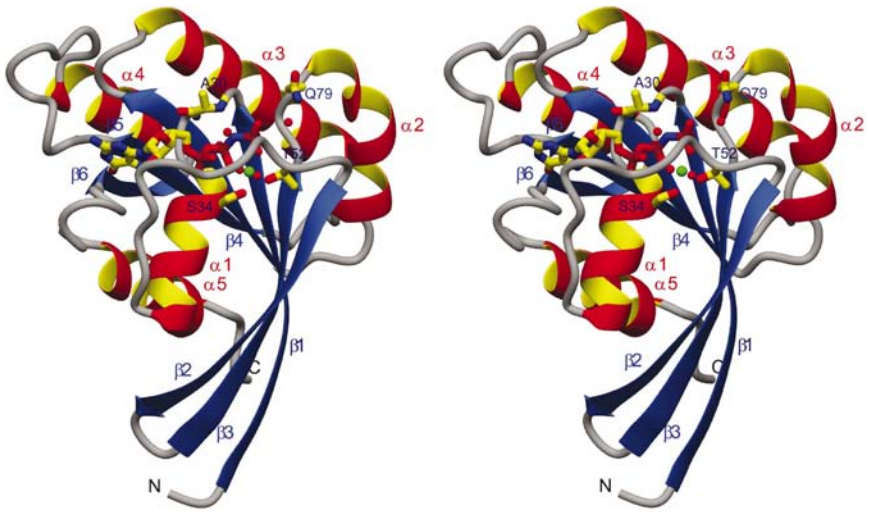

Figure 1
Stereo ribbon diagram of the Rab5 catalytic domain. Elements of secondary structure are labeled (helices as $\alpha1$–$\alpha5$ and strands as $\beta1$–$\beta6$), as well the N- and C-termini. Some catalytically important residues, *i.e.* Ala30, Ser34, Thr52 and Gln79, together with the GppNHp molecule, are shown as stick models. The Mg^{2+} ion and four water molecules that are considered important for GTP hydrolysis are represented as green and red spheres, respectively.

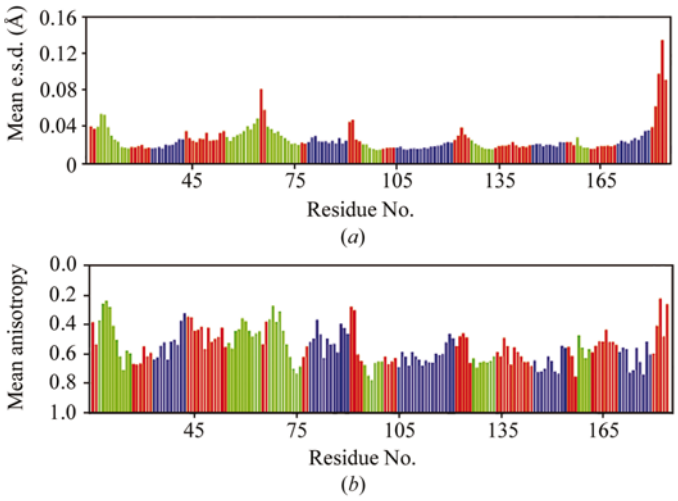

Figure 2
Properties of the refined model. (*a*) Distribution of mean r.m.s. coordinate error for the backbone atoms along the peptide chain. (*b*) Distribution of anisotropy (*i.e.* the ratio of the minimum to maximum eigenvalues of anisotropic *B* factors) of the backbone atoms along the peptide chain. Residues in β-strands are shown in green, α-helices in blue and others in red.

the atoms modeled in two alternative conformations were refined, while the occupancies of the atoms in triple conformations were fixed at 0.33. The final values of $R_{\rm work}$ and $R_{\rm free}$ for the model with implemented riding H atoms were 0.117 and 0.159, respectively, for reflections with $|F_o| > 4\sigma(F_o)$ and were 0.124 and 0.169, respectively, for all reflections in the working data set. In the last 30 cycles of refinement, the working and validation sets were merged and all reflections were used, resulting in an *R* factor of 0.118 for reflections with $|F_o| > 4\sigma(F_o)$ and 0.124 for all 68 700 reflections. In order to calculate the estimated standard uncertainties, a cycle of blocked-matrix least-squares refinement with fixed parameters was performed. The protein molecule was divided into nine blocks of 22 residues (except the last block), each with a two-residue overlap with the previous one.

Data-collection and refinement statistics are summarized in Table 1. Structural analysis was carried out with the programs *PROCHECK* (Laskowski *et al.*, 1993), *EdPDB* (Zhang & Matthews, 1995) and *PARVATI* (Merritt, 1999).

3. Results and discussion

3.1. Overall structure

The crystal structure of the GTPase domain of human Rab5a has been refined with anisotropic displacement parameters using low-temperature synchrotron diffraction data extending to 1.05 Å resolution. The overall structure (Fig. 1), which possessed a typical Ras-like GTPase fold (Schweins & Wittinghofer, 1994; Vetter & Wittinghofer, 2001), extended the results of previous crystallographic studies on Rab5 (Merithew *et al.*, 2001; Zhu *et al.*, 2003). The following secondary-structure elements were determined from the current crystal structure: β-strand $\beta1$ consisted of residues 17–26, $\beta2$ 55–64, $\beta3$ 67–76, $\beta4$ 95–100, $\beta5$ 127–133 and $\beta6$ 158–161; helix $\alpha1$ consisted of residues 33–42, $\alpha2$ 79–90 (partly 3_{10}-helical and partly α-helical), $\alpha3$ 105–121, $\alpha4$ 145–154 and $\alpha5$ 170–179. The overall positional uncertainties estimated from blocked-matrix refinement with *SHELX*97 were 0.032 Å for all non-H protein atoms, 0.022 Å for main-chain atoms and 0.041 Å for side-chain atoms (Fig. 2). Residues 15−184 of the recombinant Rab5a protein could be located from the final electron-density map, an improvement on the previously reported structures. As a result, the final structure is composed of all 170 residues of the recombinant protein, one GppNHp, one Mg^{2+} ion, five glycerol and 205 water molecules, including eight water molecules modeled in double positions. Three residues were modeled with triple side-chain rotamers, while 22 more residues, including a catalytically important residue, Gln79, showed two rotamers (Table 2). Of these, residues Leu23, Leu25 and Val37 displayed cooperative conformations in the packing of a hydrophobic core. Furthermore, the main-chain atoms of residues Leu25, Leu36, Phe48 and Tyr89 showed double or triple conformations, of which Tyr89 is proposed to be involved in Rab5-specific effector binding (Merithew *et al.*, 2001). While no peptide backbone was found to have a disallowed conformation, Asp65 possessed normally less favorable backbone torsion angles (φ, $\psi = 65, -125^\circ$). This residue was located in the $\beta2$–$\beta3$ connection loop. Its average main-chain *B* factor was 46 Å^2, more than double of that of

Table 2
Residues with alternative conformations.

Residue	Ras No.	Atoms	Occupancy	Location
Ile18	1	$C^{\gamma1}$, $C^{\gamma2}$, $C^{\delta1}$	0.68, 0.32	
Cys19	2	S^{γ}	0.63, 0.37	
Leu23	6	C^{β} and beyond	0.63, 0.37	Hydrophobic core
Leu25	8	Whole residue	0.33, 0.33, 0.33	Hydrophobic core
Leu36	19	Whole residue	0.74, 0.26	
Val37	20	$C^{\gamma1}$, $C^{\gamma2}$	0.81, 0.19	Hydrophobic core
Phe48	31	Whole residue	0.56, 0.44	
Ser51	34	O^{γ}	0.33, 0.67	SWI
Thr59	41	$C^{\gamma2}$,$O^{\gamma1}$	0.33, 0.33, 0.33	
Gln79	61	C^{γ} and beyond	0.50, 0.50	SWII
Met88	70	C^{ε}	0.06, 0.94	
Tyr89	71	Whole residue	0.52, 0.48	Effector-binding site†
Arg91	73	C^{γ} and beyond	0.41, 0.59	
Gln94	76	C^{γ} and beyond	0.39, 0.61	
Glu106	88	C^{γ} and beyond	0.62, 0.38	
Arg120	102	C^{γ} and beyond	0.33, 0.33, 0.33	
Ser123	105	O^{γ}	0.47, 0.53	
Gln146	128	C^{γ} and beyond	0.52, 0.48	
Glu147	129	C^{γ} and beyond	0.57, 0.43	
Ser150	132	O^{γ}	0.62, 0.38	
Ser156	138	O^{γ}	0.60, 0.40	
Leu158	140	C^{γ} and beyond	0.54, 0.46	
Met168	150	C^{γ} and beyond	0.44, 0.56	
Lys180	162	C^{γ} and beyond	0.45, 0.55	

† Merithew *et al.* (2001).

the overall backbone; nevertheless, the corresponding electron density was well defined. Omitting the residues of multiple main-chain conformations, the C^{α} atom root-mean-square deviation (r.m.s.d.) was 0.18 Å between the current and previous 1.5 Å Rab5a structures (PDB code 1n6h). The positions of Rab5 H atoms could not be identified from a difference electron-density map at the current resolution.

3.2. Structural analysis

3.2.1. Anisotropic displacement analysis. One advantage of refinement at atomic resolution is the ability to determine the anisotropic displacement parameters for individual atoms. In our case, introducing anisotropic refinement resulted in a drop of 5.5% in $R_{\rm work}$, accompanied by a 0.1 Å r.m.s. shift of C^{α} atoms and a 3 Å^2 increase in the average $B_{\rm iso}$. The average value of the anisotropy factors (*i.e.* the ratio of the minimum to the maximum eigenvalues of anisotropic displacement, ranging from 0.0 to 1.0) of the final refined Rab5 model was 0.55 and 0.47 for main-chain and side-chain atoms, respectively (Fig. 2*b*). The overall value was 0.51, slightly higher than the average anisotropy of 0.45 previously reported for anisotropically refined structures in the PDB (Merritt, 1999). Interestingly, protein atoms that were more anisotropic (*i.e.* with a lower anisotropy factor) were more concentrated on one side of the protein molecule (*i.e.* the front, right and lower portions in Fig. 1), which contained the N- and C-termini and the switch I and II regions, consistent with their potential conformational changes during membrane association and GTP hydrolysis.

3.2.2. Peptide-bond planarity. Excluding the terminal peptides from both termini and residues that had alternate backbone conformations, the peptide torsion angles ω (*i.e.* $C^{\alpha}-C-N-C^{\alpha}$), were found to range between 163 and 197°. The average ω and its σ were 178.7 ± 6.1°, which agree well with the statistics from surveys of small-molecule databases (MacArthur & Thornton, 1996) and other high-resolution protein crystal structures (Sandalova *et al.*, 1999). Of the 19 peptide planes in Rab5 that had an ω angle more than 10° away from planarity, two (Val41 and Arg120) were located at C-termini of helices, while 11 were in β-strands, most of which had ω angles smaller than 170°. These observations are consistent with the notion that α-helices, which are more likely to contain regular hydrogen bonds, tend to maintain better planarity than other types of secondary structure (MacArthur & Thornton, 1996). Another measurement of the peptide-bond planarity is the so-called pyramidalization angle $\theta_{\rm N}$ (Ramachandran & Kolaskar, 1973), *i.e.* the angle between the planes defined by the peptide atoms $C_iN_{i+1}H_{i+1}$ and $C_iN_{i+1}C^{\alpha}_{i+1}$. Since most H-atom positions could not be determined crystallographically at the current resolution, we chose an alternative torsion angle $\theta_{\rm C}$, defined by the planes of $C^{\alpha}_iC_iN_{i+1}$ and $O_iC_iN_{i+1}$, to measure the extent of pyramidalization. In the current structure, $\langle\theta_{\rm C}\rangle \pm \sigma$ was 0.4 ± 2.6°, very close to the planar value of zero. Nevertheless, six residues (32, 48, 106, 134, 141 and 145) possessed a $\theta_{\rm C}$ more than 2σ away from the average, including two (residues 106 and 145) in helices. Only one of them (Arg141) was also in the above-mentioned 19 peptide planes that had a large $\Delta\omega$ from the planar value, suggesting that the two types of deviations are caused by different physical reasons. While the large $\Delta\omega$ can usually be rationalized by local backbone strain, we found that in the cases of large $\theta_{\rm C}$, either the carbonyl O or amide N atoms were involved in hydrogen bonding.

3.2.3. Short hydrogen bond. Near the phosphate-binding loop (P-loop, *i.e.* residues 27–34), on the side opposite to the nucleotide-binding site, two surface glutamate residues, 28 and 106, formed one 2.40 Å (O—O distance) short hydrogen bond between one pair of their side-chain carboxyl O atoms. Neither of them participated in a direct crystal-packing interaction (*i.e.* within 4.0 Å). The Glu106 side chain assumed a double conformation. A short strong hydrogen bond, in which a proton is shared equally between two electronegative groups, is more likely to be formed when the p*K* values of the two participating groups are similar (Cleland, 1992), which is likely to be true in our case. Because of this strict requirement, short hydrogen bonds remain rare even in high-resolution protein crystal structures. Although such hydrogen bonds have been proposed to play functional roles in some cases (Tronrud *et al.*, 1992; Usher *et al.*, 1994), potential functions of the short hydrogen bond in Rab5a are unclear, since the participating groups are not conserved even among Rab5 isoforms.

3.2.4. C—H···O hydrogen bonds. C—H···O hydrogen bonds have been proposed to contribute to protein stability (Derewenda *et al.*, 1995) and to be involved in protein–protein (Jiang & Lai, 2002) and protein–nucleotide interactions (Mandel-Gutfreund *et al.*, 1998) and their existence has been verified in high-resolution protein crystal structures. C—H···O hydrogen bonds occur more frequently between

Table 3
C—H···O hydrogen-bond distribution compared with N—H···O hydrogen-bond distribution.

	d_H (Å)	d (Å)	ζ (°)	ξ (°)	θ (°)
Total C—H···O (49)†‡	2.45 ± 0.12	3.27 ± 0.12	142 ± 11	136 ± 12	7 ± 26
Backbone C^{α}—H···O═C (27)†	2.45 ± 0.08	3.27 ± 0.10	142 ± 9	138 ± 11	−2 ± 21
β-Sheet (22)†	2.43 ± 0.08	3.28 ± 0.10	145 ± 7	137 ± 10	0 ± 20
Parallel β-strands (17)†	2.42 ± 0.08	3.27 ± 0.11	145 ± 7	134 ± 10	−3 ± 19
α-Helices (0)†					
Total N—H···O (96)‡§	2.11 ± 0.10	2.93 ± 0.09	161 ± 9	149 ± 12	−8 ± 22
Backbone N—H···O═C (77)§	2.10 ± 0.09	2.93 ± 0.09	162 ± 9	151 ± 11	−11 ± 21
β-Sheet (32)§	2.07 ± 0.09	2.91 ± 0.08	165 ± 8	156 ± 11	1 ± 18
Parallel β-strands (24)§	2.08 ± 0.09	2.91 ± 0.08	165 ± 9	155 ± 10	0 ± 19
α-Helices (30)§	2.14 ± 0.09	2.95 ± 0.09	159 ± 7	150 ± 5	−24 ± 8

† In parentheses are the number of potential bonds that satisfy the following selection criteria: d_H (H···O) < 2.6 Å, d (C···O) no explicit limit, ζ (C—H···O) > 120°, ξ (H···O═C) between 90 and 150° and θ (elevation angle) between −45 and 45° (Derewenda *et al.*, 1995). ‡ Of O atoms, only carboxyl O and peptide carbonyl O atoms were included in this calculation. [c.] § Same as C—H···O criteria, except that d_H (H···O) < 2.4 Å.

C^{α}—H and carbonyl O atoms, presumably because of the more acidic property of C^{α} atoms (Senes *et al.*, 2001). Hydrogen-bond formation is different to a general electrostatic interaction in that directionality plays an important role. With a set of rather strict geometric criteria (see legend to Table 3), we found 27 main chain–main chain C^{α}—H···O═C hydrogen bonds, compared with 77 main chain–main chain conventional N—H···O═C hydrogen bonds. The geometrical statistics of the C—H···O hydrogen bonds in Rab5a are summarized in Table 3, accompanied by those for N—H···O hydrogen bonds. Our data are comparable with data extracted from a more general collection (Derewenda *et al.*, 1995). Of the C—H···O hydrogen bonds, 22 were between β-strands, including 17 between parallel β-strands (five of the six β-strands in Rab5 are parallel to each other). Although not all carbonyl O and C^{α} atoms of the central β-sheet of Rab5 formed this type of bond, it appears to us that the backbone C—H···O═C hydrogen bond is more likely to be enforced by the overall geometry of the β-sheet secondary structure, which is dictated by the stronger main chain–main chain N—H···O═C hydrogen bonds. In all cases where main-chain O atoms were engaged in a C^{α}—H···O hydrogen bond, they also formed a stronger interaction with another proton donor. Given their weaker bond strength (one third to half of that of the conventional hydrogen bond; Jiang & Lai, 2002) and less frequent occurrence in the protein structure, the contribution of C—H···O hydrogen bonds to the overall stability of the globular protein is likely to be less significant than that of conventional hydrogen bonds, although they may provide some interaction specificity because of the directionality of the bonds (Senes *et al.*, 2001).

3.2.5. Solvent structure. The current crystal form of Rab5a contained about 39%(*v*/*v*) solvent, with a V_M value of 2.0 Å^3 Da^{-1} (Matthews, 1968), which translated into ∼490 water molecules per asymmetric unit, if the density of bulk water is assumed to be the same inside and outside the crystal (*i.e.* 1 g cm^{-3} or ∼0.033 water molecules per Å^3). Of the 205 ordered water molecules built into the final refined model, 14 were found to be completely buried inside the protein molecule, *i.e.* with less than 1% solvent-accessible surface exposed to bulk solvent calculated in the presence of protein atoms but in the absence of other potential solvent molecules. Two of them coordinate the Mg^{2+} cofactor. Between the central β-sheet and helices α2 and α3, near the switch II region (residues 77–80) and its interface to switch I (residues 49–51), there were seven buried (or partially buried) water molecules, forming the only extensive network of buried water molecules in Rab5. Most of these water molecules were hydrogen bonded to main-chain polar groups; an exception is one water molecule that formed a hydrogen bond to a buried hydroxyl group from Thr76, which is part of the DTAGQE motif conserved among Ras-like GTPase proteins (Liang *et al.*, 2000). A similar but not identical buried water network is also observed in the Ras crystal structure (PDB code 1ctq). These water molecules are likely to play roles in the switching process in response to GTP binding and/or hydrolysis. High-resolution crystal structures of GDP-bound (inactive) forms of Ras-like GTPases would be desirable in order to address this question. Furthermore, 175 water molecules were found in the first hydration shell (using a 3.5 Å cutoff from any protein or GppNHp atoms), excluding the 14 buried waters; this number increased to 249 if multiple occurrences of the same water molecule were allowed based on crystal symmetry. Table 4 summarizes the solvent-accessible surface (SAS) of the protein molecule and that shielded by ordered water molecules as well as by crystal packing. About 60% of the protein SAS was shielded from bulk solvent by ordered water molecules. About 40% of SAS was shielded from bulk solvent by symmetry-related protein neighbors, which overlapped with the SAS shielded by ordered water molecules. About one fifth of the protein SAS was not covered by either ordered water molecules or atoms from neighboring protein molecules. In any case, the ratio of hydrophobic to hydrophilic buried SAS is about half:half, consistent with a previous study on the high-resolution crystal structure of dethiobiotin synthetase (Sandalova *et al.*, 1999).

3.2.6. Active site. One conserved structural feature of small GTPases is the P-loop, which directly interacts with the β- and γ-phosphate groups of the substrate GTP. The peptide amide group of Rab5 Ala30 in the P-loop (Fig. 3) or its equivalent in other GTPases has been hypothesized to play a critical role in the dissociative-intermediate catalysis mechanism (Maegley *et al.*, 1996; Zhu *et al.*, 2003). This mechanism predicts that the GTPase stabilizes the accumulation of negative charge at the β–γ bridge oxygen in the transition intermediate through an important hydrogen bond between this O atom and the peptide amide group at position 30 in the P-loop. In the current structure, the β–γ bridge O atom of the nucleotide was replaced with an imino group, thus becoming non-hydrolysable. Nevertheless, the N atom assumes the same tetrahedral geometry as the O atom by forming hydrogen bonds with both the Ala30 amide group and a water molecule. The hydrogen bond with the Ala30 amide group

Table 4
Solvent-accessible surface of Rab5a.

Lee and Richards solvent-accessible surface (SAS, in Å^2; Lee & Richards, 1971) was calculated with the program *EdPDB* using a 1.4 Å probe radius. All calculation was performed without explicit H atoms. Glycerol molecules are included in the protein-neighbor category.

	Total†	Hydrophobic‡	Hydrophilic‡
Protein SAS	8430	4090 (49%)	4330 (51%)
Buried by ordered water molecule	5230 (62%)	2490 (48%)	2740 (52%)
Buried by protein neighbors	3620 (43%)	1790 (50%)	1830 (50%)
Buried by either water or neighbors	6750 (80%)	3320 (49%)	3430 (51%)

† Percentages are of the total protein SAS. Note that there was an overlap between SAS buried by ordered water molecules and that buried by neighboring protein (or glycerol) molecules. Protein SAS buried by internal water molecules was negligible (<40 Å^2). ‡ Percentage of the total of each category (*i.e.* the same row).

exhibited good geometry [d_H = 2.22 Å, $d_{(N-N)}$ = 3.04 Å, $\zeta_{(N-H\cdots N)}$ = 159°, $\xi_{(P\gamma-N\cdots H)}$ = 108°]; the hydrogen bond to the water molecule had a N—O distance of 2.86 Å, with the longest axis of the anisotropic ellipsoid of the water molecule, which had an anisotropy factor of 0.36, perpendicular to the hydrogen bond. Another catalytically important water molecule, located close to a γ-phosphate O atom of the substrate, serves as a nucleophile in the hydrolytic catalysis and presumably attacks the γ-phosphate from the direction collinear with the $^{\beta}O^{\gamma}—P^{\gamma}$ axis (Milburn *et al.*, 1990; Pai *et al.*, 1990). In the current structure, this water molecule formed one hydrogen bond to a O^{γ} atom (2.83 Å O—O distance) and four more potential hydrogen bonds to protein atoms, *i.e.* the carbonyl O atom of Thr52 (3.11 Å), the hydroxyl (3.19 Å) and the backbone amide (3.14 Å) groups of Ser51, and $N^{\varepsilon 2}$ (3.26 Å) of Gln79, which could not be satisfied simultaneously because of the directionality requirement of hydrogen bonding. The shortest axis of the anisotropic ellipsoid of the water molecule (with an anisotropy factor of 0.48) was collinear with the shortest hydrogen bond. This water molecule does not superimpose well with the corresponding water in the Ras crystal structure (they are 1.76 Å away from each other after overall structural superposition) and in neither case is it positioned on the putative reaction axis. Although the side chain of Gln79 is considered to be important for aligning the nucleophilic water molecule during catalysis (Milburn *et al.*, 1990; Pai *et al.*, 1990), in the current structure the hydrogen bond between $N^{\varepsilon 2}$ and the nucleophilic water molecule is weak [d_{N-O} = 3.26 Å, $\xi_{(C-N\cdots O)}$ = 95.81°]. Thus, alignment of the nucleophilic water molecule does not seem to occur spontaneously, but requires some pre-adjustments in other parts of the protein–ligand complex structure. This is consistent with the typically low intrinsic GTPase-hydrolysis rate of small GTPases.

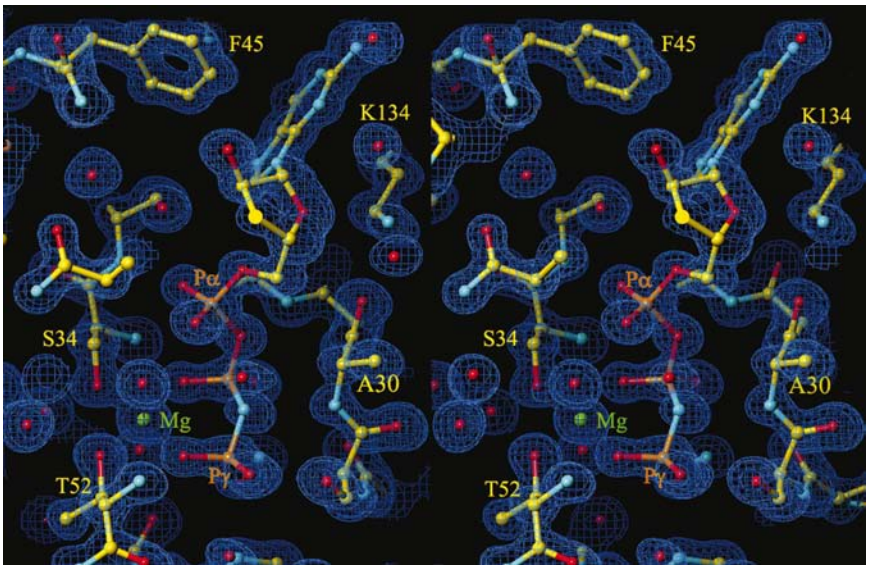

Figure 3
A representative region of the electron-density map. The $2m|F_o| - D|F_c|$ electron-density map phased with the final refined model is contoured at the 1.5σ level and superimposed with a ball-and-stick model of the structure. Coloring corresponds to atom types: red for oxygen, blue for nitrogen, yellow for carbon, brown for phosphate and green for magnesium. Selected residues, Mg^{2+} ion and GppNHp are labeled.

3.3. Comparison with the high-resolution Ras structure

As the second reported atomic resolution crystal structure of the Ras-like GTPase family, Rab5 was compared with the 1.26 Å resolution Ras crystal structure (PDB code 1ctq). Rab5 has two single-residue insertions in the $\beta 2$–$\beta 3$ and $\beta 5$–$\alpha 5$ connection loops and one single-residue deletion in the $\alpha 3$–$\beta 5$ connection loop in its primary sequence compared with Ras. Consistent with their conserved primary sequences, the overall crystal structure r.m.s.d. was 0.93 Å for 146 C^{α} atom pairs using a 2.0 Å cutoff. Bearing the same biochemical function, the two structures essentially share an identical GTP-binding site. By simple superposition of their GppNHp (substrate-analog) molecules, the two structures could be brought together with an r.m.s.d. of 0.96 Å for 142 C^{α} atom pairs, indicating a very similar geometry in their GTP-binding pockets. The two GTP-analog molecules superimposed well on each other (0.16 Å r.m.s.d.), with the only noticeable difference, an ∼10° rotation about the $^{\beta}N^{\gamma}—P^{\gamma}$ bond, occurring between their γ-phosphate groups. This rotation appeared to be caused by the hydrogen bonds between the O^{γ} atoms and the serine side chains of residues 29 and 51 of Rab5, which correspond to Gly12 and Pro34 in Ras, respectively. In both structures, the GTP-binding specificity for the most closely related common nucleotide triphosphate, ATP, was determined by a hydrogen bond between GTP O_6 and the amide group of Ala164 (Rab5a) and two hydrogen bonds between the N_1/N_2 groups of GTP and the side-chain carboxyl groups of Asp136 (Rab5a), where the GTP groups are different from their ATP counterparts. Of the total of 18 hydrogen bonds between GTP and Rab5a (using a 3.1 Å cutoff for distances between polar groups), 16 are also observed in Ras, forming a conserved and extensive interaction

network, while the two missing ones are caused by Ser12 to Gly and Ser18 to Ala substitutions in Ras. In small GTPases, two regions, named switches I and II, typically change their conformations in response to the hydrolytic status of the bound nucleotide, thus conveying the information to the downstream effectors (Vetter & Wittinghofer, 2001). As expected from their difference in selecting regulators and effectors, the switch regions of Rab5 and Ras were among the regions of largest structural deviation after an overall structural superposition. In total, seven regions in Rab5 were found to have C^{α} r.m.s.d.s larger than 1.40 Å compared with Ras, including the N- and C-termini, switch I (residues 44–53, 1.55 Å r.m.s.d.), β2 and β3 (residues 59–71, 2.09 Å), switch II and α2 (residues 80–93, 1.47 Å), α3 and the following loop (residues 109–126, 1.48 Å) and the α4–β6 connecting loop (residues 154–159, 1.41 Å).

4. Conclusions

The atomic resolution crystal structure of Rab5 provides us with unprecedented details of the Ras-like GTPase structure in the GTP-bound form. It allows us to analyze, for example, the internal water network and anisotropic *B*-factor distribution in the protein structure. This structure confirmed with high certainty the previous observation that the catalytic apparatus of the protein has not been optimized for fast turnover, consistent with the timer function of this protein family in regulating molecular processes (Li & Qian, 2002; Rybin *et al.*, 1996).

The authors thank N. Wakeham for technical assistance, the staff of the F1 beamline at the CHESS synchrotron facility for help in data collection and Drs T. Mather and G. Air for helpful comments. This work was supported in part by a CAREER Award from the National Science Foundation to GL.

References

Bourne, H. R., Sanders, D. A. & McCormick, F. (1991). *Nature (London)*, **349**, 117–127.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Bucci, C., Lutcke, A., Steele-Mortimer, O., Olkkonen, V. M., Dupree, P., Chiariello, M., Bruni, C. B., Simons, K. & Zerial, M. (1995). *FEBS Lett.* **366**, 65–71.
Bucci, C., Parton, R. G., Mather, I. H., Stunnenberg, H., Simons, K., Hoflack, B. & Zerial, M. (1992). *Cell*, **70**, 715–728.
Cleland, W. W. (1992). *Biochemistry*, **31**, 317–319.
Derewenda, Z. S., Lee, L. & Derewenda, U. (1995). *J. Mol. Biol.* **252**, 248–262.
Gorvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. (1991). *Cell*, **64**, 915–925.
Jiang, L. & Lai, L. (2002). *J. Biol. Chem.* **277**, 37732–37740.
Lanzetti, L., Rybin, V., Malabarba, M. G., Christoforidis, S., Scita, G., Zerial, M. & Di Fiore, P. P. (2000). *Nature (London)*, **408**, 374–377.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, B. & Richards, F. M. (1971). *J. Mol. Biol.* **55**, 379–400.
Li, G. (1996). *Biocell*, **20**, 325–330.
Li, G. & Liang, Z. (2001). *Biochem. J.* **355**, 681–689.
Li, G. & Qian, H. (2002). *Traffic*, **3**, 249–255.
Li, G. & Stahl, P. D. (1993). *Arch. Biochem. Biophys.* **304**, 471–478.
Liang, Z., Mather, T. & Li, G. (2000). *Biochem. J.* **346**, 501–508.
Liu, K. & Li, G. (1998). *J. Biol. Chem.* **273**, 10087–10090.
MacArthur, M. W. & Thornton, J. M. (1996). *J. Mol. Biol.* **264**, 1180–1195.
Maegley, K. A., Admiraal, S. J. & Herschlag, D. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 8160–8166.
Mandel-Gutfreund, Y., Margalit, H., Jernigan, R. L. & Zhurkin, V. B. (1998). *J. Mol. Biol.* **277**, 1129–1140.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merithew, E., Hatherly, S., Dumas, J. J., Lawe, D. C., Heller-Harrison, R. & Lambright, D. G. (2001). *J. Biol. Chem.* **276**, 13982–13988.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1997–2004.
Milburn, M. V., Tong, L., deVos, A. M., Brunger, A., Yamaizumi, Z., Nishimura, S. & Kim, S.-H. (1990). *Science*, **247,** 939–945.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W. & Wittinghofer, A. (1990). *EMBO J.* **9**, 2351–2359.
Pereira-Leal, J. B. & Seabra, M. C. (2000). *J. Mol. Biol.* **301**, 1077–1087.
Ramachandran, G. N. & Kolaskar, A. S. (1973). *Biochim. Biophys. Acta*, **303**, 385–388.
Roussel, A. & Cambillau, C. (1989). *Silicon Graphics Geometry Partners Directory*, pp. 77–79. Silicon Graphics, Mountain View, CA, USA.
Rybin, V., Ullrich, O., Rubino, M., Alexandrov, K., Simon, I., Seabra, M. C., Goody, R. & Zerial, M. (1996). *Nature (London)*, **383**, 266–269.
Sandalova, T., Schneider, G., Kack, H. & Lindqvist, Y. (1999). *Acta Cryst.* D**55**, 610–624.
Schweins, T. & Wittinghofer, A. (1994). *Curr. Biol.* **4**, 547–550.
Senes, A., Ubarretxena-Belandia, I. & Engelman, D. M. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 9056–9061.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **227**, 319–343.
Simon, I., Zerial, M. & Goody, R. S. (1996). *J. Biol. Chem.* **271**, 20470–20478.
Tronrud, D. E., Roderick, S. L. & Matthews, B. W. (1992). *Matrix Suppl.* **1**, 107–111.
Usher, K. C., Remington, S. J., Martin, D. P. & Drueckhammer, D. G. (1994). *Biochemistry*, **33**, 7753–7759.
Vetter, I. R. & Wittinghofer, A. (2001). *Science*, **294**, 1299–1304.
Xiao, G. H., Shoarinejad, F., Jin, F., Golemis, E. A. & Yeung, R. S. (1997). *J. Biol. Chem.* **272**, 6097–6100.
Zhang, X. & Matthews, B. W. (1995). *J. Appl. Cryst.* **28**, 624–630.
Zhu, G., Liu, J., Terzyan, S., Zhai, P., Li, G. & Zhang, X. C. (2003). *J. Biol. Chem.* **278**, 2452–2460.