# Definition of the Switch Surface in the Solution Structure of Cdc42Hs[†,‡]

Joanna L. Feltham,[§,||] Volker Dötsch,[⊥] Sami Raza,[#] Danny Manor,[§] Richard A. Cerione,[§] Michael J. Sutcliffe,[#] Gerhard Wagner,[⊥] and Robert E. Oswald*[,§,⊥]

*Department of Pharmacology, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, and Department of Chemistry, University of Leicester, Leicester LE1 7RH, U.K.*

*Received March 25, 1997; Revised Manuscript Received May 5, 1997*[⊗]

ABSTRACT: Proteins of the rho subfamily of ras GTPases have been shown to be crucial components of pathways leading to cell growth and the establishment of cell polarity and mobility. Presented here is the solution structure of one such protein, Cdc42Hs, which provides insight into the structural basis for specificity of interactions between this protein and its effector and regulatory proteins. Standard heteronuclear NMR methods were used to assign the protein, and approximately 2100 distance and dihedral angle constraints were used to calculate a set of 20 structures using a combination of distance geometry and simulated annealing refinement. These structures show overall similarity to those of other GTP-binding proteins, with some exceptions. The regions corresponding to switch I and switch II in H-ras are disordered, and no evidence was found for an α-helix in switch II. The 13-residue insertion, which is only present in rho-subtype proteins and has been shown to be an important mediator of binding of regulatory and target proteins, forms a compact structure containing a short helix lying adjacent to the $\beta$4−α3 loop. The insert forms one edge of a "switch surface" and, unexpectedly, does not change conformation upon activation of the protein by the exchange of GTP analogs for GDP. These studies indicate the insert region forms a stable invariant "footrest" for docking of regulatory and effector proteins.

The rho subfamily (Barbacid, 1987; Boguski & McCormick, 1993; Bourne *et al.*, 1991) of ras-like GTP-binding[1] proteins has been shown in recent years to control various cellular signaling pathways leading to changes in cell morphology and polarity, as well as increased DNA synthesis and cell cycle progression (Chant & Stowers, 1995; Ridley, 1995; Vojtek & Cooper, 1995). Initial indications of rho family involvement in the regulation of cytoskeletal architecture and cell polarity came primarily from studies of homologous proteins in yeast, where Cdc42 is a crucial facilitator of bud site assembly (Adams *et al.*, 1990; Johnson & Pringle, 1990). Microinjection studies in fibroblasts have shown rho proteins to be involved in similar cytoskeletal reorganizations in higher cell types, although the various family members appear to act downstream of different extracellular signals and lead to distinct cytoskeletal responses (Kozma *et al.*, 1995; Nobes & Hall, 1995). Microinjection of Cdc42Hs, for instance, mimics the fibroblast response to bradykinin by generating actin-containing membrane protrusions called filopodia. Similarly, Rac1 acts downstream of growth factors such as PDGF to cause membrane ruffles called lamellipodia, and RhoA microinjection results in the formation of stress fibers and focal adhesion plaques. These studies have also shown that, at least in fibroblasts, the rho GTPases are capable of regulating each other in a signaling cascade where Cdc42Hs activates Rac1, which in turn activates RhoA.

More recently, however, it has become apparent that regulation of cytoskeletal architecture is only one function of the rho subfamily members. In fact, these proteins have long been suggested to be enhancers of cell growth and division, based on the fact that the *dbl* oncogene product is capable of stimulating nucleotide exchange on both Rho and Cdc42Hs (Hart *et al.*, 1994). Since common domains of the *dbl* protein are required for both its nucleotide exchange and cell transformation capabilities, rho subfamily proteins may be responsible for mediating the unregulated growth signal from *dbl*. Cdc42Hs and Rac1 have since been shown to be capable of activating the mitogen-activated protein kinases JNK and p38 *via* direct stimulation of the p21-activated serine/threonine kinases (PAKs; Bagrodia *et al.*,

[†] Supported by a grant from the American Cancer Society and the John Simon Guggenheim Memorial Foundation to R.E.O. and a grant from the NIH (GM38608) to G.W. J.L.F. was supported by an NSF graduate research fellowship and an NIH predoctoral training grant (T32GM08210). V.D. was supported in part by an EMBO fellowship. M.J.S. is a Royal Society University Research Fellow. S.R. was supported by a Leicester University Research Studentship.

[‡] The coordinates of the 20 final structures have been deposited in the Brookhaven Protein Data Bank as entry 1AJE.

* Author to whom correspondence should be addressed. Phone: 607-253-3877. Fax: 607-253-3659. E-mail: reo1@cornell.edu.

[§] Cornell University.

[||] Present address: Department of Chemistry, University of Massachusetts, Amherst, MA 01003.

[⊥] Harvard Medical School.

[#] University of Leicester.

[⊗] Abstract published in *Advance ACS Abstracts,* July 1, 1997.

[1] Abbreviations: DSS, 3-(trimethylsilyl)-1-propanesulfonic acid; DTT, DL-dithiothreitol; EDTA, ethylenediaminetetraacetic acid; GAP, GTPase activating protein; GDI, GDP dissociation inhibitor; GEF, guanine nucleotide exchange factor; GDP, guanosine 5′-diphosphate; GMPPCP, $\beta,\gamma$-methylene derivative of GTP; GMPPNP, $\beta,\gamma$-imido derivative of GTP; GSH, glutathione; GST, glutathione *S*-transferase; GTP, guanosine 5′-triphosphate; HEPES, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; HMQC, heteronuclear multiple-quantum correlation; HSQC, heteronuclear single-quantum correlation; IPTG, isopropyl $\beta$-D-thiogalactopyranoside; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser effect spectroscopy; PAK, p21 activated kinase; PDGF, platelet-derived growth factor; PMSF, phenylmethanesulfonyl fluoride; sNBD, succinimidyl 6-[(7-nitrobenz-2-oxa-1,3-diazol-4-yl)amino]hexanoate; Tris-HCl, tris(hydroxymethyl)-aminomethane hydrochloride; TOCSY, total correlation spectroscopy.

S0006-2960(97)00694-6 CCC: $14.00 

1995; Coso *et al.*, 1995; Minden *et al.*, 1995; Zhang *et al.*, 1995). Cdc42Hs apparently can act downstream of the cytokine IL-1 to this end. It is not yet clear how or if these signals to the nucleus relate to the cytoskeletal rearrangements also controlled by rho GTPases.

As with all G proteins, the activity of rho proteins is controlled by the type of nucleotide bound in the active site (Barbacid, 1987; Bourne *et al.*, 1991). With GDP bound, they are quiescent or nonsignaling. Exchange of the GDP for GTP activates the protein by causing structural rearrangements which allow it to interact with downstream effectors. The signal from the active protein is eventually turned off by an intrinsic GTP hydrolytic activity, resulting in a return to the inactive GDP-bound state. In the case of Cdc42Hs, the active protein can interact with a number of effectors in addition to the PAK serine/threonine kinase already mentioned, including the $p120^{ACK}$ tyrosine kinase (Manser *et al.*, 1993), Wiskott−Aldrich syndrome protein (WASP; Symons *et al.*, 1996), IQGAP's 1 and 2 (Brill *et al.*, 1996; Hart *et al.*, 1996; Kuroda *et al.*, 1996; McCallum *et al.*, 1996), $pp70^{S6K}$ (Chou & Blenis, 1996), phospholipase D (Singer *et al.*, 1995), kinectin (Hotta *et al.*, 1996), and the p85 regulatory subunit of PI-3 kinase (Zheng *et al.*, 1994). These effector proteins in turn participate in a variety of different cellular functions.

The GTPase cycle of rho proteins is controlled by several classes of regulatory proteins. The rate of activation by exchange of bound GDP for GTP is enhanced through interaction with upstream signaling molecules known as guanine nucleotide exchange factors (GEFs), which catalyze the exchange of nucleotide. Both *ost* and *dbl* oncogene products are capable of stimulating nucleotide exchange on Cdc42Hs (Hart *et al.*, 1991a; Horii *et al.*, 1994). The rate of GTP hydrolysis can be enhanced by interaction with a GAP, such as Cdc42GAP (Barfod *et al.*, 1993; Hart *et al.*, 1991b), which terminates the active signal. Additional regulation of the GTPase cycle of rho subfamily proteins arises from interaction with GDIs, which bind to both GTP- and GDP-bound proteins and are capable of solubilizing them from membranes as well as inhibiting GTP hydrolysis and GDP dissociation (Leonard *et al.*, 1992).

Extensive structural studies of H-ras (Kraulis *et al.*, 1994; Pai *et al.*, 1990; Tong *et al.*, 1991), Arf (Amor *et al.*, 1994; Greasley *et al.*, 1995), and Ran (Scheffzek *et al.*, 1995), as well as the heterotrimeric G proteins ($G_t$ and $G_i$; Lambright *et al.*, 1996; Wall *et al.*, 1995) and bacterial elongation factors (EF-Tu and EF-G; Aevarsson *et al.*, 1994; Czworkowski *et al.*, 1994; Polekhina *et al.*, 1996), have illustrated that the core GTP-binding domain is a conserved structural unit into which insertions can be built at a number of locations. In most cases, these structural insertions confer unique attributes upon each protein, such as dimerization of Arf proteins through an additional $\beta$-strand, and the GAP activity of the helical insertion (Gail) of heterotrimeric $G_\alpha$ subunits (Markby *et al.*, 1993). It is the unique structural features of each GTP-binding protein which result in specificity in its interactions with effector and regulatory proteins. Of particular interest in rho subfamily proteins is a 13-residue insertion which has been shown to be an important mediator of target binding as well as GDI function (Freeman *et al.*, 1996; McCallum *et al.*; 1996, Wu *et al.*, 1997). Currently, structural information about rho subfamily proteins includes a model of Cdc42Hs built on the basis of homology to H-ras (Sutcliffe *et al.*, 1994) and a crystal structure of a mutant Rac1 protein complexed with a GTP analog (Hirshberg *et al.*, 1997). Presented here are the results of an investigation into the solution structure of wild-type Cdc42Hs in both active and inactive conformations using NMR methods.

## EXPERIMENTAL PROCEDURES

*Protein Expression.* Cdc42Hs·GDP used in these experiments is expressed as a GST fusion protein in *Escherichia coli* from a pGEX-CDC42Hs expression plasmid (Shinjo *et al.*, 1990). Nonenriched protein samples were expressed in the TG1 strain of *E. coli* in M9 minimal medium (Muchmore *et al.*, 1989). For samples labeled homogeneously with $^{15}N$ or $^{15}N/^{13}C$, $^{15}NH_4Cl$ and [$^{13}C$]glucose were substituted for their unlabeled counterparts in the M9 media. Proteins selectively labeled with specific [$^{15}N$]amino acids (Tyr, Phe, Leu, Val, Ala, Lys) were expressed from the DL39 auxotrophic strain of *E. coli* (Muchmore *et al.*, 1989) in M9 minimal medium supplemented with L-Asp, L-Tyr, L-Phe, L-Leu, L-Ile, and L-Val. $^{15}N$-Labeled amino acids were substituted for their unlabeled counterparts (Tyr, Phe, Leu, and Val) or added (Lys, Ala) at the following concentrations: L-Tyr, L-Phe, and L-Lys, 100 $\mu$g/mL; L-Leu, 150 $\mu$g/mL; L-Val and L-Ala, 200 $\mu$g/mL. Protein was expressed using the following protocol: 500 mL of a cell culture grown to late log phase were used to inoculate 4 L of media in a Hi-Density Fermentor (Lab-Line Instruments, Melrose Park, IL). Cells were induced at mid-log phase with 0.3 mM IPTG and harvested 6 h after induction. Proteins expressed with a suppressed-labeling protocol for identification of amino acid type were expressed in the BL21 *E. coli* strain as described (Shortle, 1994), except a minimal medium was used which contained 2× M9 salts (with $^{15}NH_4Cl$), 2 mM $MgSO_4$, 1−4 g/L glycerol, 0.1 mM $CaCl_2 \cdot H_2O$, and 35 $\mu$g/mL ampicillin.

*Protein Purification*. All procedures were performed at 4 °C. Cells were harvested from a 4 L fermentor culture by centrifugation at 14300*g* for 10 min, washed in 100 mL of lysis buffer (20 mM Tris-HCl, pH 8.0, 6 mM EDTA, 1 mM DTT, 1 mM $NaN_3$, 0.1 mM GDP, 2 $\mu$g/mL aprotinin, leupeptin and pepstatin, and 10 $\mu$g/mL benzamidine and PMSF), and recentrifuged. Cell pellets were frozen in liquid nitrogen and stored at −80 °C until use. For preparation of cellular lysates, frozen pellets were thawed in 100 mL of lysis buffer and homogenized. Cells were treated with lysozyme (2 mg/mL) for 30 min, followed by the addition of $MgCl_2$ (50 mM) and DNase (2 mg/mL) and subsequent cell disruption with a Polytron homogenizer. Deoxycholate (40 mg) and an additional aliquot of PMSF (to 20 $\mu$g/mL final) were added, followed again by Polytron homogenization. The insoluble fraction was removed by centrifugation at 96000*g* for 45 min. After addition of a final aliquot of PMSF (to 30 $\mu$g/mL final), the supernatant was mixed with GSH−agarose beads by rocking for 1 h. The beads were subsequently extensively washed in a column with HMAG buffer (20 mM HEPES, pH 8.0, 5 mM $MgCl_2$, 1 mM $NaN_3$, and 0.1 mM GDP), and fusion protein was eluted in 10 mM glutathione in HMAG buffer at pH 8.0. The protein was cleaved with thrombin and applied to a MonoQ anion-exchange column. Pooled flow-through containing Cdc42Hs·GDP was concentrated and applied to a Sephadex G-25 gel filtration column preequilibrated with NMR buffer (5 mM $NaH_2PO_4$, pH 5.5, 25 mM NaCl, 5 mM $MgCl_2$, and 1 mM

Table 1: Acquisition Parameters for NMR Experiments

| | | | nucleus | | | no. of complex points | | | spectral width (ppm) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ligand | experiment | protein label | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_1$ | $\omega_2$ | $\omega_3$ |
| GDP | 2D TOCSY | unlabeled | $^1$H | $^1$H | | 962 | 2048 | | 16 | 16 | |
| | 2D NOESY | unlabeled | $^1$H | $^1$H | | 512−896 | 2048−4096 | | 16 | 16 | |
| | $^1$H, $^{15}$N HSQC | $^{15}$N | $^{15}$N | $^1$H | | 322−1024 | 2048 | | 33.6−59.2 | 16 | |
| | $^1$H, $^{13}$C, HSQC | 8−10% $^{13}$C | $^{13}$C | $^1$H | | 806 | 1024 | | 63 | 16 | |
| | SE HNCO | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 58 | 160 | 1024 | 47.4 | 31.8 | 16 |
| | SE HNCA | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 64 | 148 | 1024 | 47.4 | 47.8 | 16 |
| | SE HN(CO)CA | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 68 | 128 | 1024 | 47.4 | 47.8 | 16 |
| | SE CBCA(CO)NH | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 76 | 100 | 1024 | 33.6 | 60.9 | 16 |
| | HCACO | $^{13}$C, $^{15}$N | $^{13}$C | $^{13}$C | $^1$H | 62 | 150 | 1024 | 30.3 | 20.2 | 11.74 |
| | $^1$H, $^{15}$N TOCSY HSQC | | | | | | | | | | |
| | 30 ms mix | $^{15}$N | $^{15}$N | $^1$H | $^1$H | 64 | 144 | 1024 | 39.5 | 16 | 16 |
| | 50 ms mix | $^{15}$N | $^{15}$N | $^1$H | $^1$H | 54 | 144 | 1024 | 37.0 | 15 | 15 |
| | 70 ms mix | $^{15}$N | $^{15}$N | $^1$H | $^1$H | 64 | 144 | 1024 | 33.6 | 16 | 16 |
| | HCCH TOCSY | | | | | | | | | | |
| | 7 ms mix | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 116 | 144 | 1024 | 67.7 | 12 | 12 |
| | 14 ms mix | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 106 | 144 | 1024 | 33.4 | 12 | 12 |
| | 21 ms mix | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 58 | 176 | 1024 | 69.6 | 12 | 12 |
| | $^1$H, $^{15}$N NOESY HSQC | $^{15}$N | $^{15}$N | $^1$H | $^1$H | 60 | 144 | 1024 | 49.3 | 16 | 16 |
| | $^1$H, $^{13}$C NOESY HSQC | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 144 | 160 | 1024 | 19.9 | 16.67 | 16.67 |
| | aromatic $^1$H, $^{13}$C NOESY HSQC | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 90 | 200 | 1024 | 31.8 | 16.67 | 16.67 |
| | 4D $^1$H, $^{13}$C HMQC NOESY HMQC | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 34 | 32 | 100, 512 | 25 | 25 | 7; 7 |
| GMPPCP | $^1$H, $^{15}$N HSQC | $^{15}$N | $^{15}$N | $^1$H | | 256 | 2048 | | 47.4 | 16 | |
| | SE HNCO | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 62 | 144 | 1024 | 34.5 | 15.9 | 16.67 |
| | HNCA with $C_\beta$ decoupling | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 66 | 128 | 1024 | 36.2 | 26.5 | 16.67 |
| | HNCOCA with $C_\beta$ decoupling | $^{13}$C, $^{15}$N | $^{15}$N | $^{13}$C | $^1$H | 68 | 128 | 1024 | 36.2 | 26.5 | 16.67 |
| | $^1$H, $^{15}$N TOCSY HSQC | $^{15}$N | $^{15}$N | $^1$H | $^1$H | 64 | 144 | 1024 | 33.6 | 16 | 16 |
| | HCCH TOCSY | $^{13}$C, $^{15}$N | $^{13}$C | $^1$H | $^1$H | 120 | 144 | 1024 | 42.4 | 6.67 | 6.67 |
| | $^1$H, $^{15}$N NOESY HSQC | $^{15}$N | $^1$H | $^{15}$N | $^1$H | 256 | 64 | 1024 | 16 | 49.3 | 16 |

$NaN_3$) to effect buffer exchange and to remove residual glutathione. Purified Cdc42Hs·GDP was dialyzed extensively against NMR buffer and concentrated. Yields of pure protein varied widely according to the *E. coli* strain used, with TG1 providing about 6 mg/L of culture, and BL21 and DL39 only about 1 mg/L.

*Exchange of GDP for GMPPCP.* For experiments on the active conformation of Cdc42Hs, bound GDP was exchanged for the nonhydrolyzable GTP analog, GMPPCP, by a variation of the method in John *et al.* (1990). Specifically, 15 mg of Cdc42Hs·GDP was placed in a buffer containing 50 mM HEPES, pH 8.0, 20 mM EDTA, and a 10-fold molar excess of GMPPCP (relative to Cdc42Hs). This mixture was added to 150 $\mu$L of alkaline phosphatase beads and rocked for 3 h on ice, resulting in specific degradation of GDP. Removal of alkaline phosphatase by low-speed centrifugation was followed by the addition of $MgCl_2$ to 25 mM. This sample was then chromatographed on a Sephadex G-25 gel filtration column equilibrated in NMR buffer to remove excess unbound GMPPCP. The extent of the exchange reaction was verified by HPLC under ion-pairing conditions.

*Preparation of NMR Samples.* Protein samples for NMR were 0.5−0.6 mL in NMR buffer with $D_2O$ added to 10% for deuterium lock. The concentration of protein in unlabeled and homogeneously labeled samples was 0.8 mM, while selectively labeled and suppressed label samples were roughly 0.2 mM. Protein concentrations were assayed by Bradford assay (Bradford, 1976), which was normalized to the results of a quantitative amino acid analysis. Samples in 100% $D_2O$ were prepared by lyophilizing a sample prepared as described above and resuspending to the original volume with ampule $D_2O$. No pH correction was applied to counter isotope effects.

*NMR Spectroscopy.* All NMR experiments were conducted at 25 °C primarily on Varian Unity 500 or Unity Inova 600 spectrometers with triple resonance pulsed-field gradient probes, except where otherwise noted. In all cases, data were acquired in the States−TPPI mode (Marion *et al.*, 1989a; States *et al.*, 1982) for quadrature detection. Proton chemical shifts are referenced to an external 1 mM DSS standard (in $D_2O$ at 25 °C) at 0 ppm. Carbon chemical shifts are referenced to an external methyl iodide standard at −20.7 ppm. Carrier frequencies were as follows, except where noted otherwise: $^1$H, 4.75 ppm; $^{15}$N, 118−125 ppm; aliphatic $^{13}$C, 41.1−43.3 ppm; $^{13}C_\alpha$, 51.6−56.7 ppm; and $^{13}C'$, 172.2−179.8 ppm. Unless otherwise noted, parameters for these experiments are listed in Table 1.

*(A) 2D Homonuclear and Heteronuclear Experiments.* 2D homonuclear TOCSY (Braunschweiler & Ernst, 1983) and NOESY (Kumar *et al.*, 1980) spectra were acquired with presaturation during the recycle period to reduce the water signal, followed by SCUBA recovery (Brown *et al.*, 1988). In the TOCSY, a DIPSI-2 (Rucker & Shaka, 1989) mixing sequence of 30 ms duration was applied to transfer magnetization. The 75 ms mixing period in the 2D NOESY included an additional SCUBA recovery sequence.

2D $^1$H, $^{15}$N, and $^1$H, $^{13}$C heteronuclear single-quantum coherence (HSQC; Bodenhausen & Ruben, 1980) spectra were acquired on Cdc42Hs samples which were homogeneously labeled with $^{15}$N, as well as protein samples with specific [$^{15}$N]amino acids incorporated and samples partially labeled (8−10%) with $^{13}$C. Deuterium-exchange HSQCs were collected with 192−512 complex points in $t_1$ and 2048 in $t_2$ and with spectral widths of 16 ppm in $\omega_2$ and 59.2 ppm in $\omega_1$ at intervals of 0.5, 1.5, 2.5, 4, 8, 12, 24, 272, and 864 h after resuspension in $D_2O$.

*(B) 3D Experiments for Sequential Backbone Assignments.* HNCO experiments (SE HNCO) were collected essentially as in Kay *et al.* (1994), except that the selective water pulse was excluded, and SEDUCE decoupling of $C_\alpha$s during

nitrogen evolution was replaced by a 180° refocusing pulse. An HNCA (SE HNCA) experiment collected on the GDP-bound form of Cdc42Hs used the same pulse sequence except for the positioning of the carbon carrier and the direction of shifted pulses. An HN(CO)CA experiment [SE HN(CO)-CA] also collected on Cdc42Hs·GDP was acquired with a pulse sequence similar to that of the SE HNCO but modified as suggested in Grzesiek and Bax (1992). On the GMPPCP-bound form of Cdc42Hs, an HNCA was collected (HNCA with $C_\beta$ decoupling) as in Yamazaki *et al.* (1994b) except for the removal of sensitivity enhancement pulses. Also, during $t_1$, WALTZ-16 decoupling was applied to protons, allowing the removal of $T_c$ ($C_\alpha$ constant time period), and WURST decoupling (Kupce & Freeman, 1996; Kupce & Wagner, 1995) is applied to both the carbonyls and $C_\beta$s (Matsuo *et al.*, 1996). An HN(CO)CA also collected on Cdc42Hs·GMPPCP used a nonsensitivity-enhanced version of the pulse sequence in Yamazaki *et al.* (1994a), except for the addition of SEDUCE decoupling of $C_\alpha$s during $\tau_c$ (in which coupling develops between nitrogen and the carbonyl) and WURST decoupling of $C_\beta$s in $t_1$, as well as WALTZ-16 decoupling of protons in $t_1$, allowing the removal of $T_c$ ($C_\alpha$ constant time period). An HCACO collected on Cdc42Hs·GDP with a Bruker AMX-600 spectrometer used the pulse sequence in Dötsch *et al.* (1995).

*(C) 3D Experiments for Side Chain Assignments.* $^{15}$N, $^1$H TOCSY HSQC experiments (Norwood *et al.*, 1990) were recorded with various mixing times on Cdc42Hs with both GDP and GMPPCP bound. The 50 ms mixing time experiment on Cdc42Hs·GDP was collected on a Varian Unity Plus 400 MHz spectrometer. HCCH TOCSY spectra were collected on Cdc42Hs·GDP and Cdc42Hs·GMPPCP with a pulse sequence exactly as described in Kay *et al.* (1993). The proton carrier in the GMPPCP HCCH TOCSY was shifted upfield of the water resonance (2.37 ppm). A CBCA(CO)NH spectrum was acquired on Cdc42Hs·GDP exactly as described (Muhandiram & Kay, 1994).

*(D) 3D and 4D NOESY Experiments.* Several multidimensional NOESY experiments were collected with mixing periods of 100 ms. A $^1$H, $^{15}$N NOESY HSQC (Fesik & Zuiderweg, 1988; Marion *et al.*, 1989b) was acquired on Cdc42Hs·GDP using a pulse sequence which features water suppression by a selective pulse on the water resonance followed by dephasing with a gradient during the mixing period, as well as by a gradient applied while the magnetization of interest is spin-ordered during the final INEPT transfer step. Similar data were acquired on Cdc42Hs·GMPPCP but with the pulse sequence of Talluri and Wagner (1996). Both spectra were acquired on a Varian Unity Plus 750 MHz spectrometer. A $^1$H, $^{13}$C NOESY HSQC spectrum was recorded on Cdc42Hs·GDP in 100% $D_2O$. The pulse sequence was as described above for the GDP-bound $^1$H, $^{15}$N NOESY HSQC, except that water suppression was accomplished by presaturation during the recycle period rather than by a selective proton pulse, and nitrogen decoupling was added during the carbon chemical shift evolution period. An aromatic carbon-edited $^1$H, $^{13}$C NOESY HSQC was recorded with a slight modification of the same pulse sequence: SEDUCE decoupling of both aliphatic and aromatic carbons occurred during proton evolution and the carbon carrier was set on the aromatics (125.7 ppm) for all other carbon pulses.

A 4D $^{13}$C/$^{13}$C-separated HMQC NOESY HMQC spectrum was collected on Cdc42Hs·GDP in 100% $D_2O$ as in Vuister *et al.* (1993), except that a 180° decoupling pulse on nitrogen was added in the middle of both carbon evolution periods and presaturation of water was used in the recycle period. The parameters listed under $\omega_3$ in Table 1 include both $\omega_3$ and $\omega_4$ dimensions, respectively. In this experiment the carbon carrier was set in the methyl region (19.1 ppm) and the proton carrier at 3.2 ppm.

*Processing of NMR Experiments.* Data were processed with versions 2.1 and 2.3 of Felix software (Molecular Simulations, Inc.) on Sun UltraSparc 140 and Sparc 5 workstations. Sensitivity-enhanced experiments were initially processed as in Kay *et al.* (1992). Data corrected in this way were then processed like nonenhanced data. When necessary, reduction of the water peak in the directly detected dimension was accomplished by convolution of the FID with a sine-bell function of empirically determined width to identify the lowest frequency component, followed by subtraction of this component from the FID. In the typical case, data were zero-filled to double the original number of data points and apodized by convolution with a squared sine-bell window function shifted by 60° to 90°, depending on the relative amount of signal in the data. For experiments with a limited number of sampling points in $\omega_1$, linear prediction was sometimes applied during the transformation of this final dimension to increase resolution. Data visualization and spectral assignment utilized the XEASY (Bartels *et al.*, 1995) program. Peaks were integrated with the Peakint program (N. Schäfer, diploma thesis, ETH, Zürich, Switzerland).

*Structure Calculation.* Distance constraints for structure calculation were derived from both $^1$H, $^{15}$N and $^1$H, $^{13}$C NOESY HSQC experiments. In some cases ambiguities in the $^{13}$C-dispersed spectrum were resolved by consulting a 4D $^1$H, $^{13}$C HMQC NOESY HMQC spectrum, which was of fairly low sensitivity. The relationship between peak volume and the distance between the two atoms involved was derived in different ways for each of the NOESY spectra used. In the $^1$H, $^{15}$N correlation spectrum, the peak volumes of 13 well-dispersed sequential $d_{\alpha N}$ NOE peaks and 6 well-dispersed $d_{\alpha N(i,j)}$ NOE peaks (resulting from correlations within and between parallel $\beta$-strands, respectively) were compared to the distances expected in ideal $\beta$-sheets (2.2 and 3.0 Å, respectively). In the $^1$H, $^{13}$C NOESY HSQC, the same type of relationship was derived by comparison of the peak volumes of 3 well-dispersed leucine and valine intraresidue methyl-to-methyl NOEs to ideal distance values within leucine or valine side chains. In this manner, it was possible to correlate quantitatively peak volumes in each of these spectra with distances between pairs of protons. Approximately 1850 constraints from NOESY spectra were then classified into five categories depending on the relative strength of the NOESY cross peak ($<$2.4, $<$2.9, $<$3.4, $<$4.0, and $<$5.5 Å). In addition to these distance constraints, 50 hydrogen bonds were included (as indicated by the secondary structures presented below, based on deuterium exchange rates, chemical shift indices, and characteristic NOE patterns), each as two separate restraints. For every hydrogen bond, the amide proton to carbonyl oxygen distance was constrained between 1.8 and 2.3 Å and the amide nitrogen to carbonyl oxygen between 2.5 and 3.3 Å. Since experiments designed to generate dihedral angle restraints were unsuc-

cessful, the backbone dihedral angles of those residues involved in α-helices and β-strands were loosely constrained to "favorable" regions of $\phi$, $\psi$ space in the following way: α-helix, $\phi$ −80° ± 50°, $\psi$ −20° ± 50°; β-strand, $\phi$ −105° ± 65°, $\psi$ 145° ± 45° (Laskowski *et al.*, 1993).

This list of roughly 2100 distance and dihedral angle constraints was used to calculate several preliminary structures using X-PLOR 3.843 (Brunger, 1996). Structure calculation involved four main steps. First, 50 substructures containing only N, HN, $C_\alpha$, $H_\alpha$, C′, $C_\beta$, and $C_\gamma$ atoms were calculated using metric matrix distance geometry. Next, missing atoms were added, and the resultant structures were refined by simulated annealing and energy minimization, using center averaging (the pseudoatom approach) to represent constraints to nonstereospecifically assigned atoms and a square well potential to hold constrained atoms together. A second phase of simulated annealing and energy minimization was then used which also included "Ramachandran refinement" (this constrains dihedral angles to those that are known to be physically realizable; Kuszewski *et al.*, 1996). Twenty "best" structures were selected on the basis of low overall energy values and low numbers of constraint violations. These 20 structures were then refined once again by energy minimization but with the additional incorporation of $H_\alpha$, $C_\alpha$, and $C_\beta$ chemical shifts as restraints (Kuszewski *et al.*, 1995a,b).

Those constraints which were violated consistently and significantly within the set of calculated structures (*i.e.*, those for which the distance between the two protons in the refined structures exceeded the input constraint value) were reexamined to see if other possibilities existed for that particular NOESY peak assignment. The violation of a constraint could arise from one of three possibilities, which were addressed in the following order: (1) If the violation corresponded to a misassignment, it was reassigned. (2) If the violation did not correspond to a misassignment but could have arisen as a result of the misassignment of another peak, it was left in the constraint set. (3) In several cases, some ambiguity was present due to the inability to assign all of the side chains. In these few cases, the constraint was removed. In addition, the calculated structures were used to help resolve previously ambiguous NOESY assignments. An amended constraints list was then used to calculate a new set of structures, and this entire procedure was iterated several times.

Due to a paucity of intermolecular restraints between the GDP and the protein, it was not possible to constrain the GDP satisfactorily within the calculation protocol used to determine the Cdc42Hs structures. Instead, the position of the GDP was modeled in conjunction with two intermolecular NOEs: Phe28 HZ···GDP H1 and Ile117 HB···GDP (H21 and H22). The starting position of the GDP was determined to a first approximation by superposition of the H-ras·GDP crystal structure onto the "core" $C_\alpha$ atoms of each of the 20 structures [determined by NMRCORE (Kelley, 1997) to consist of residues 3−9, 16−28, 41−47, 49−56, 77−85, 88−102, 109−114, 117, 118, and 141−176]. The GDP was then copied directly into the Cdc42Hs structure from its position in the H-ras·GDP structure. The Cdc42Hs·GDP structures were subsequently energy minimized using X-PLOR to remove steric conflicts in the complex, initially keeping the atoms in the protein fixed, and then, once the position of the GDP was more favorable, minimizing the whole complex (this resulted in slight rearrangement of some of the protein side chains lining the GDP-binding site). Structures were visualized and analyzed using Insight II (Molecular Simulations, Inc.), MOLMOL (Koradi *et al.*, 1996), PROCHECK-NMR (the NMR version of the PROCHECK program; Laskowski *et al.*, 1993), and NMRCLUST (Kelley *et al.*, 1996).

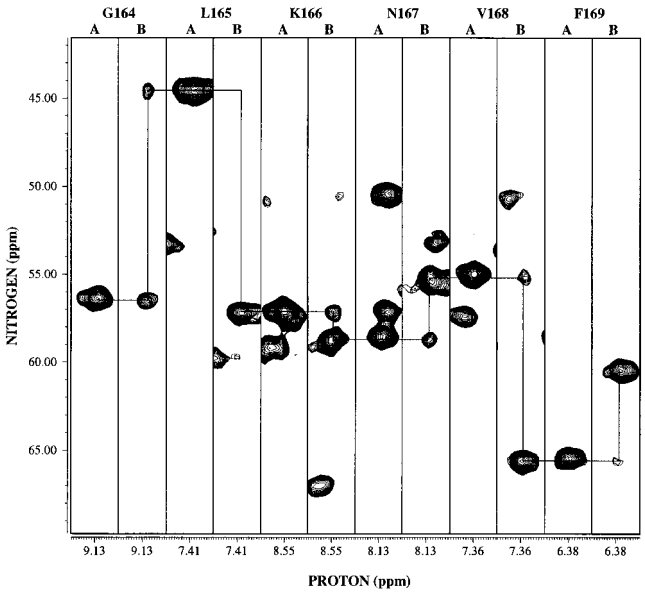


FIGURE 1: $F_2$,$F_3$ slices extracted from HN(CO)CA (A) and HNCA (B) spectra collected on Cdc42Hs·GDP, illustrating the sequential backbone assignment of residues Gly164 to Phe169.

## RESULTS AND DISCUSSION

*Backbone Assignments for Cdc42Hs·GDP.* Several types of heteronuclear experiments were used to assign main chain nuclei sequentially in Cdc42Hs·GDP. The most useful of these experiments was an HNCA and HN(CO)CA pair, which provides correlations between a given amide proton and nitrogen and both the intraresidue and the (*i - 1*) $C_\alpha$s (Figure 1). In some cases, where the $C_\alpha$ chemical shifts were well dispersed, the correlations in these two spectra allowed the unambiguous identification of sequential residues and sometimes even the exact identification as specific amino acids in the protein sequence due to characteristic $C_\alpha$ chemical shifts. However, in most cases, unambiguous identification of sequential residues using the HNCA and HN(CO)CA spectra alone was complicated by overlap in the $C_\alpha$ chemical shifts. In these cases additional information from selective amino acid labeling and 3D TOCSY and NOESY spectra was used to aid in the selection of the correct neighbor. For instance, assuming that neighboring residues in the protein sequence will encounter a similar complement of protons and will therefore have in common a large percentage of NOESY cross peaks, comparison of the $^1$H, $^{15}$N NOESY HSQC slices for potentially neighboring spin systems [identified in the HNCA/HN(CO)CA spectra] often narrowed the choices considerably. Several different methods were used to assign residues which were correlated in this way to specific amino acids in the protein sequence. First, inspection of the corresponding slices from $^1$H, $^{15}$N TOCSY HSQC spectra sometimes allowed the identification of a particular spin system to amino acid type. Second, six protein samples were made in which selective [$^{15}$N]amino acids were incorporated (Lys, Val, Leu, Tyr, Phe, Ala). $^1$H,

$^{15}$N HSQC spectra collected on these samples illustrated which amide peaks belong to each amino acid type. Finally, some additional assignments to amino acid type were accomplished by a suppressed-labeling protocol (Shortle, 1994), in which proteins are expressed in $^{15}$N-containing minimal media supplemented with unlabeled amino acids to block specifically the incorporation of $^{15}$N into those residues. Many threonines, arginines, and glycines were uniquely identified using this method.

Since proline residues lack an amide proton, they are absent from all spectra mentioned above. Backbone assignment of most proline residues employed a set of three experiments: HNCO, HN(CO)CA, and HCACO. The first two of these provide correlations from the amide of a given residue to the (*i - 1*) C′ and $C_\alpha$ chemical shifts, respectively. The HCACO then provides an intraresidue correlation from these $C_\alpha$ and C′ chemical shifts to the $H_\alpha$ chemical shift, which completes the backbone assignment for proline residues.

*Side Chain Assignments for Cdc42Hs·GDP.* Assignment of side chain nuclei relied on several 3D TOCSY experiments as well as a CBCA(CO)NH. Most $H_\alpha$ and $H_\beta$ nuclei, as well as some $H_\gamma$, were assigned in $^1$H, $^{15}$N TOCSY HSQC spectra collected with a variety of mixing times (30−70 ms). Extension of assignments further along the side chain required the additional use of HCCH TOCSY experiments with 7, 14, and 21 ms mixing periods. The CBCA(CO)NH aided in the interpretation of these spectra by providing $C_\beta$ chemical shifts correlated through the (*i* + *1*) amide resonance.

*(A) Stereospecific Assignment of γ- and δ-Methyls.* The methyls of 29 out of 35 leucine and valine residues were stereospecifically assigned using the 10% $^{13}$C-labeling method described in Szyperski *et al.* (1992). Due to the biosynthetic pathways for valine and leucine, a $^1$H, $^{13}$C HSQC (without carbon decoupling) collected on this 10% $^{13}$C-labeled sample exhibits singlets for $\gamma_2$- and $\delta_2$-methyls, whereas the $\gamma_1$- and $\delta_1$-methyls are split in the carbon dimension.

*(B) Assignment of Asn, Gln, Phe, Tyr, Met, His, and Trp Side Chains.* Some residues are impossible to assign completely with spectra using TOCSY transfers only since portions of the side chains are isolated by the presence of a carbon atom without a bound proton. Asn and Gln side chain amides were easily correlated to the rest of the residue because the amido moiety resembles the arrangement of nuclei in the peptide backbone. As a result these side chain amides are present in HN(CO)CA and CBCA(CO)NH spectra, which provide intraresidue correlations to the Asn $C_\beta$, $C_\alpha$ and Gln $C_\gamma$, $C_\beta$ chemical shifts.

Phenylalanine and tyrosine $\delta$, $\epsilon$, and $\zeta$ protons were assigned within a particular side chain using a 2D homonuclear TOCSY taken in $D_2O$, as well as by comparison of cross peak patterns in slices from an aromatic carbon-edited $^1$H, $^{13}$C NOESY HSQC. Aromatic side chain protons were then matched with the backbone resonances of a particular phenylalanine or tyrosine by comparison of the $^1$H, $^{15}$N NOESY HSQC slice for the backbone amide with the side chain aromatic slices from the $^1$H, $^{13}$C NOESY HSQC.

The two methionine $\epsilon$-methyls were located in a region of the $^1$H, $^{13}$C HSQC unique to this type of methyl group and then matched with the Met1 and Met45 residues by comparison of the cross peak pattern in the methyl $^1$H, $^{13}$C NOESY HSQC slices to those of the backbone amides in the $^1$H, $^{15}$N NOESY HSQC. The $\delta$- and $\epsilon$-protons of the two histidine side chains (103 and 104) were not assigned. Finally, the tryptophan $\epsilon$-amide was easily located in $^1$H, $^{15}$N HSQC spectra due to the large downfield shift of the $\epsilon$-proton (11.46 ppm), and the neighboring $\delta$-proton was correlated with it using a $^1$H, $^{15}$N TOCSY HSQC experiment.

*Summary of Cdc42Hs·GDP Assignments.* These assignments are summarized in an HSQC spectrum collected on $^{15}$N-Cdc42Hs·GDP (Figure 2A). It was not possible to fully assign Cdc42Hs. Three residues that remain completely unassigned are individual prolines (P34, P87, and P99), which are generally difficult to characterize because they lack an amide proton. In addition, the four C-terminal residues of Cdc42Hs are missing from all spectra and have been shown by mass spectrometry to have been removed by proteolysis. However, an additional ten unassigned residues fall into three contiguous stretches of the protein sequence (F37−N39, T58−G60, and R66−P69), which correspond to residues in the switch I and II regions of H-ras. On the basis of the fact that the resonances of residues flanking the unassigned regions are of substantially decreased intensity (Figure 2B), we propose that conformational dynamics on an intermediate time scale in these regions result in line broadening to such a degree as to make the peaks disappear into background noise. Preliminary measurements of $^{15}$N relaxation times (data not shown) support this proposition, indicating motion in the millisecond to microsecond time scale for residues flanking the unassigned regions. Both switch I and switch II of H-ras have been shown to undergo slow conformational exchange in solution (Kraulis *et al.*, 1994).

*Secondary Structure.* The location and extent of secondary structures were determined on the basis of amide proton exchange rates, ensemble Ramachandran plots, the chemical shift index (Wishart & Sykes, 1994; Wishart *et al.*, 1992), and patterns of NOE connectivities. This information is summarized in Figure 3.

The secondary structures indicated in Figure 3 are for the most part similar to those of H-ras and other low molecular weight GTP-binding proteins. One notable exception is that no evidence is found either from experimental data (the chemical shift indices, the amide exchange rates, and the NOE connectivities; Figure 3) or from the calculated structures for the existence of the α2 helix, which spans residues 69−75 of switch II in the H-ras·GDP crystal structure (Tong *et al.*, 1991) and residues 66−75 in the solution structure of that protein (Kraulis *et al.*, 1994). This short helix is unlikely to occur in a similar location in Cdc42Hs due to the presence of prolines at positions 69 and 73. However, a recent crystal structure of a mutant Rac1 protein bound to GMPPNP (Hirshberg *et al.*, 1997) contains two short $3_{10}$-helices spanning residues 62−64 and 68−71. Considering that the sequences of Cdc42Hs and Rac1 are identical between residues 57 and 79, it is surprising that their structures would differ in this region. It might be supposed, since the mutant Rac1 structure represents the active conformation, that these short helices form upon binding of GTP (or a GTP analog). However, in the present study chemical shifts of Cdc42Hs bound to GMPPCP have been assigned using a strategy similar to that used for the GDP-bound form, and the chemical shift indices of Cdc42Hs bound to GMPPCP (Figure 3) also do not support the

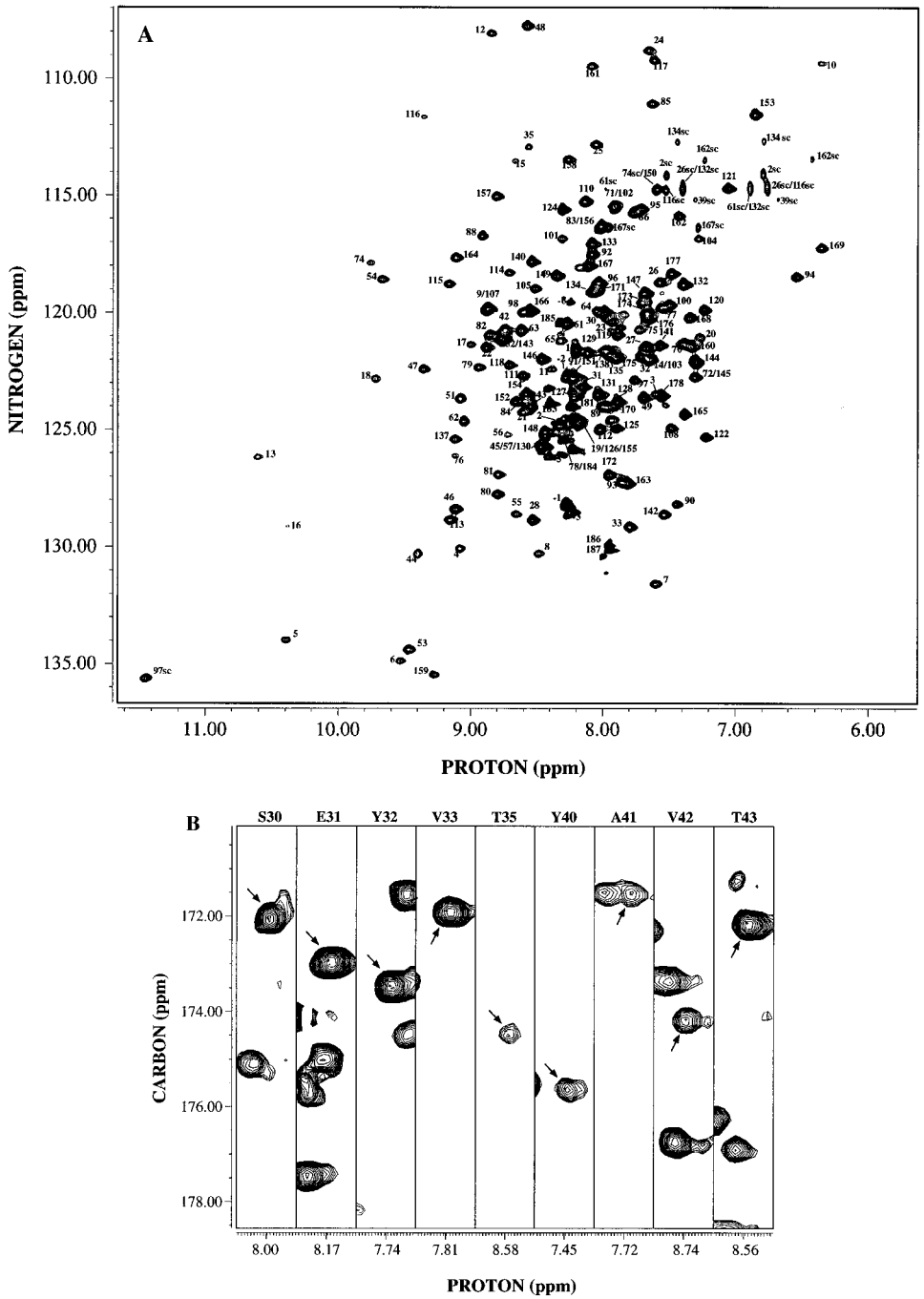


FIGURE 2: (A) $^1$H, $^{15}$N HSQC spectrum collected on Cdc42Hs·GDP with peaks labeled with assignments. (B) $F_2$,$F_3$ slices extracted from an HNCO spectrum collected on Cdc42Hs·GDP representing residues Ser30 to Thr43. In each strip, the peak for that residue is indicated by an arrow. No strips are shown for Pro34, which lacks an amide proton and therefore does not appear in the HNCO, or for Val36, for which no peak is found in this spectrum.

presence of any helix in this region. In addition, similar to the behavior observed in the GDP-bound protein, peak intensities are reduced for residues corresponding to the switch I and switch II domains of Cdc42Hs·GMPPCP, suggesting that these resonances may be broadened by conformational exchange even in the active conformation. Although the reason for the presence of these $3_{10}$-helices in Rac1 *vs* their absence in Cdc42Hs is unclear, one possibility is that it reflects a difference in conformation in this region between the crystal and solution states.

Another difference between Cdc42Hs and other low molecular weight GTP-binding proteins concerns the insert region which is only found in members of the rho subfamily. In the Rac1·GMPPNP structure, two helices are observed in this region. The first, which extends from residue 117 to residue 120, is not present in Cdc42Hs·GDP despite a nearly identical amino acid sequence, nor does it form upon activation of Cdc42Hs by GMPPCP binding, since only a single backbone chemical shift change occurs within these residues upon protein activation (D118 $H_\alpha$ shifted by 0.05 ppm). On the other hand, residues 124−128 within the insert region of Cdc42Hs may adopt a helical structure (αI) as indicated by short-range NOEs, chemical shift indices, and the $\phi$, $\psi$ dihedral angles of these residues in the final calculated structures (see below). This helix corresponds fairly closely to helix H3b in the Rac1 structure (residues 123−130); however, in Cdc42Hs the stability of this helix appears to be fairly low, since no amide protons are

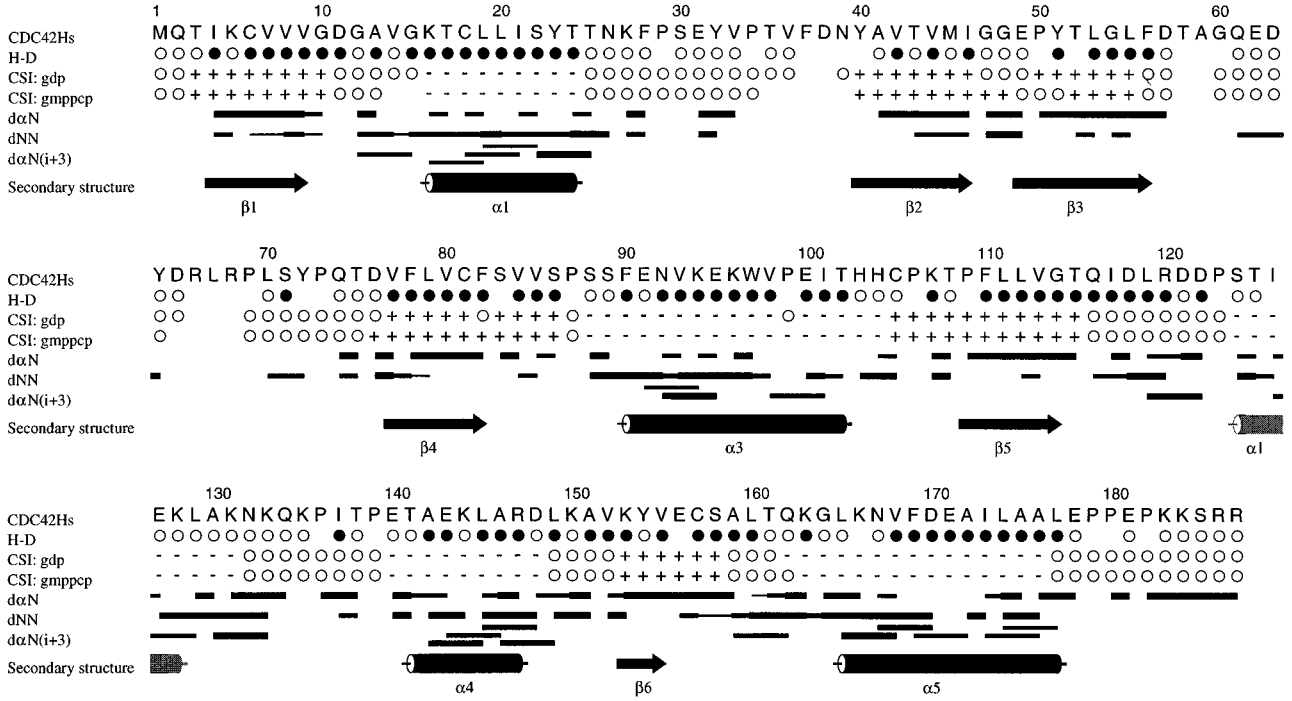


FIGURE 3: Summary of the assignments and the secondary structure of Cdc42Hs. Amide protons in Cdc42Hs·GDP exhibiting deuterium exchange (H−D) with half-lives of less than 1 h in $D_2O$ are shown using open circles; closed circles indicate exchange half-lives of greater than 1 h. The second and third rows show the composite chemical shift indices of the $H_\alpha$, $C_\alpha$, $C_\beta$, and C′ protons (CSI; Wishart & Sykes, 1994; Wishart *et al.*, 1992) in the GDP-bound and GMPPCP-bound forms, respectively, where (+) is correlated with $\beta$-strand, (−) with helix, and (○) with coil. NOE intensities are represented in the rows labeled $d_{\alpha N}$, $d_{NN}$, and $d_{\alpha N(i+3)}$, with thick bars indicating strong NOEs (<2.9 Å), medium bars indicating intermediate NOEs (<4 Å), and thin bars indicating weak NOEs (<5.5 Å). Secondary structural elements are indicated in the bottom row.

measurably protected from deuterium exchange. The low stability of this helix could perhaps explain why, in the crystal structure of mutant Rac1 (Hirshberg *et al.*, 1997), the electron density in this region required enhancement using a "growing density" method.

A number of other more subtle differences also exist between Cdc42Hs and both the H-ras and the Rac1 structures. For instance, the α3 helix in Cdc42Hs is truncated by three residues N-terminally and two residues C-terminally relative to the corresponding helices in H-ras and Rac1. Once again, the sequences of Cdc42Hs and Rac1 are highly conserved in this region, and differences between the Rac1 and Cdc42Hs structures cannot be attributed to nucleotide-induced conformational changes since the chemical shift indices of Cdc42Hs bound to GMPPCP are identical to those in the GDP-bound form. Both Cdc42Hs and Rac1 structures are in agreement, however, that a proline residue at position 99 does not break the helix but only causes a small degree of distortion. The final α5 helix is truncated in Cdc42Hs by two turns at the C-terminus compared to H-ras, presumably due to a cluster of three prolines between residues 179 and 182 in Cdc42Hs which are absent from H-ras. As a result, residues in the C-terminal tail of Cdc42Hs exhibit extremely narrow line widths characteristic of fast exchange and presumably form a flexible tether to the membrane *via* posttranslational prenylation as in H-ras. Using medium-range NOE cross peaks, the packing order of the six-stranded $\beta$-sheet was established as $\beta2-\beta3-\beta1-\beta4-\beta5-\beta6$, with all strands parallel except $\beta2$ (Figure 4).

*Description of the Final Structures.* From the 20 aligned structures in Figure 5, it can be seen that the "core" of the protein is well-defined, with an rms deviation of 0.57 ± 0.08 Å for all non-hydrogen atoms. Most of the structural disorder is confined to four regions of the protein sequence. First, seven non-native amino-terminal residues, which remain after cleavage of the fusion protein, are completely flexible in solution, as are all C-terminal residues after E178. In addition, regions corresponding to switch I and switch II of H-ras also show a high level of disorder (residues 31−40 and 57−74, respectively). While the conformational heterogeneity in these two regions is primarily caused by a lack of constraints since they contain some residues which could not be assigned, $^{15}N$ relaxation data as well as peak intensity differences suggest that these regions are indeed more flexible than the remainder of the protein. Corresponding regions in the H-ras·GDP solution structure (residues 30−40 and 58−66) also show a higher level of disorder than the rest of the protein, which suggests that flexibility in these domains may be characteristic of ras-like proteins in solution. When considering the entire structure, including all disordered regions, an rms deviation of 3.8 ± 0.5 Å is calculated for all non-hydrogen atoms. Statistics on the stereochemical quality, restraint violations, and the energies of the different components in the potential function across the ensemble of 20 structures are presented in Table 2.

In summary, Cdc42Hs consists of a central six-stranded $\beta$-sheet which is curled about the axis of helix α5. Helix α1 lies perpendicular to the $\beta$-strands, also on the concave side of the sheet. Flanking the convex surface are helices α3 and α4, which lie parallel to the $\beta$-strands. As discussed above, no evidence is found for the α2 helix in Cdc42Hs. The short αI helix of the insert region is apparently less stable than other helices and lies adjacent to the loop between $\beta4$ and α3, perpendicular to the $\beta4$ and α3 axes. The insert region itself forms a relatively compact loop structure which does not occlude the nucleotide binding site as predicted

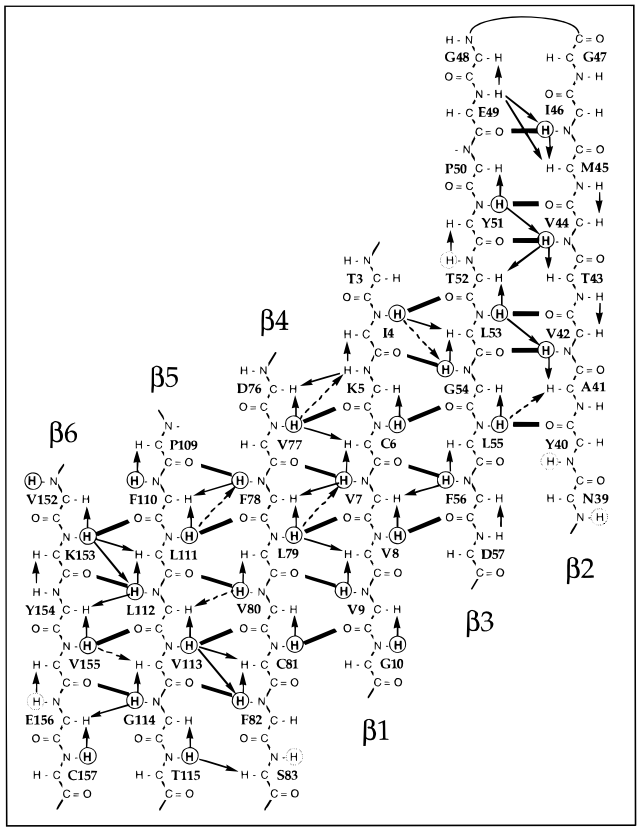


FIGURE 4: Illustration of the alignment and packing of the $\beta$-sheet. Amide protons exhibiting deuterium exchange (H−D) with half-lives of greater than 1 h in $D_2O$ are indicated by solid circles, whereas dotted circles indicate those amide protons for which it was not possible to measure the exchange rate due to peak overlap. Hydrogen bonds are indicated by heavy bars, and strong NOEs (<2.9 Å) and weak NOEs (<5.5 Å) are indicated by solid and dashed arrows, respectively.

(Sutcliffe *et al.*, 1994) but appears to flank portions of Cdc42Hs which do bind the nucleotide. Close contacts are made between residues 120−129 in the insert region and residues 84−89 which lie in the loop between $\beta$4 and α3. While a fair amount of disorder is seen in the insert region of the 20 fully aligned structures, alignment of the insert region alone shows it to be well-defined structurally (Figure 5). The disorder seen with respect to the remainder of the protein is due either to molecular motion or to a lack of NOE constraints.

*Definition of a Switch Surface.* Assignment of an active Cdc42Hs protein bound to the nonhydrolyzable GTP analog, GMPPCP, illustrated which regions of the protein undergo conformational changes upon activation, as evidenced by chemical shift changes between the active and inactive states. As expected, the P-loop between $\beta$1 and α1, as well as those portions of switch I and switch II which were assignable, is dramatically affected by activation of the protein. However, the α3−$\beta$4 loop and helix α3 also experience significant chemical shift changes when GMPPCP binds (labeled switch III in Figure 6), forming a large "switch surface" (Figure 6) in Cdc42Hs. The unspecified conformational change in helix α3 is the likely cause of changes in the quantum yield of intrinsic tryptophan fluorescence (at position 97) and extrinsic sNBD fluorescence (covalently attached at K150 in the neighboring α4 helix) upon GTP binding and hydrolysis (Leonard *et al.*, 1994; Nomanbhoy *et al.*, 1996). Interestingly, the chemical shifts of the nuclei within the insert region

Table 2: Structural Statistics

(a) Stereochemical Quality[a]

| | av value | std dev |
|---|---|---|
| Ramachandran plot | | |
| residues in most favored regions (%) | 63 | 3 |
| residues in additional allowed regions (%) | 28 | 3 |
| residues in generously allowed regions (%) | 6 | 2 |
| residues in disallowed regions (%) | 3 | 1 |
| nonbonded contacts (no. within 2.1 Å) | 0 | 0 |
| overall quality (*G*-factor) | −0.59 | 0.02 |

(b) NOE Constraint Violations

| | no. of violations | |
|---|---|---|
| | av value | std dev |
| violations | | |
| >0.3 Å | 58 | 5 |
| >0.4 Å | 22 | 3 |
| >0.5 Å | 4 | 2 |
| >0.6 Å | 1.0 | 1.1 |
| >0.7 Å | 0.3 | 0.6 |
| >0.8 Å | 0 | 0 |

(c) Energy Values in Potential Function

| | av value | std dev |
|---|---|---|
| total energy (kcal/mol) | 4360 | 120 |
| bonded energy terms | | |
| bond length (kcal/mol) | 153 | 7 |
| bond angle (kcal/mol) | 720 | 40 |
| improper (kcal/mol) | 130 | 10 |
| nonbonded energy terms | | |
| simplified van der Waals repulsive (kcal/mol) | 790 | 50 |
| NOE (kcal/mol) | 840 | 40 |
| carbon chemical shift (kcal/mol) | 630 | 40 |
| proton chemical shift (kcal/mol) | 380 | 40 |
| Ramachandran dihedral restraints (kcal/mol) | 700 | 20 |

[a] This analysis was performed using Procheck-NMR (Laskowski *et al.*, 1993).

are almost entirely invariant, suggesting that no conformational changes are induced within the insert region upon activation of the protein by GTP binding. Nevertheless, the location of the insert region relative to the Cdc42Hs switch surface helps to explain its role in the mediating the binding of target and regulatory proteins.

The structure of the insert region is of critical importance for understanding the function of rho subfamily members. Studies in another rho subfamily protein, Rac, have shown that deletion of, or point mutations within, the insert region disrupts normal binding to a primary target, the NADPH oxidase complex (Freeman *et al.*, 1996). Interactions of activated Rac with the PAK serine/threonine kinase, on the other hand, are not affected by changes in this region, whereas both PAK binding and stimulation of NADPH oxidase activity are disrupted by mutations within the "effector" domain residues 32−38. While our studies on Cdc42Hs·GMPPCP rule out the possibility that the insert region functions as an additional switch domain that determines selectivity of target proteins for the active conformation, it is clear that both the insert region and the effector domain are crucially involved in the binding of the NADPH oxidase effector complex. If both the insert and the effector domain contribute to the binding of the NADPH oxidase effector complex, then the most likely orientation for the (bound) effector complex would cover the nucleotide binding site. This suggests that the insert may be viewed as a region that is either important as a secondary binding site for the effector protein or as a transduction element which mediates the interaction between Rac and the NADPH oxidase system.

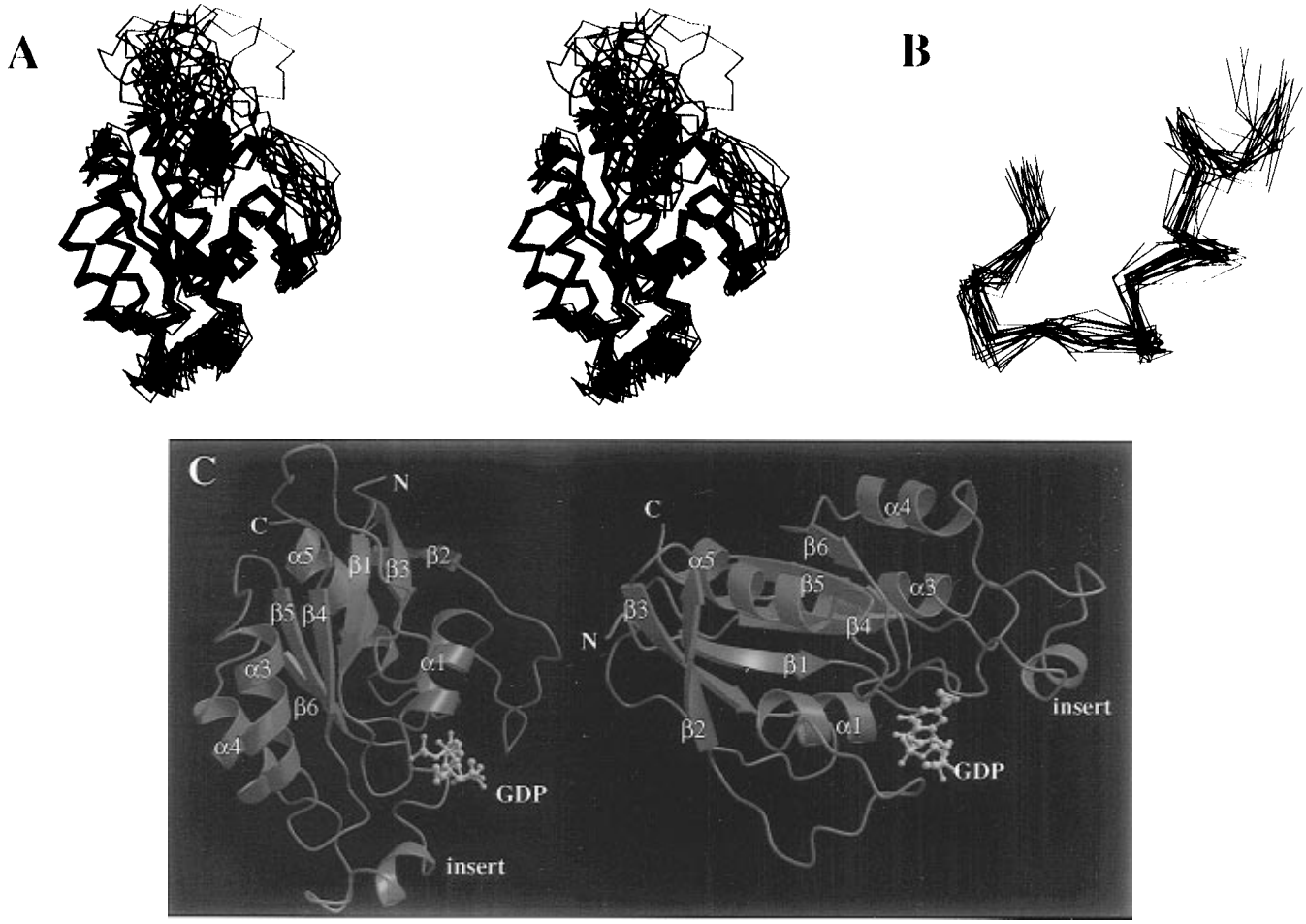


FIGURE 5: (A) Stereoview of the $C_\alpha$ atoms of 20 aligned Cdc42Hs·GDP structures. (B) Insert region residues Asp118 to Lys135 from the 20 final structures are aligned against each other and displayed as a $C_\alpha$ plot. (C) The most representative structure, determined by NMRCLUST (Kelley *et al.*, 1996), shown in two orientations. This figure was produced with MOLSCRIPT and Raster3D (Kraulis, 1991; Merritt & Murphy, 1994).

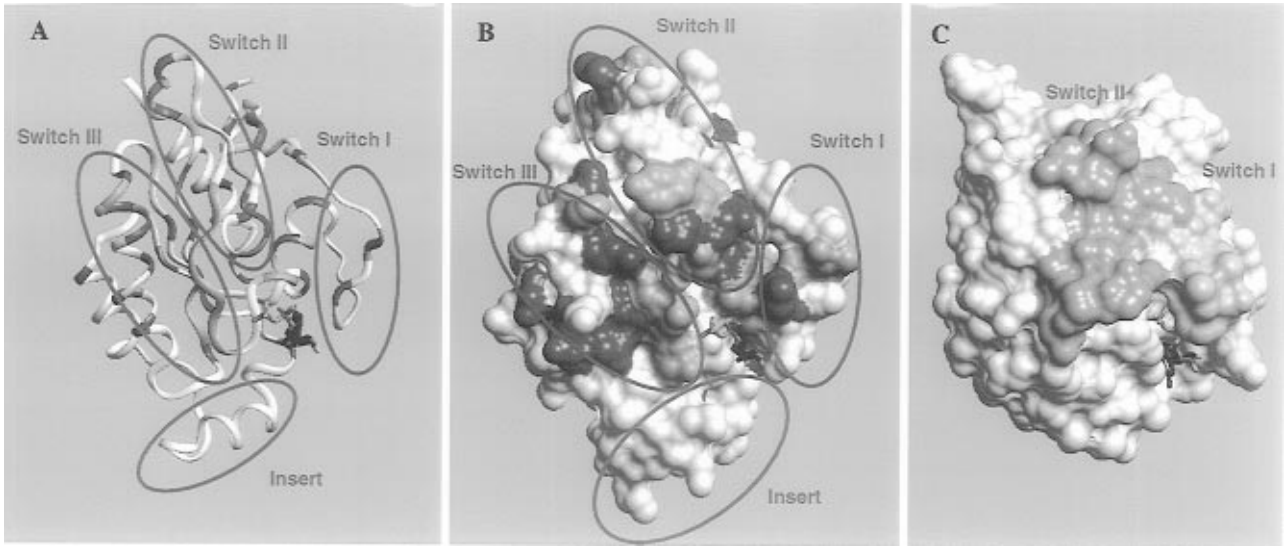


FIGURE 6: Switch surface of Cdc42Hs shown as (A) a schematic representation of the backbone and (B) a rendering using Insight II (MSI). In both cases GDP is shown in the "stick" representation. The colors indicate the degree to which chemical shift changes are observed between the GDP-bound state and the GMPPCP-bound state. Chemical shift changes were placed in three groups: large (red), medium (magenta), and small (blue). The categories were based on chemical shift changes in backbone resonances only. (C) Structure of GDP-bound H-ras (Tong *et al.*, 1991) rendered using Insight II showing the positions of switch I and switch II.

Similarly, the insert is involved in the binding of a primary target protein, IQGAP, to Cdc42Hs (McCallum *et al.*, 1996). Replacement of the insert region of Cdc42Hs with the smaller H-ras loop 8 leads to decreased affinity of this chimera for IQGAP. It is expected that IQGAP binds primarily to those regions of Cdc42Hs which change conformation upon GTP binding. Indeed, the fluorescence of an extrinsic sNBD reporter group covalently attached at K150, which monitors conformational changes within the Switch surface, is also sensitive to IQGAP binding. Whereas in H-ras most of the

changes which occur upon activation are confined to switch I and switch II (Wittinghofer & Pai, 1991), in Cdc42Hs regions of α3 and the adjacent $\beta$4 strand also exhibit significant chemical shift changes between the active and inactive forms of the protein. The insert region may play a role in the binding interaction with IQGAP, precisely due to its orientation relative to both the loop between $\beta$4 and α3 and the P-loop between $\beta$1 and α1.

The insert region is also intimately involved in the mechanism of action of the RhoGDI. While the RhoGDI protein is capable of binding to the chimeric Cdc42Hs protein in which H-ras sequences replace the insert, the binding event has no effect on the dissociation rate of GDP from the chimera. In addition, whereas the fluorescence of mant-GDP bound to wild-type Cdc42Hs is quenched by 20% upon GDI binding, no quench is observed in similar experiments with the Cdc42Hs chimera (Wu *et al.*, 1997). These studies suggest a regulatory role for the insert region in the function of rho subfamily proteins, namely, transducing the signal from GDI binding to inhibition of GDP dissociation. While it is known that GDI binding to Cdc42Hs requires the prenylated C-terminus (Leonard *et al.*, 1992), the high mobility of the C-terminal region and its location on the opposite side of the protein from the GDP binding site suggest that more proximal residues, such as the insert region, would be required for GDI-induced inhibition of GDP dissociation. However, because of the high flexibility of the C-terminal region, the exact interaction cannot be predicted. An attractive hypothesis is that, after binding to the prenylated C-terminus, the flexibility of the C-terminal region would allow RhoGDI to interact with the insert region, which lies adjacent to residues that line the nucleotide binding pocket. On the basis of modeling studies which suggested that the insert region could cover the guanine nucleotide binding site (Sutcliffe *et al.*, 1994), it was originally assumed that the effect of RhoGDI binding would be to stabilize this conformation. The possibility remains that binding of RhoGDI could induce the insert to occlude the nucleotide binding site, thereby impeding diffusion of GDP out of the binding site, or that RhoGDI binding induces the insert to contribute side chains to the binding of the GDP (an interesting proposition given the basic nature of the insert sequence and the acidic nature of the nucleotide). Although the exact details of this interaction cannot be determined without the structure of the complex, it is important to note that no known GDIs exist for Ras proteins, which lack the insert.

The H-ras structure can be viewed as a minimal scaffolding for guanine nucleotide binding and hydrolysis into which structural insertions can be built at a number of locations (Bourne *et al.*, 1991). The solution structure of Cdc42Hs is no exception to this rule; however, it is the structural differences between Cdc42Hs and other GTP-binding proteins that confer the specificity of Rho proteins and ultimately their unique biological activity. Clearly the switch surface with the Rho-specific insert along one edge is a crucial element in the specificity of these proteins, as is the flexible C-terminal tail which binds GDI proteins. Building on the solution structure of Cdc42Hs, further work on the complexes between Cdc42Hs and its regulatory and effector proteins will be essential for understanding the molecular details of the biological function of this important signaling pathway.

## ACKNOWLEDGMENT

We are extremely grateful to Linda Nicholson, Kwaku Dayie, Greg Heffron, and Sekhar Talluri for many helpful discussions and to Greg Warren (Yale University) for advice in setting up the chemical shift and Ramachandran refinement protocols.

## SUPPORTING INFORMATION AVAILABLE

Resonance assignments of the GDP- and GMPPCP-bound forms of Cdc42Hs (10 pages). Ordering information is given on any current masthead page.

## REFERENCES

Adams, E. M. A., Johnson, D. I., Longnecker, R. M., Sloat, B. F., & Pringle, J. R. (1990) *J. Cell Biol. 111*, 131−142.
Aevarsson, A., Brazhnikov, E., Garber, M., Zheltonosova, J., Chirgadze, Y., Al-Karadaghi, S., Svensson, L. A., & Liljas, A. (1994) *EMBO J. 13*, 3669−3677.
Amor, J. C., Harrison, D. H., Kahn, R. A., & Ringe, D. (1994) *Nature 372*, 704−708.
Bagrodia, S., Derigard, B., Davis, R. J., & Cerione, R. A. (1995) *J. Biol. Chem. 270*, 27995−27998.
Barbacid, M. (1987) *Annu. Rev. Biochem. 56*, 779−827.
Barfod, E. T., Zheng, Y., Kuang, W.-J., Hart, M. J., Evans, T., Cerione, R. A., & Ashkenazi, A. (1993) *J. Biol. Chem. 268*, 26059−26062.
Bartels, C., Xia, T.-H., Billeter, M., Guntert, P., & Wuthrich, K. (1995) *J. Biomol. NMR 5*, 1−10.
Bodenhausen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185−189.
Boguski, M. S., & McCormick, F. (1993) *Nature 366*, 643−654.
Bourne, H. R., Sanders, D. A., & McCormick, F. (1991) *Nature 349*, 117−127.
Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
Brill, S., Li, S., Lyman, C. W., Church, D. M., Wasmuth, J. J., Weissbach, L., Bernards, A., & Snijders, A. J. (1996) *Mol. Cell. Biol. 16*, 4869−4878.
Brown, S. C., Weber, P. L., & Mueller, L. (1988) *J. Magn. Reson. 77*, 166−169.
Brunger, A. T. (1996) *X-PLOR Manual Version 3.843*, Yale University, New Haven, CT.
Chant, J., & Stowers, L. (1995) *Cell 81*, 1−4.
Chou, M. M., & Blenis, J. (1996) *Cell 85*, 573−583.
Coso, O. A., Chiariello, M., Yu, J.-C., Teramoto, H., Crespo, P., Xu, N., Miki, T., & Gutkind, J. S. (1995) *Cell 81*, 1137−1146.
Czworkowski, J., Wang, J., Steitz, T. A., & Moore, P. B. (1994) *EMBO J. 13*, 3661−3668.
Dötsch, V., Oswald, R. E., & Wagner, G. (1995) *J. Magn. Reson., B 108*, 285−288.
Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.
Freeman, J. L., Abo, A., & Lambeth, J. D. (1996) *J. Biol. Chem. 271*, 19794−19801.
Greasley, S. E., Jhoti, H., Teahan, C., Solari, R., Fensome, A., Thomas, G. M. H., Cockcroft, S., & Bax, B. (1995) *Nat. Struct. Biol. 2*, 797−806.
Grzesiek, S., & Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
Hart, M. J., Eva, A., Evans, T., Aaronson, S. A., & Cerione, R. A. (1991a) *Nature 354*, 311−314.
Hart, M. J., Shinjo, K., Hall, A., Evans, T., & Cerione, R. A. (1991b) *J. Biol. Chem. 266*, 20840−20848.
Hart, M. J., Eva, A., Zangrilli, D., Aaronson, S. A., Evans, T., Cerione, R. A., & Zheng, Y. (1994) *J. Biol. Chem. 269*, 62−65.
Hart, M. J., Callow, M. G., Souza, B., & Polakis, P. (1996) *EMBO J. 15*, 2997−3005.
Hirshberg, M., Stockley, R. W., Dodson, G., & Webb, M. R. (1997) *Nat. Struct. Biol. 4*, 147−152.
Horii, Y., Beeler, J. F., Sakaguchi, K., Tachibana, M., & Miki, T. (1994) *EMBO J. 13*, 4776−4786.

Hotta, K., Tanaka, K., Mino, A., Kohno, H., & Takai, Y. (1996) *Biochem. Biophys. Res. Commun. 225*, 69−74.
John, J., Sohmen, R., Feuerstein, J., Linke, R., Wittinghofer, A., & Goody, R. S. (1990) *Biochemistry 29*, 6058−6065.
Johnson, D. I., & Pringle, J. R. (1990) *J. Cell Biol. 111*, 143−152.
Kay, L. E., Keifer, P., & Saarinen, T. (1992) *J. Am. Chem. Soc. 114*, 10663−10665.
Kay, L. E., Xu, G.-Y., Singer, A. U., Muhandiram, D. R., & Forman-Kay, J. D. (1993) *J. Magn. Reson., B 101*, 333−337.
Kay, L. E., Xu, G. Y., & Yamazaki, T. (1994) *J. Magn. Reson., A 109*, 129−133.
Kelley, L. A., et al. (1997) *Protein Eng.* (in press).
Kelley, L. A., Gardner, S. P., & Sutcliffe, M. J. (1996) *Protein Eng. 9*, 1063−1065.
Koradi, R., Billeter, M., & Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
Kozma, R., Ahmed, S., Best, A., & Lim, L. (1995) *Mol. Cell. Biol. 15*, 1942−1952.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Kraulis, P. J., Domaille, P. J., Campbell-Burk, S. L., Van Aken, T., & Laue, E. D. (1994) *Biochemistry 33*, 3515−3531.
Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1−6.
Kupce, E., & Wagner, G. (1995) *J. Magn. Reson., B 109*, 329.
Kupce, E., & Freeman, R. (1996) *J. Magn. Reson., A 118*, 299.
Kuroda, S., Fukata, M., Kobayashi, K., Nakafuku, M., Nomura, N., Iwamatsu, A., & Kaibuchi, K. (1996) *J. Biol. Chem. 271*, 23363−23367.
Kuszewski, J., Gronenborn, A. M., & Clore, G. M. (1995a) *J. Magn. Reson., B 107*, 293−297.
Kuszewski, J., Qin, J., Gronenborn, A. M., & Clore, G. M. (1995b) *J. Magn. Reson., B 106*, 92−96.
Kuszewski, J., Gronenborn, A. M., & Clore, G. M. (1996) *Protein Sci. 5*, 1067−1080.
Lambright, D. G., Sondek, J., Bohm, A., Skiba, N. P., Hamm, H. E., & Sigler, P. B. (1996) *Nature 379*, 311−319.
Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283−291.
Leonard, D., Hart, M. J., Platko, J. V., Eva, A., Henzel, W., Evans, T., & Cerione, R. A. (1992) *J. Biol. Chem. 267*, 22860−22868.
Leonard, D. A., Evans, T., Hart, M., Cerione, R. A., & Manor, D. (1994) *Biochemistry 33*, 12323−12328.
Manser, E., Leung, T., Salihuddin, H., Tan, L., & Lim, L. (1993) *Nature 363*, 364−367.
Marion, D., Ikura, M., Tschudin, R., & Bax, A. (1989a) *J. Magn. Reson. 85*, 393−399.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989b) *J. Am. Chem. Soc. 111*, 1515−1517.
Markby, D. W., Onrust, R., & Bourne, H. R. (1993) *Nature 262*, 1895−1901.
Matsuo, H., Kupce, E., Li, H., & Wagner, G. (1996) *J. Magn. Reson., B 113*, 91.
McCallum, S. J., Wu, W. J., & Cerione, R. A. (1996) *J. Biol. Chem. 271*, 21732−21737.
Merritt, E. A., & Murphy, M. E. P. (1994) *Acta Crystallogr. D50*, 869−873.
Minden, A., Lin, A., Claret, F.-X., Abo, A., & Karin, M. (1995) *Cell 81*, 1147−1157.
Muchmore, D. C., McIntosh, L. P., Russell, C. B., Anderson, D. E., & Dahlquist, F. W. (1989) *Methods Enzymol. 177*, 44−73.
Muhandiram, D. R., & Kay, L. E. (1994) *J. Magn. Reson., B 103*, 203−216.
Nobes, C. D., & Hall, A. (1995) *Cell 81*, 53−62.
Nomanbhoy, T., Leonard, D. A., Manor, D., & Cerione, R. A. (1996) *Biochemistry 35*, 4602−4608.
Norwood, T. J., Boyd, J., Heritage, J. E., Soffe, N., & Campbell, I. D. (1990) *J. Magn. Reson. 87*, 488−501.
Pai, E. F., Krengel, U., Petsko, G. A., Goody, R. S., Kabsch, W., & Wittinghofer, A. (1990) *EMBO J. 9*, 2351−2359.
Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C., & Nyborg, J. (1996) *Structure 4*, 1141−1151.
Ridley, A. (1995) *Curr. Opin. Gen. Dev. 5*, 24−30.
Rucker, S. P., & Shaka, A. J. (1989) *Mol. Phys. 68*, 509−517.
Scheffzek, K., Klebe, C., Fritz-Wolf, K., Kabsch, W., & Wittinghofer, A. (1995) *Nature 374*, 378−381.
Shinjo, K., Koland, J. G., Hart, M. J., Narasimhan, V., Johnson, D. I., Evans, T., & Cerione, R. A. (1990) *Proc. Natl. Acad. Sci.* U.S.A. *87*, 9853−9857.
Shortle, D. (1994) *J. Magn. Reson., B 105*, 88−90.
Singer, W. D., Brown, H. A., Bokoch, G. M., & Sternweis, P. C. (1995) *J. Biol. Chem. 270*, 14944−14950.
States, D. J., Haberkorn, R. A., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286−292.
Sutcliffe, M. J., Feltham, J., Cerione, R. A., & Oswald, R. E. (1994) *Protein Pept. Lett. 1*, 84−91.
Symons, M., Derry, J. M. J., Karlak, B., Jiang, S., Lemahieu, V., McCormick, F., Francke, U., & Abo, A. (1996) *Cell 84*, 723−734.
Szyperski, T., Neri, D., Leiting, B., Otting, G., & Wüthrich, K. (1992) *J. Biomol. NMR 2*, 323−334.
Talluri, S., & Wagner, G. (1996) *J. Magn. Reson., B 112*, 200−205.
Tong, L., de Vos, A. M., Milburn, M. V., & Kim, S.-H. (1991) *J. Mol. Biol. 217*, 503−516.
Vojtek, A. B., & Cooper, J. A. (1995) *Cell 82*, 527−529.
Vuister, G. W., Clore, G. M., Gronenborn, A. M., Powers, R., Garrett, D. S., Tschudin, R., & Bax, A. (1993) *J. Magn. Reson., B 101*, 210−213.
Wall, M. A., Coleman, D. E., Lee, E., Iniguez-Lluhi, J. A., Posner, B. A., Gilman, A.G., & Sprang, S. R. (1995) *Cell 83*, 1047−1058.
Wishart, D. S., & Sykes, B. D. (1994) *J. Biomol. NMR 4*, 171−180.
Wishart, D. S., Sykes, B. D., & Richards, F. M. (1992) *Biochemistry 31*, 1647−1651.
Wittinghofer, A., & Pai, E. (1991) *Trends Biochem. Sci. 16*, 382−387.
Wu, W. J., Leonard, D. A., Paar, J. M., Cerione, R., & Manor, D. (1997) *J. Biol. Chem.* (submitted for publication).
Yamazaki, T., Lee, W., Arrowsmith, C. H., Muhandiram, D. R., & Kay, L. E. (1994a) *J. Am. Chem. Soc. 116*, 11655−11666.
Yamazaki, T., Lee, W., Revington, M., Mattiello, D. L., Dahlquist, F. W., Arrowsmith, C. H., & Kay, L. E. (1994b) *J. Am. Chem. Soc. 116*, 6464−6465.
Zhang, S., Han, J., Sells, N. A., Chernoff, J., Knaus, U. G., Ulevitch, R. J., & Bokoch, G. M. (1995) *J. Biol. Chem. 270*, 23934−23936.
Zheng, Y., Bagrodia, S., & Cerione, R. A. (1994) *J. Biol. Chem. 269*, 18727−18730.

BI970694X