Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Analysis of protein hydration in ultrahigh-resolution structures of the SRP GTPase Ffh

Ursula D. Ramirez and Douglas M. Freymann*

Department of Molecular Pharmacology and Biological Chemistry, Northwestern University, Chicago, IL 60611, USA

Correspondence e-mail: freymann@northwestern.edu

Received 8 June 2006
Accepted 3 October 2006

PDB References: apo Ffh NG, structure I, 2j46, r2j46sf; structure II, 2j45, r2j45sf.

Two new structures of the SRP GTPase Ffh have been determined at 1.1 Å resolution and provide the basis for comparative examination of the extensive water structure of the apo conformation of these GTPases. A set of well defined water-binding positions have been identified in the active site of the two-domain 'NG' GTPase, as well as at two functionally important interfaces. The water hydrogen-bonding network accommodates alternate conformations of the protein side chains by undergoing local rearrangements and, in one case, illustrates binding of a solute molecule within the active site by displacement of water molecules without further disruption of the water-interaction network. A subset of the water positions are well defined in several lower resolution structures, including those of different nucleotide-binding states; these appear to function in maintaining the protein structure. Consistent arrangements of surface water between three different ultrahigh-resolution structures provide a framework for beginning to understand how local water structure contributes to protein–ligand and protein–protein binding in the SRP GTPases.

1. Introduction

The interaction between water and protein has been shown to contribute to a variety of functionally important phenomena, providing the basis for structural stability, binding flexibility in the active site, proton transfer, adaptability to environmental change, motion, both in terms of side-chain conformational changes (Mattos, 2002; Nicholls, 2000) and large-scale protein motions (Mattos & Ringe, 2001), and enzymatic mechanism (Mattos, 2002; Nicholls, 2000) and playing a key role in many protein–protein interactions (Lo Conte *et al.*, 1999). In hemoglobin, the redistribution of a cluster of water molecules at the subunit interface is a mediator of the allosteric transition (Knapp *et al.*, 2006; Pardanani *et al.*, 1998; Royer *et al.*, 1996). Water has also been shown to be involved in large-scale surface hydrogen-bonding networks (Nakasako, 1999). Since water has been shown to play roles in binding and protein dynamics, demonstration of the existence of these networks and description of their features on the surfaces of other proteins is likely to elucidate roles in protein structure and function.

Water is often poorly resolved in protein structures at moderate resolution (*e.g.* ∼2.0 Å), so that only the most well ordered molecules are visualized in the electron-density map. Different structures of the same or related proteins suggest that water molecules that are crucial to the structure and/or

function of the protein are well ordered (Mattos, 2002) and are thus observed in multiple structures of a given protein, while those that are not crucial occur only in a subset of those structures (Mattos & Ringe, 2001). Two studies, one of a family of microbial ribonucleases (Loris *et al.*, 1999) and the other of the legume lectin family (Loris *et al.*, 1994), found that the locations of only a small number of water positions are independent of crystal-packing environment and solution conditions. However, as the resolution of crystallographic data increases, the ability to resolve and identify loosely bound or partially occupied water molecules increases significantly. This implies that identification of water molecules which are essential to the structure, dynamics and function of a protein should include a comparison of multiple structures of the same or related proteins at ultrahigh resolution (Zhang *et al.*, 2000).

Crystals of one of the GTPase components of the signal recognition particle (SRP), Ffh NG, diffract to >1.0 Å resolution. The SRP GTPases function in co-translational targeting of nascent membrane and secreted proteins to the membrane (Keenan *et al.*, 2001). SRP binds the hydrophobic signal sequence of the nascent protein and targets it and the ribosomal translation machinery to the endoplasmic reticulum, where SRP binds its receptor (SR) to form an SRP–SR complex. GTP binding by both the SRP subunit, SRP54, and by SR is required for their interaction. The nascent protein is subsequently transferred to a membrane translocon and the SRP cycle completes when SRP and SR each hydrolyze one molecule of GTP and disengage (Focia, Shepotinovskaya *et al.*, 2004). Homologues of these two components are found throughout evolution and are termed Ffh and FtsY, respectively, in prokaryotes (Keenan *et al.*, 2001). The GTPase domains of Ffh and FtsY are structurally homologous and are termed the NG domain: the N domain is a four-α-helical bundle that appears to provide a contact between the SRP and the ribosome (Halic *et al.*, 2004) and its G domain is structurally similar to other members of the GTPase superfamily (Freymann *et al.*, 1997). Structures of the NG domains of Ffh and FtsY in the apo state (Freymann *et al.*, 1997; Gariani *et al.*, 2006; Montoya *et al.*, 1997, 2000) and in nucleotide-bound states (Freymann *et al.*, 1999; Padmanabhan & Freymann, 2001) have been determined, as has the structure of their nucleotide-dependent heterodimeric complex (Egea *et al.*, 2004; Focia, Shepotinovskaya *et al.*, 2004).

Previously, an analysis of mobility in the structure of the apo Ffh NG at 1.1 Å resolution (Ramirez *et al.*, 2002) allowed the identification of a small set of waters between the N and G domains that contribute to interface mobility; however, no thorough investigation of the water structure of this protein has been carried out previously. Here, we present two new ultrahigh-resolution structures of the apo Ffh NG, in a different crystal form from that previously determined, and use them to describe more fully the water structure of Ffh NG. We analyze the position and dynamics of the water molecules in these new structures with reference to those observed in the previous ultrahigh-resolution apo structure as well as the lower resolution apo (Freymann *et al.*, 1997), Mg^{2+}-GDP-bound (Freymann *et al.*, 1999) and GMPPNP-bound structures (Padmanabhan & Freymann, 2001). The structures reveal substantial disorder at the protein surface, visualized as alternative conformations of side chains that are coupled to clearly assigned shifts in the positions of surrounding water molecules. Despite this disorder and differences in crystal packing and solvent conditions, we can identify several networks of water molecules that are conserved between each of the 1.1 Å structures. The comparison reveals how the water network can adjust to binding of a small solute without disruption to the overall pattern of hydrogen bonding. Well ordered networks are associated with several conserved sequence motifs of the SRP GTPases: the first occupies the nucleotide-binding pocket, the second is associated with the N/G domain interface and contributes to interdomain mobility (Ramirez *et al.*, 2002) and the third is poised to contribute to the conformational changes that accompany the formation of the SRP–receptor complex (Focia, Shepotinovskaya *et al.*, 2004). These 1.1 Å crystal structures thus provide a snapshot of what are likely to be important water arrangements at the surface of the SRP GTPase Ffh and provide a framework for beginning to understand their functional roles.

2. Materials and methods

2.1. Crystallization and data collection

Ffh NG was expressed and purified as described previously (Freymann *et al.*, 1997; Shepotinovskaya *et al.*, 2003). Crystals of the apo Ffh NG domain were grown under two related conditions, which we term I and II. For structure I, 6 µl protein at 17 mg ml^{-1} in 2 m*M* $MgCl_2$ was added to an equal volume of the crystallization mother liquor: 0.1 *M* sodium acetate pH 4.7, 0.1 *M* $MnCl_2$, 30% MPD. For structure II, 4 µl of the protein at 34 mg ml^{-1} in 2 m*M* $MgCl_2$ was combined with an equal volume of mother liquor: 0.1 *M* MES pH 6.1, 20 m*M* $CaCl_2$, 30% MPD. Crystallization was performed by sitting-drop vapour diffusion at 277 K. In both cases a 'cross-dilution' streak-seeding protocol was used (Stura & Wilson, 1990, 1991): after 1 d incubation at 277 K, a nylon fibre was touched to seed crystals and swiped through a drop of protein-free mother liquor and a second fibre swiped across the trace of the first before swiping it through the protein-containing crystallization drop, thereby minimizing the number of nuclei transferred. For data collection, the crystals were mounted directly from their respective MPD mother liquors using a nylon loop and flash-cooled in liquid nitrogen (Teng, 1990).

Data for structure I were measured on a MAR 345 area detector at SSRL BL 9-1 using a wavelength of 0.785 Å in three overlapping passes to allow scaling. The high-resolution data were collected in dose mode at a crystal-to-detector distance of 188 mm (1.08 Å at the edge of the detector) with an exposure time of ∼6 min per 0.5° oscillation frame. The medium-resolution data were collected with an exposure time of approximately 2.5 min per 1.5° oscillation frame. The low-resolution data set was collected with an exposure time of 10 s per 2° oscillation. Data for structure II were measured at APS DND-CAT beamline 5ID-B using a wavelength of 1.00 Å and

Table 1
Crystallographic statistics.

Values in parentheses are for the last resolution shell.

	I	II
Data collection		
Space group	*C*2	*C*2
Unit-cell parameters		
a (Å)	109.47	108.60
b (Å)	54.50	54.30
c (Å)	98.80	99.78
β (°)	97.72	99.36
Resolution range (Å)	18.5–1.10	30.0–1.14
Unique reflections	199458	195214
$R_{\rm sym}$ (%)	3.8 (26.2)	5.1 (30.8)
Completeness (%)	94.6 (80.0)	93.0 (88.8)
Redundancy	2.6 (1.8)	3.8 (3.2)
Average $I/\sigma(I)$	30.6 (2.6)	24.6 (4.0)
Refinement		
No. of test-set reflections	9676	13120
$R_{\rm cryst}$ (%)	14.9	11.9
$R_{\rm free}$ (%)	17.2	15.4
No. of protein atoms†	5307	6756
No. of water molecules†	554	893
No. of atoms in alternate conformations	744	2155
Average temperature factor (Å^2)		
Protein	23.0	16.6
Water	35.6	29.7
Alternate conformations	23.8	19.9
R.m.s.d. bond lengths (Å)	0.015	0.015
R.m.s.d. bond angles (°)	1.621	2.099

† No. of atoms includes atoms in alternate conformations.

a 165 cm MAR CCD detector. In initial images multiple lattices were apparent in the diffraction pattern; however, following *in situ* annealing (blocking the cryostream twice for 2–3 s and allowing the crystal to re-freeze in the cryostream; Kriminski *et al.*, 2002) only one clean diffraction lattice was visible. Data were collected in three passes: the high-resolution data were collected at a crystal-to-detector distance of 52 mm (1.05 Å at the edge of detector) with an exposure time of 3.2 s per 0.5° oscillation frame. The medium-resolution data were collected at a crystal-to-detector distance of 120 mm with an exposure time of 2.5 s per 1.5° oscillation frame. The low-resolution data set was collected at a crystal-to-detector distance of 200 mm with an exposure of 1.65 s per 2° oscillation. Data for both structures were integrated using *DENZO* with a -3σ cutoff and were scaled using *SCALEPACK* (Otwinowski, 1993). Data-collection statistics are shown in Table 1.

2.2. Structure refinement

The initial model for crystallographic refinement of both structures was the Ffh NG–GMPPNP complex (PDB code 1jpn; Padmanabhan & Freymann, 2001) with all solvent and ligand atoms removed. This crystal form is obtained under similar crystallization conditions in space group *C*2, with two monomers (termed *A* and *B*) in the asymmetric unit. Rigid-body refinement in the resolution range 15–3 Å was carried out first, using *REFMAC* (Murshudov *et al.*, 1997, 1999), and was followed by 30 cycles of refinement over the resolution range 18.5–1.10 Å for structure I and 30.0–1.135 Å for structure II. An initial set of solvent atoms was then added automatically using *ARP*/*wARP* (Lamzin *et al.*, 2001; Perrakis *et al.*, 1999). Subsequent refinement with *REFMAC* (Murshudov *et al.*, 1997, 1999) incorporated anisotropic displacement parameters, explicit H atoms and a bulk-solvent model. Electron-density maps were calculated using the *CCP*4 suite (Collaborative Computational Project, Number 4, 1994) and the model was inspected using *O* (Jones *et al.*, 1991). Alternate conformations were modeled when supported both by well defined features in the difference map contoured at 3σ and by the presence of persistent features in the $2F_{\rm o} - F_{\rm c}$ map; they were managed using locally written scripts (Ramirez, 2005) and the program *ACONIO* (Kleywegt *et al.*, 2001). The occupancies of residues in alternate conformations were adjusted manually following inspection of difference maps. Test refinements using *REFMAC* demonstrated an increase in $R_{\rm free}$ of 0.5–1.0% when alternate conformations were refined using an isotropic *B*-factor model (Wilson & Brunger, 2000), so all atoms, including main-chain atoms, were modeled with anisotropic temperature factors. During the final stages of refinement, anisotropic temperature factors were evaluated using *PARVATI* (Merritt, 1999) and the sphericity, rigid-bond and *B*-factor restraints ('SPHE', 'RBON' and 'BFAC', respectively; Murshudov *et al.*, 1997, 1999) adjusted such that the optimized anisotropy profiles of each closely resembled those of other proteins (Merritt, 1999). This procedure resulted in an $\sim$1% decrease in $R_{\rm cryst}$ and $R_{\rm free}$ for each structure.

All residues in each structure are in the allowed regions of the Ramachandran plot (Lovell *et al.*, 2003). More interestingly, evaluation of the ω-angle distributions revealed that 92% of residues are in the 'core' ω-angle region, 6% are in the 'allowed' region and five residues in each structure are in the 'generous' region (Morris *et al.*, 1992; Willard *et al.*, 2003). A striking deviation from ω planarity at Phe102/Leu103 was noted previously (Ramirez *et al.*, 2002) and occurs here in all monomers (ω ranges from 159 to 163°). This deviation occurs within the hydrophobic core of the G domain; a second outlier occurs in structure II, but not structure I, at Leu106/Gln107 of the motif I P-loop (ω values of 160 and 163° in the two monomers). Both deviations are real features of the structures, as evidenced by their reproducibility across the independently refined monomers in each crystal form.

A number of features of the electron-density map were not interpretable in terms of water molecules and were modeled as mother-liquor solutes. In structure I, two acetate groups, two manganese ions and two molecules of MPD were easily discerned in the electron-density map and two additional features were tentatively assigned as chloride ions. In structure II, four calcium ions, one MPD and two MES molecules, one of which is modeled in two conformations, were readily identified. In addition, one feature is tentatively identified as a sodium ion and four distinct additional features in the electron-density map are modeled as partially disordered or uninterpretable solutes. Difference peaks that could be interpreted as H atoms were observed in both structures, but because they were not systematically resolved in either of the

Table 2
Statistics for structures compared.

	Structure I	Structure II	1ls1	1ffh	1ng1	1jpn
Binding state	Apo	Apo	Apo	Apo	Mg^{2+}-GDP	GMPPNP
Crystallization conditions	30% MPD, 0.1 *M* NaOAc pH 4.7, 0.1 *M* $MnCl_2$, 2 m*M* $MgCl_2$	30% MPD, 0.1 *M* MES pH 6.1, 20 m*M* $CaCl_2$, 2 m*M* $MgCl_2$	30% PEG MME 550, 0.2 *M* TAPS pH 9.0, 0.2 *M* $MgCl_2$	30% PEG MME 550, 0.1 *M* TAPS pH 9.0, 0.2 *M* $MgCl_2$	10% PEG 8K, 0.1 *M* sodium cacodylate pH 6.5, 0.2 *M* $Mg(OAc)_2$, 50 m*M* $CdSO_4$	29% MPD, 0.1 *M* NaOAc pH 4.7, 20 m*M* $CaCl_2$, 2 m*M* $MgCl_2$
Space group	*C*2	*C*2	*C*2	*C*2	$P4_32_12$	*C*2
Unit-cell parameters						
a (Å)	109.47	108.60	99.73	99.90	100.07	108.82
b (Å)	54.50	54.30	53.67	53.91	100.07	54.53
c (Å)	98.80	99.78	57.84	57.36	73.49	99.08
β (°)	97.72	99.36	119.92	119.77		97.42
No. of molecules in ASU	2	2	1	1	1	2
Crystallographic refinement						
Resolution (Å)	1.10	1.14	1.10	2.05	2.03	1.90
R_{cryst} (%)	14.9	11.9	13.7	18.6	24.2	19.0
R_{free} (%)	17.2	15.4	16.9	24.8	26.3	24.0
No. of non-H protein atoms	5307	6756	2794	2226	2275	4604
No. of residues in alternate conformations	90	261	33	0	0	7
Solvent content (%)	45.3	45.2	47.9	47.8	56.2	45.1
No. of waters (unique)	548	809	279	116	165	424
Waters built† (%)	24.8	36.6	22.6	9.4	9.7	19.2
No. of waters in alternate conformations	6	84	29	0	0	0

† [(Unit-cell volume) × (% solvent)]/[(No. of asymmetric units in unit cell) × (30 Å^3)].

structures, further analysis of hydrogen positions was not pursued. Refinement statistics are presented in Table 1.

2.3. Regularization and characterization of water structure

Water molecules were classified as belonging to monomer *A* or *B* in each structure using *SORTWATER* (Collaborative Computational Project, Number 4, 1994). Waters at crystal contacts were identified using *ACT* and in-house scripts (Collaborative Computational Project, Number 4, 1994; Ramirez, 2005) to apply a cutoff distance of 3.5 Å; these were excluded from further analysis, except as noted below. The two monomers (*A* and *B*) of each crystal structure were considered individually (and are referred to in the text as I*A*, I*B*, II*A*, II*B*). Water molecules were assigned to hydration shells as defined conventionally (Beamer *et al.*, 2005). Waters of the first hydration shell were identified using in-house scripts to extract water–protein contacts (3.2 Å cutoff) using the *CCP*4 programs *ACT* and *CONTACT* (Collaborative Computational Project, Number 4, 1994). Higher hydration shells were identified by recursively flagging hydrogen bonds output by the programs *REDUCE* (Word *et al.*, 1999) and *PROBE* (Word *et al.*, 1999) using in-house scripts to identify water molecules whose interactions were limited to the next lower hydration shell. Waters that could not be classified according to these criteria (*e.g.* 12 undefined in PDB code 1ls1) generally occurred distant from the protein surface and were excluded from further analysis. Water-molecule accessibilities were calculated using the program *AREAIMOL* (Collaborative Computational Project, Number 4, 1994).

The apo Ffh NG structures I and II presented here were also compared with four previously published crystal structures of Ffh NG: 1ls1 (Ramirez *et al.*, 2002) and 1ffh (Freymann *et al.*, 1997), which are the same crystal form of the apo protein determined at 1.1 and 2.0 Å resolution, respectively, 1jpn (Padmanabhan & Freymann, 2001), GMPPNP-bound, but crystallized under conditions similar to those of structures I and II and determined at 2.0 Å, and 1ng1 (Freymann *et al.*, 1999), an Mg^{2+}-GDP complex in an unrelated crystal form determined at 2.0 Å (Table 2). Since several of the structures exhibit noncrystallographic symmetry, nine different monomers in all, each treated separately as described above, were compared. The program *LSQMAN* (Kleywegt *et al.*, 2001) was used for superimposition and to evaluate the overlap of water positions between structures. Monomers were superimposed on the reference structure (II*A*, see below) using an explicit definition of the overlap region identified by the 'lsq' options in *O* (Jones *et al.*, 1991; Kleywegt & Jones, 1997). Waters in each probe structure were expanded by symmetry prior to the alignment and a 1.0 Å distance cutoff was used as the matching criterion (Bottoms *et al.*, 2006). A parallel analysis was carried out by generating maps of hydrogen-bonding partners using *ACT* (Collaborative Computational Project, Number 4, 1994) and matching the conservation of water-interaction partners using in-house scripts (Ramirez, 2005). Although Ffh NG exhibits interdomain motion (Freymann *et al.*, 1999), the apo and GMPPNP-bound structures superimpose with an r.m.s. deviation of ∼0.4 Å on C^{α} positions, so that a simple superposition suffices for matching water positions throughout the structure (the r.m.s. deviation is similar for the superimposition of the N and G domains separately). In the case of the Mg^{2+}-GDP complex, however, the superposition is problematic. Large deviations in the position of the N domain relative to the G domain interfere with any simple

matching algorithm and hydrogen-bonding interactions were used to confirm the matching assignments; where the two methods disagreed, the appropriate match was determined by inspection in *O* (Jones *et al.*, 1991).

Water accessibilities were obtained using *AREAIMOL* (Collaborative Computational Project, Number 4, 1994) in HOH mode (*i.e.* ignoring other water molecules in the calculation), using the noncrystallographic pair and treating alternate water positions as separate cases. Waters with an accessible surface area less than 10 Å^2 were classified as 'buried'. Water clusters were mapped from water contact lists generated by *CONTACT* (Collaborative Computational Project, Number 4, 1994) to build a recursive interaction tree using in-house scripts (Ramirez, 2005). Those clusters within hydrogen-bonding distance of conserved residues of the GTPase sequence motifs I–IV were inspected using *O* (Jones *et al.*, 1991) in order to manually expand the interaction network by including protein hydrogen-bonding partners that bridged clusters.

Finally, when well defined water molecules were found to be conserved in most of the ultrahigh-resolution structures, but not all of them, difference maps contoured at 2σ and 3σ were reviewed to check for errors or omissions during building. A similar analysis of the absence of conserved water positions was carried out with respect to the lower resolution structures. Water numbering was defined to be consistent among struc-

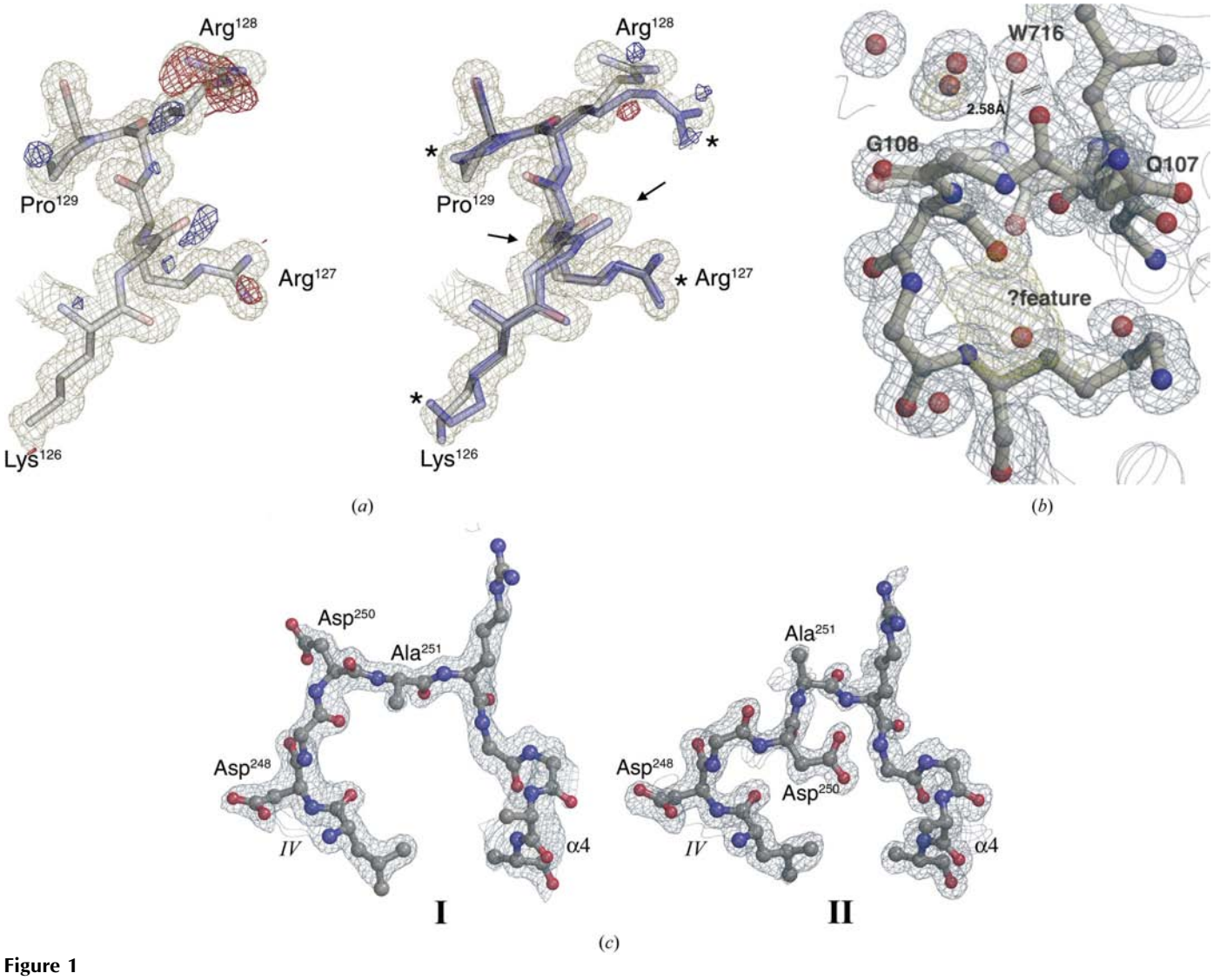

Figure 1
Features of the electron-density maps that reflect alternate conformations. (*a*) Before (left) and after (right) modeling alternate configurations of residues 126–129 in structure II*A*. The $2F_o - F_c$ map contoured at 1σ is shown in light blue, the negative difference map ($F_o - F_c$ contoured at -3σ) shown in red and the positive difference map (contoured at 3σ) shown in bright blue. The same maps are shown at the same contour levels after the addition of alternates and are superimposed on the model of residues 126–129 with alternate conformations (shaded blue). The alternate side-chain positions are indicated by asterisks. The 'backrub' motion of Arg127 (Davis *et al.*, 2006) is indicated by two arrows. A significant translation of the peptide perpendicular to the viewer is obscured in this orientation (see Fig. 2*a*). (*b*) A peptide flip between Gln107 and Gly108 of the motif I P-loop in structure II*A*. Note the water hydrogen bonded to the 'flipped' configuration of the Gly108 amide nitrogen. There is an 'unexplained' density (likely to be a buffer molecule) bound within the P-loop (at the center). (*c*) Two conformations of the conserved DARGG loop, the first in structure I*A* and the second in structure II*A*.

tures following the superposition, first assigning the water numbering already present in the 1ls1 PDB coordinate set and subsequently (for water molecules not present in 1ls1) recursively assigning numbering to match within themselves. Thus, waters present in 1ls1 are assigned 401+, waters present in I*A* assigned 1001+, I*B* 2001+, II*A* 3001+ and unassigned (or unmatched) waters given numbers 4001+.

3. Results and discussion

3.1. Quality of the ultrahigh-resolution structures

The two structures reported here have been determined to ∼1.1 Å resolution and refined to crystallographic *R* factors of 0.149 (structure I; $R_{\rm free}$ = 0.172) and 0.119 (structure II; $R_{\rm free}$ = 0.154) (Table 1). Our goal in this work was to exploit multiple ultrahigh-resolution crystal structures to identify those aspects of the water structure that are maintained between different crystal forms and to gain some insight into both what that water structure was and its possible functional roles. The rationale was that at ultrahigh resolution water molecules are particularly well defined in the electron-density map and features of the water structure are more easily discerned. However, a number of constraints on this project arise, the most significant in our work being (i) the relative disorder of the solvation interactions actually observed, which limits the descriptive power of the analysis, (ii) the presence of common and extensive crystal-packing interactions shared between the different crystal forms, which limits the generalizability of our observations, and (iii) the propensity of the protein, particularly within its active site, to bind solutes (*e.g.* buffer molecules), such that any 'pristine' water structure there must necessarily be disrupted.

While the electron-density maps for both structures I and II are very clear overall, they show extensive evidence of distinct structural substates, not only for side chains exposed at the surface, but also for the underlying main-chain structure and the associated water molecules (Fig. 1). This result is not unexpected, as it has been shown that an electron-density map can generally be best modeled by multiple conformationally distinct substate structures (DePristo *et al.*, 2004). However, the atomic models for the two structures incorporate markedly different numbers of residues in alternate conformations (90 in I, 261 in II; Table 2). For structures I and 1ls1 the number of well defined alternates (15 and 11%, respectively, Table 2) is typical of that seen in other structures determined at similar resolution (typically ∼10–15%; Simonovic & Volz, 2001; Terzyan *et al.*, 2004). The number of residues exhibiting alternate conformations in structure II is notably high (261 out of 594, or 44%). Alternate conformations were generally introduced adjacent to residues that exhibited side-chain torsion rearrangements and we found that this strategy provided the best model based both on the resulting electron-density maps and on the improvement in the geometry of the protein model; thus, while some main-chain shifts are somewhat subtle, such as the two main-chain conformations for Arg127, which are 0.7 Å apart at the C^{α} atoms (Fig. 1*a*), and the ring pucker flip seen at Pro129 (see Fig. 2*a*), modeling these shifts explicitly improves the fit between the model and the data. Further, many of the alternate conformations occur in regions known to exhibit conformational plasticity from comparison of structures of the Ffh NG domain determined at lower resolution (Focia, Alam *et al.*, 2004; Freymann *et al.*, 1999; Ramirez *et al.*, 2002). The structure presented here exhibits the lowest value of $R_{\rm free}$ (15.4%) obtained during an extensive period of model building and refinement that

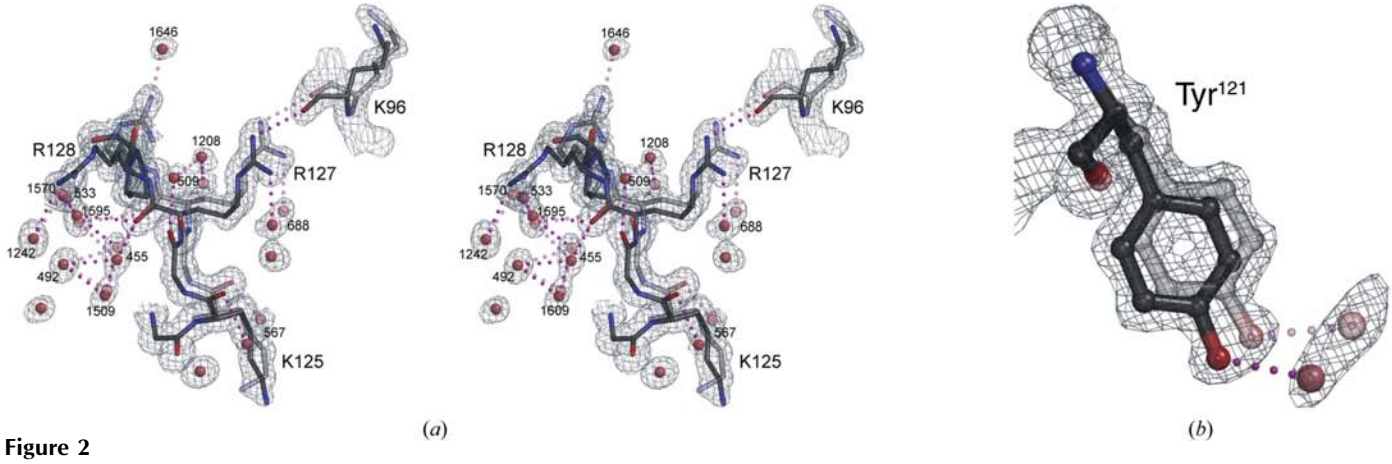

Figure 2
Hydrogen-bonding interactions of alternate and water structure. (*a*) $2F_{\rm o} - F_{\rm c}$ density at 1σ over a stretch of residues (in structure II*A*) that have alternate main-chain conformations (124–129), the neighbouring Lys96 and several water molecules. Protein, water and hydrogen bonding are shown in a bright colour for the first (alt *A*) conformation and in a muted colour for the second (alt *B*) conformation. Broken magenta lines indicate hydrogen bonds between water positions and protein-atom positions. The hydrogen bond between the carbonyl O atom of Lys96 and the side-chain NH2 of Arg127 is maintained by the two conformations of each of these residues. Peanut-shaped density is likely to reflect alternate positions of two hydrogen-bonded water molecules (waters 455, 1509, 533, 688 and 1595) which shift in order to maintain hydrogen-bonding interactions with adjacent protein atoms. Note the water molecule that substitutes in the bonding pattern for a protein atom that changes conformations, such as Wat1570 when Arg128 is in its *B* conformation. Wat1242 bonds to either water Arg128 NH1 or Wat1570, depending on the conformational state of Arg128. Wat1208 is within hydrogen-bonding distance of both conformations of Wat509. (*b*) $2F_{\rm o} - F_{\rm c}$ electron density contoured at 1σ over alternate conformations of the side chain of Tyr121 and its associated water (1163) in structure II*A*. The water density exhibits a characteristic 'peanut' shape.

included tests of a number of alternate strategies, including assignment of isotropic temperature factors to main-chain atoms in alternate positions (as they tend not to undergo large torsional displacements) and the assignment of alternative conformations only to residues that exhibited clearly defined torsional jumps (*e.g.* side-chain χ-angle rotamers). In each case $R_{\rm free}$ increased by 0.5% or more. As the $R_{\rm cryst}$ and $R_{\rm free}$ for structure II are 3.0 and 1.8% less, respectively, than for structure I, the average temperature factors are $\sim$6 Å^2 lower (Table 1) and the electron-density map is of better quality, reflected both by better definition of protein alternate conformations and water positions (Table 2), this structure (II) was taken as the reference for the analysis presented here.

The number of water molecules observed in the two structures I and II is somewhat less than observed in other crystal structures at this resolution (Terzyan *et al.*, 2004); the fraction of the water molecules contained within the asymmetric unit (assuming 0.033 water molecules per Å^3) that are observed in the structures ranges from 22 to 37% (Table 2). This suggests that the solvation structure within the crystal is by no means fully resolved and indeed (as shown below) solvent molecules exhibit multiple substate arrangements that tend to disorder at the higher solvation shells (Teeter & Kossiakoff, 1984). Therefore, while the lack of solvent structure in the crystal limits its description to the first solvation shell adjacent to the protein surface, conversely, the ultrahigh-resolution electron-density maps provide an image of correlated protein and solvent motions that can be interpreted as coupling between protein structural substates and discrete surrounding solvent substates (Yamano *et al.*, 1997; Fig. 2).

Water structure around proteins can be characterized as a series of 'shells' defined by the number of hydrogen bonds traversed between an observed water molecule and a protein atom (Fig. 3*a*). Generally, only the first hydration shell is well

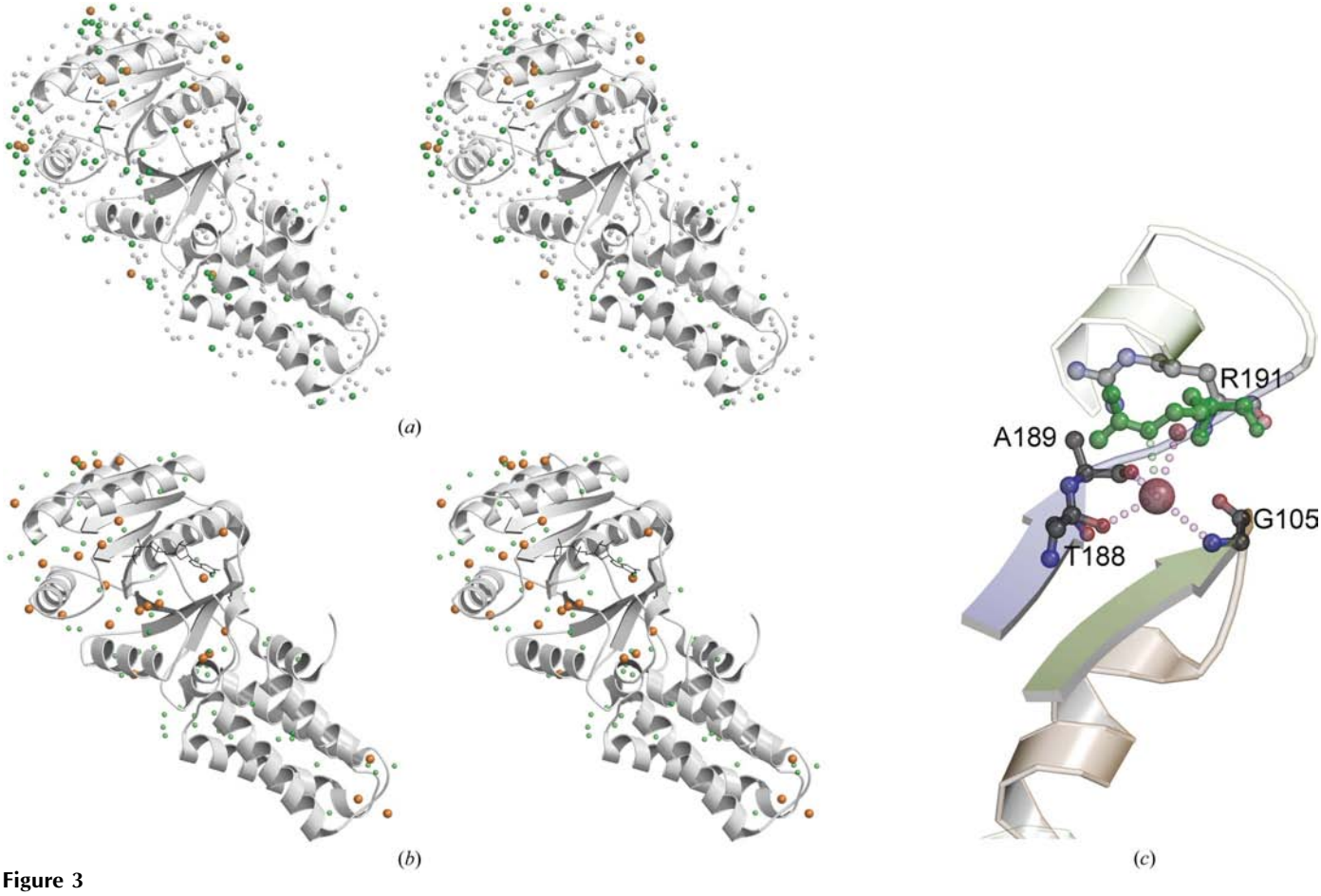

Figure 3
Water structure of Ffh NG. (*a*) Stereo image of the shell distribution of water molecules in structure II*A* is shown, colored by hydration shell, superimposed over a ribbon diagram of Ffh NG. The 344 first hydration-shell waters are shown as small light grey dots (to de-emphasize them). The 67 waters in the second hydration shell are shown as medium-sized green dots. The 14 waters in third hydration shell are shown as large orange dots. Second-shell and third-shell waters generally occur at crystal contacts (see text). (*b*) Stereo image of conserved water positions. 94 water positions that are conserved in every monomer of the 1.1 Å resolution Ffh NG apo structures are shown as small green dots on the ribbon diagram of Ffh NG. 28 of those waters are conserved across all binding states and crystal forms of *T. aquaticus* Ffh NG, including apo, Mg^{2+}-GDP-bound and GMPPNP-bound states and are shown as large orange dots. A stick representation of GTP in black indicates the binding site. (*c*) A water molecule that plays a structural role (shown as a large pink ball) makes hydrogen bonds that bridge the main-chain N atom of conserved Gly105 of motif I and the side-chain hydroxyl of conserved Thr188 and the carbonyl O atom of Ala189, both from motif III. The water is present in all structures of the Ffh NG domain. The water's final hydrogen-bonding partner is another water, shown as a small pink ball; in the structure of the SRP GTPase heterodimer (Focia, Shepotinovskaya *et al.*, 2004), its position becomes occupied by the side chain of Arg191 (shown in green).

ordered in protein structures and the extent of observed hydration is highly correlated with diffraction resolution (Britton *et al.*, 2006); water to the third hydration shell can be observed in the highest resolution protein structures available (Teeter, 1984, 1991). In structure II (which is taken as the reference) there are 504 first-shell waters. This is an average of 0.86 waters per residue and ∼80% of solvent-exposed polar protein atoms have a clearly defined solvent molecule within hydrogen-bonding distance. There are 156 well defined second-shell waters in the structure, but relatively few third-shell waters. (Note that waters directly involved in crystal-packing interactions are excluded.) Of the second-shell waters, ∼50% are members of pentameric ring-like hydrogen-bonding arrangements (Fig. 4*a*) that occur near hydrophobic solvent-exposed residues or in pockets formed by groups of hydrophobic side chains (Nakasako, 1999; Teeter, 1984). Such interactions are even more extensive if the hydrogen-bonding 'pentamer' is defined to include polar protein atoms as well as other water molecules and we use this definition in our discussion of water-bonding networks below. Not unexpectedly, second-shell and third-shell waters in this structure tend to occur in surface grooves, bridging first-shell waters that hydrogen bond to protein atoms located 9–11 Å apart. Indeed, ∼90% of second-shell waters and ∼30% of third-shell waters are thus located, generally between α-helices or between an α-helix and a neighbouring β-strand.

3.2. Comparison of the three structures

The two structures I and II reported here crystallized in the same form using MPD as the precipitant, but at different pH values (sodium acetate pH 4.7, MES pH 6.1), and each has two monomers in the asymmetric unit. The previously reported 1.1 Å resolution structure of the apo Ffh protein (PDB code 1ls1; Ramirez *et al.*, 2002) was crystallized under different solution conditions (TAPS pH 9.0, PEG MME 550) and contains one monomer in the asymmetric unit. The overall structure of each monomer is very similar; the r.m.s. difference between structures I*A* and II*A* is 0.408 Å (293 C^{α} atoms), that between structures I*B* and II*B* is 0.425 Å (295 C^{α} atoms) and that between II*A* and I*B* is 0.628 Å (293 C^{α} atoms). Interestingly, the largest difference between structures I and II occurs in the conserved DARGG loop between motif IV and helix α4, which is seen in these structures in two different configurations in the two crystal forms (Fig. 1*c*). One conformation is characteristic of the apo Ffh (Freymann *et al.*, 1999) and is similar to that in 1ls1; the other is similar to the conformation adopted during the GTP-dependent assembly of its heterodimeric complex with FtsY (Focia, Shepotinovskaya *et al.*, 2004). Although the loop is well ordered in one of two configurations in the *A* monomer of each crystal structure (I and II), in the *B* monomers of each the electron density is more poorly defined, suggesting that the loop adopts alternate conformations throughout the crystal that cannot be resolved. Because the DARGG loop appears to be relatively unconstrained in the monomeric Ffh, it is unlikely that it functions to directly couple the N and G domains as proposed previously (Freymann *et al.*, 1997).

Each of the crystal forms (I, II, 1ls1) shares two common crystal-packing motifs. The crystal-packing interactions of monomers I*A* and II*A* are very similar to that of the one monomer in the asymmetric unit of 1ls1. The crystal-packing context of the *B* monomers is distinct, sharing only one of the three crystal-packing interactions with the others (Fig. 5). Most striking is that each monomer in each structure engages in a head-to-tail association across a crystallographic twofold axis that buries approximately 1238 $Å^2$ of the 'back' surface (opposite the active-site pocket) of the protein, an interaction

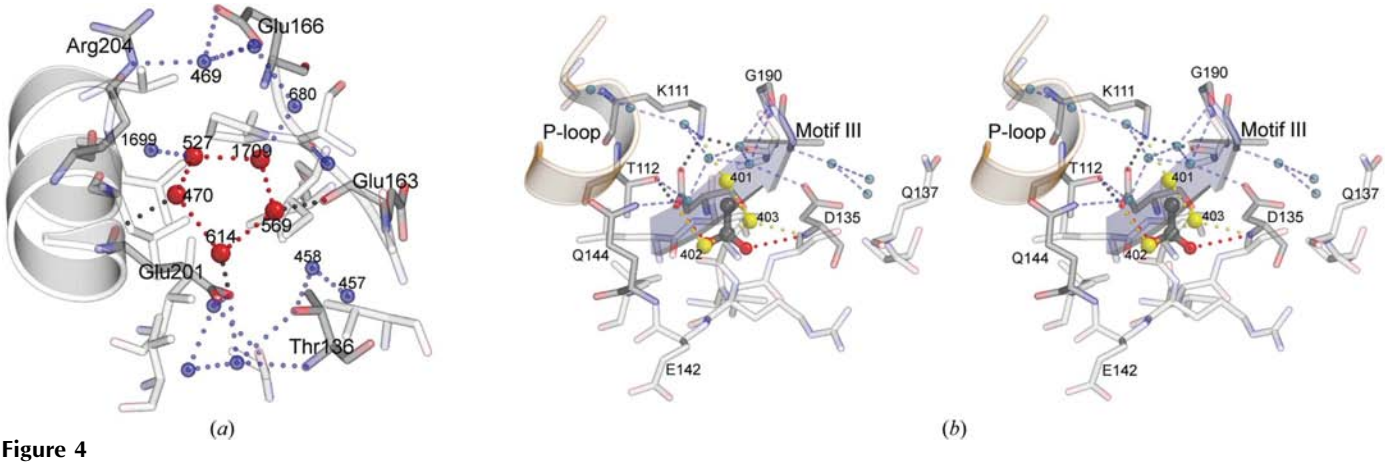

Figure 4
Details of the water structure. (*a*) The water structure at the solvent-exposed edge of a pocket of hydrophobic residues is shown. Water molecules shown in red (470, 527, 569, 614 and 1709) participate in a water pentamer. The pentamer fills the groove along the surface of the protein that is formed by the side chains of residues extending into the hydrophobic pocket below the surface. Hydrogen bonds made among the members of the pentamer are shown as red dotted lines and hydrogen bonds from pentamer members to protein residues are shown as black dotted lines. The surrounding water structure and the hydrogen bonds these waters make are shown in blue. Alternate conformations of protein and water molecules in this region are omitted for clarity. (*b*) Stereo image of the acetate molecule bound within the active site in structure I, the waters in structure II that occupy the space of the acetate (superimposed, yellow) and the residues which hydrogen bond to them. Hydrogen bonds from the acetate O atoms are shown as red dotted lines and the waters of structure II are shown in yellow, as are the dotted lines representing the hydrogen bonds they make with each other and with protein atoms.

mediated by a divalent cation (Fig. 5*a*). The additional crystal-packing interaction of monomer *A* (but not monomer *B*) that is similar to that of 1ls1 arises from a 'head'-to-'head' twofold screw interaction common to both (Fig. 5*b*). The context of particular solvent conditions and shared crystal-packing interactions clearly limit the utility of these structures for comparative analysis of the interaction between protein and solvating water because they themselves affect its stability and organization (Mattos, 2002; Zhang & Matthews, 1994). However, the wide range of pH values sampled in the three crystal structures (I, II and 1ls1) and the slightly different packing environments for each monomer (*A*, *B* and 1ls1) somewhat offset these constraints. In our discussion of the conservation of water structure at the solvation interface, we focus on three regions relatively distant from crystal-packing contexts (a, b and c in Fig. 5*b*) in order to discern protein-directed, rather than crystal-directed, arrangements.

3.3. Solute binding within the active site

An acetate molecule (a crystallization buffer component) is bound completely buried within a deep active-site pocket in both monomers of the asymmetric unit of structure I (Fig. 4*b*). The electron density is very well defined and the greater electron density of the O atoms allows its orientation to be determined. The carboxylate O atoms hydrogen bond to the main-chain amide of Asp135 and to the side-chain carboxylate of Asp187; on this basis, we infer that the bound acetic acid species is likely to be protonated. The methyl group of the acetic acid is accommodated within a 'cage' of three well ordered waters (and the N^{ε} group of Lys111). Structure II was determined in order to eliminate the acetate from this site, substituting MES pH 6.1, 20 m*M* $CaCl_2$ for sodium acetate pH 4.7, 0.1 *M* $MnCl_2$ (the change in divalent cation was a byproduct of the optimization screen). The presence or absence of acetate has little effect on the protein structure, but comparison of the solvent structures of I and II suggests that its presence stabilizes the water structure somewhat. In monomer II*B*, three water molecules occupy the space occupied by acetate in structure I, with waters overlapping almost exactly with the positions of the carboxylate O atoms (Fig. 4*b*). However, the surrounding 'cage', which extends towards the motif I P-loop, is much less well defined, suggesting an organizing effect by the acetate methyl group. In monomer II*A*, however, the water structure is strikingly disordered within the pocket and bound water is only indicated by a tube of poorly defined electron density that extends between alternate binding positions. Thus, while each water position defined by atoms of the bound acetate can be assigned, in monomer *A* of the acetate-free structure each is modeled as occupying (at least) two additional positions.

Replacing acetate with MES buffer itself introduces an 'undesirable' consequence, the binding of MES at two well defined sites, including, adjacent to the acetate-binding site, the motif I P-loop. As in structures of other GTPases obtained in the presence of ammonium sulfate (Freymann *et al.*, 1999; Gariani *et al.*, 2005; Worthylake *et al.*, 2000), the sulfate group of MES is well defined in the electron-density map and interacts with the phosphate-binding site of the P-loop. In monomer II*A* the full MES molecule can be discerned and it

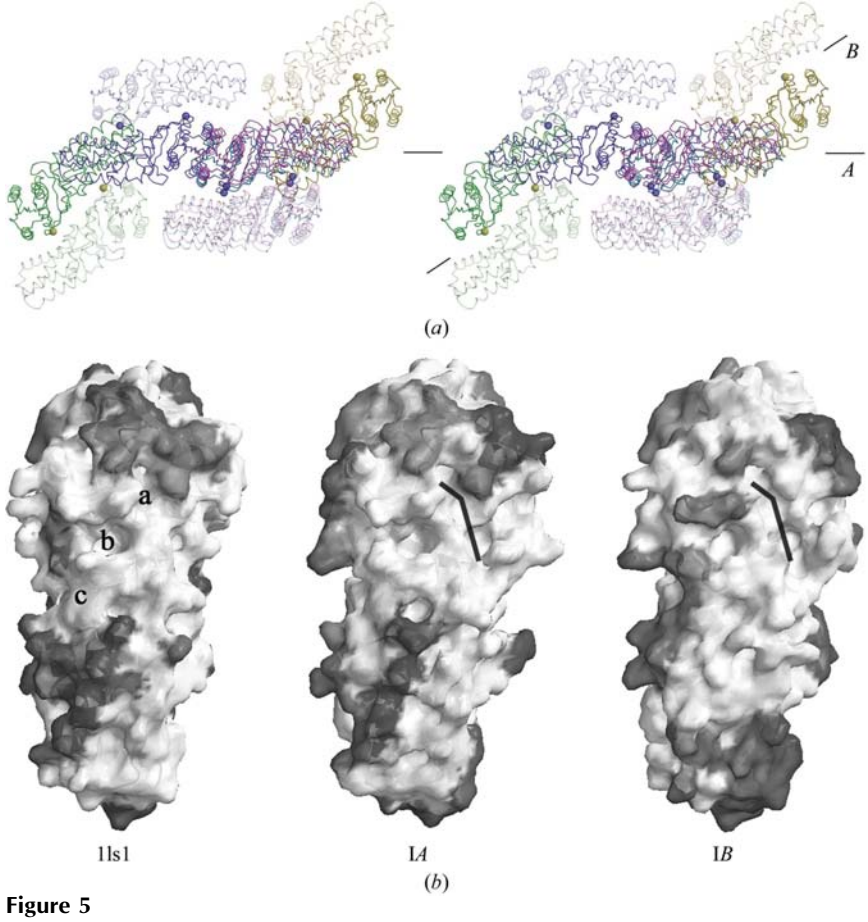

Figure 5
Crystal-packing contexts. (*a*) Stereo image of the crystal-packing arrangement in both structures. There are two monomers in the asymmetric unit. One monomer (*A*) forms a head-to-head twofold screw packed layer, running horizontally in this image. The other (*B*) packs between *A* layers. Both *A* and *B* monomers have a common twofold packing interaction along the 'back' side of the monomer; the colored balls indicate twofold-related ion-binding sites (Mn^{2+}, Ca^{2+}) that mediate the 'back' side packing. (*b*) A surface representation of one face of the NG monomer (here, structure I) oriented so that the nucleotide-binding site is towards the viewer (indicated by stick representation of GDP) and shaded to indicate regions involved in crystal-packing contacts (darker). The crystal contacts for 1ls1 and I*A* are similar (*e.g.* the ridge at the lower left and the packing surface at the top), but distinct, in this orientation, for I*B*. However, for each monomer the packing interactions of the obscured face (180° rotation around the vertical axis) are almost identical (not shown) owing to a common head-to-tail twofold packing relationship that is preserved across each monomer. The approximate locations of the three water networks (*a*, *b* and *c*) discussed in the text are indicated.

extends across the active site (see Fig. 6*a*). In monomer II*B* the remainder of the MES molecule is poorly defined and cannot be fully modeled. Additionally, while the precipitant MPD can be identified bound at sites external to the active site in both structures I and II, in structure II*A* density that we interpret as a partially disordered MPD molecule occupies the lateral edge of the active site between the IBD and the closing loop (Fig. 6*a*).

A number of studies have shown that active sites (or binding sites) exhibit distinct behavior with respect to water and solute binding (Mattos *et al.*, 2006; Ringe & Mattos, 1999). Thus, for example, crystallographic studies of porcine elastase in organic solvents have demonstrated that the solvent-binding sites are predominantly localized to the enzymatic active site (Mattos *et al.*, 2006). The consequence of promiscuous solute (or buffer) binding here is that it is impossible to acquire a 'pristine' image of water structure around the Ffh NG domain, particularly within the active site. As the active site of the GTPase can be considered to be 'designed' for substrate interaction and therefore the partially disordered water structure configured such that exchange or displacement by small solute molecules readily occurs, this behavior should not be surprising (Mattos, 2002; Mattos *et al.*, 2006; Mattos & Ringe, 1996; Ringe & Mattos, 1999). It can perhaps be contrasted to water that contributes directly to stabilization of the underlying protein structure (Petukhov *et al.*, 1999) or water that maintains hydration of hydrophobic surfaces that contribute to conformational change; both appear to exhibit greater conservation when compared between structures and more extensive order (below).

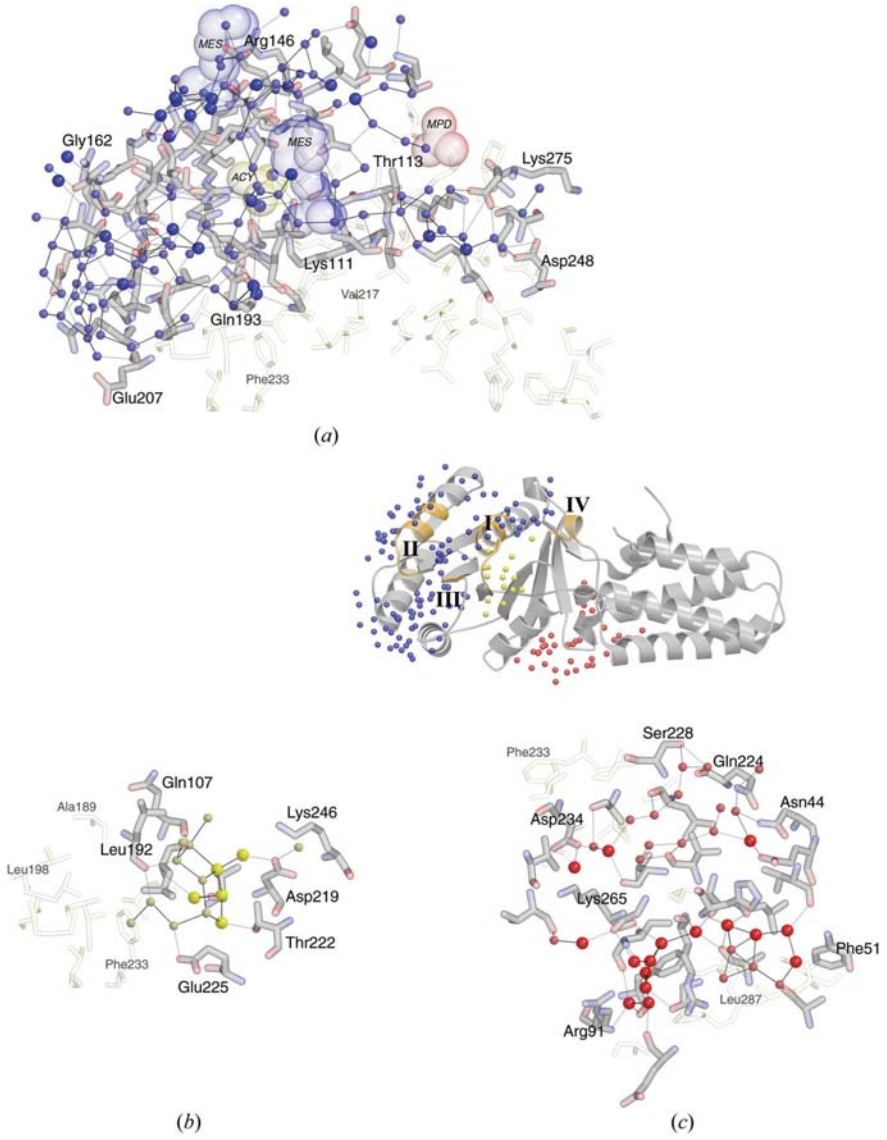

Figure 6
Three functionally important water hydrogen-bonding networks. The central ribbon diagram positions the three panels relative to each other. (*a*) Network *a*, in the GTPase active site, which includes water molecules that hydrogen bond to residues of the conserved GTPase motifs I–IV (Freymann *et al.*, 1997). The water molecules are in are blue; conserved waters (present in all structures) are shown with a larger diameter, while those that contribute to the cluster but are not conserved are shown as smaller faded spheres. The positions of three solute molecules in structure II*A* (two MES, in blue, and one partially disordered MPD molecule, in red) and the position of the acetate bound in structure I*A* (in yellow) are indicated as CPK 'ghosts': these interfere with analysis of the binding-pocket water structure. (*b*) Network *b*, at the conformationally sensitive pocket central to the 'latch' interaction of the targeting heterodimer (Focia, Shepotinovskaya *et al.*, 2004). Waters that hydrogen bond to residues in motifs I and III are in orange. Again, conserved waters are larger. Note the pentagon cluster that is present between motif I (at the left) and the DAGQ motif (at the right), stabilized against a hydrophobic pocket between them. (*c*) Network *c*, at the Ffh N/G interface, includes waters that contribute to conformational flexibility between the N and G subdomains (Ramirez *et al.*, 2002). Note the spine of highly conserved waters that traverses the N/G domain interface. Each of the three panels indicates, in faded yellow, the pocket of generally hydrophobic residues that delimits the region of the cluster.

3.4. Alternate conformations couple to water substates

The numerous residues adopting alternate conformations were generally clearly indicated by residual features in the difference electron density prior to modeling of the second conformation (Fig. 1*a*). Many exhibited only a main-chain shift without an accompanying side-chain rotamer rearrangement; most of these (51 of 261 residues with alternate conformations in structure II) occur adjacent to a residue for which there is clear evidence for an alternate side-chain rotamer in the electron-density map, consistent with a coupling propagated along the main chain to accommodate the side-chain rearrangement. Several peptides exhibit the

characteristics of the recently identified main-chain 'backbone rub' (Davis *et al.*, 2006); this torsional coupling allows movement of the C^{β} atom between two positions with minimal effect on the positions of the residues adjacent to it (see, for example, Arg127 in Fig. 1*a*). Several peptide flips are also well resolved. One at motif I, between Gln107 and Gly108, may be functionally significant; in structure II*A* the peptide adopts two clearly defined conformations, each linked to the presence of a nearby water (Fig. 1*b*). This rearrangement abrogates the phosphate interactions that can be contributed by the P-loop for nucleotide recognition there and similar conformational changes have been suggested to regulate nucleotide exchange in other G proteins (Cherfils & Chardin, 1999). Hydrogen-bond and van der Waals mediated coupling between protein backbone and side-chain conformational alternates is suggested by several correlated shifts: for example, in structure II*A* the distal end of helices αN2 and αN4 shift in cooordination with each other such that residues Arg21–Leu27 (αN2) and residues Lys62–Leu68 (αN4) adopt distinct alternate configurations that maintain the packing interactions between them. A second well defined example of residues adopting interacting alternate conformations occurs between the side chain of Arg127 and the main chain of Lys96 (Fig. 2*a*). Finally, a remarkable number of side-chain alternates are not surface-exposed but are buried within the hydrophobic core of the protein. Several of these have been noted previously in 1ls1, in particular nine leucine side chains, of which five exhibit clearly defined rotamer flips (Ramirez *et al.*, 2002). Of the alternate conformations observed in structures I and II here, 17 of 90 and 41 of 261 residues, respectively, are buried. That almost 20% of the conformational flexibility defined by these structures is exhibited by residues packed within the core of the protein is noteworthy and may reflect the conformational freedom necessary for the functional cycle of the SRP GTPase.

There is also extensive coupling between the protein conformation and the surrounding waters, as adoption of different configurations is reflected in the positions of the neighbouring solvent molecules and the overall solvent structure (Fig. 2). Elongation of the water density was common in the electron-density maps and was taken to indicate alternate water positions if it accommodated two waters too close to hydrogen bond to one another; examples include Wat1163 (Fig. 2*b*) and waters 455, 509, 567, 688, 1595 and 1646 (Fig. 2*a*). Of the latter, half interact directly with backbone carbonyl O atoms and reflect the underlying conformations of the main chain. Not surprisingly, in structure II 90%, in structure I 100% and in 1ls1 86% of waters in alternate conformations are located adjacent to protein residues that themselves adopt alternate conformations. Conversely (in structure II), ~20% of protein residues that adopt alternate configurations, including those limited to backbone shifts without accompanying side-chain rotamer rearrangment, can be associated with waters occupying alternate positions. The shifts in water positions are generally consistent with the maintenance of hydrogen-bonding interactions to polar protein atoms (Fig. 2*a*); similar shifting of water to maintain hydrogen bonding to protein atoms in alternate configurations has been described in different structures at ultrahigh resolution (Yamano *et al.*, 1997). However, alternate positions for a water do not necessarily disrupt other possible hydrogen-bonding interactions (*i.e.* between water molecules), which can therefore adjust; this is exemplified by water W1208, which is within hydrogen-bonding distance of both conformations of water W509 (Fig. 2*a*). The change in position of the first-shell water to maintain a hydrogen-bonding distance to its protein-atom bonding partner does not necessarily shift the second-shell waters it binds to. Thus, water alternate conformations can co-localize with protein alternate conformations such that the extended water structure is able to absorb local changes with minimal reorganization. Alternatively, some water positions remain occupied despite protein conformational changes near them, but the role of the waters occupying these positions changes. Wat1646, for example, is a first-shell water position only when Arg128 is in conformation *B*, but is fully occupied (see Fig. 2*a*).

3.5. Conservation of water positions

That elements of the configuration of local water hydrogen-bonding structure are maintained with respect to alternate protein conformations prompts the question of whether they are similarly maintained between structures determined under different crystallization conditions. We find that ~17% to ~30% of the water positions resolved in each ultrahigh-resolution structures can be readily superimposed within a 1.0 Å radius (Fig. 3*b*). 78 waters are conserved across each of the five distinct monomer species (1ls1, I*A* and I*B*, II*A* and II*B*); an additional 16 water positions are conserved but are located at the twofold packing interaction that is common to all of them (Fig. 5*a*). The 78 non-contact water positions represent ~28% of the water positions determined in 1ls1, the 1.1 Å structure with the fewest waters of the structures compared and ~28% of the water structure of monomers I*A* and I*B*. Almost all are first-shell waters; six are second hydration-shell waters and none are third shell (Fig. 3*b*). Of the 72 conserved first-shell water positions, approximately 47% are buried and most of the conserved water positions appear to play a structural role (Fig. 3*c*), mediating interactions between secondary-structural elements or positioned by main-chain hydrogen-bond interactions at the end of an α-helix or β-strand or at a loop region (Pujadas & Palau, 2001). While the overall distribution is fairly disperse (Fig. 3*b*), the promiscuous interactions with solute molecules across the nucleotide-binding site implies that observation of 'conserved' water positions is necessarily limited to areas that are static with respect to solvent structure, although not necessarily with respect to function. Thus, while some conservation simply reflects hydrogen bonding to conformationally restricted polar groups, much of the distribution localizes to important packing interfaces, including the N/G domain interface (Ramirez *et al.*, 2002) and the interaction between the dynamic helices α2 (switch 2 in the GTPase superfamily; Vetter & Wittinghofer, 2001) and helix α3 (which tracks the twist of the underlying

β-sheet; Freymann *et al.*, 1999). Each plays a role in the concerted conformational change that accompanies assembly of the FtsY heterodimer (Focia, Shepotinovskaya *et al.*, 2004; Padmanabhan & Freymann, 2001). Interestingly, almost half of the conserved positions can be clustered together as networks assembled in regions that are likely to be functionally important (below).

We can also ask whether a subset of the solvent arrangement is conserved in structures determined at lower resolution or in different binding states. Several structures of the apo, GDP-bound and GMPPNP-bound *T. aquaticus* Ffh NG have been determined at ~2.0 Å resolution (Freymann *et al.*, 1997, 1999; Padmanabhan & Freymann, 2001). Of the set of 78 conserved waters (72 first-shell and six second-shell waters), 28 waters, one of which is a second-shell water, are present in all compared structures (Fig. 3*b*). Of these, approximately two-thirds are buried and clearly play a structural role. Water is often organized across solvent-exposed hydrophobic surfaces (Teeter, 1984) and this behavior can be exploited in functionally important ways. For example, extensive water organization specific to a halophilic enzyme which has the role of hydrating the protein in the presence of high salt has been shown to follow from increased exposure of hydrophobic residues to the surface relative to the non-halophilic case (Britton *et al.*, 2006). There are several surface-exposed hydrophobic clusters in the Ffh NG structure, one of which is of particular interest because it contributes to the receptor-interaction surface between helices α2 and α3 and the P-loop. A pentameric water-ring structure is readily discernable (Fig. 4*a*) and is present in both the ultrahigh-resolution structures and in structures determined at lower resolution. The one second-shell water conserved, W447, is part of a water pentamer ring with two other completely conserved waters, W421 and W422. Elsewhere, three other first-shell water positions present in all structures (W464, W465 and W470) also contribute to pentamer structures.

3.6. Conserved hydrogen-bond networks

The interactions between water molecules at the protein surface generate what can be considered 'networks' (Mattos & Ringe, 2001; Nakasako, 1999; Sanschagrin & Kuhn, 1998); that is, hydrogen-bonding interactions between water molecules and the underlying protein structure that extend across the protein surface. Since these 'networks' mediate interactions between secondary-structural elements and interactions between the protein and approaching solute and because they must rearrange during any significant conformational change, it is of interest to establish for the Ffh GTPases to what extent structured water networks exist on the protein surface and to what extent they incorporate the 'conserved' waters described above. However, owing to the extent of disorder exhibited by these structures and the tendency to bind solute molecules, as described above, unambiguous definition of conserved networks in some regions of these structures is difficult. That is, while extensive hydrogen-bonding interactions can be identified readily in any one structure, defining conservation of hydrogen-bonding patterns is problematic. Here, we take the water distribution of II*A* as illustrative of the network structure, focusing on three regions that are associated with functionally important protein structural motifs, not structured by crystal-packing interactions, and relatively well defined.

Generally, conserved water positions can be identified as contributing to much of a first-shell 'core' of each network. Of the 94 water positions 'conserved' between the 1.1 Å structures, 49 can be assigned to one of three hydrogen-bonding networks *a*, *b* and *c* discussed below. Each network can be associated with a pocket on the protein surface and each is generally delimited by surface-exposed hydrophobic patches, linked *via* polar main-chain atoms and polar side-chain atoms that define their boundaries and extends into the pockets to link polar main-chain atoms to the surface of the protein (Fig. 6). In the crystal, of course, the hydrogen-bonding interactions can extend undisrupted across crystal-packing contacts (Nakasako, 1999); however, here we map only the local relationship of these clusters at the functional groups of the protein surface. The three clusters were evaluated in the context of the 'reference' structure II*A* and the descriptions refer to that structure. Subsequently, each cluster was compared with the water structure in the other monomers to evaluate its conservation across structural contexts. The 'core' of each network can be shown to be conserved across the five monomers of the three 1.1 Å apo structures.

The first network, which we term network *a*, is quite extensive, spanning ~33 Å to occupy the nucleotide-binding site. However, as this region also mediates the buffer solute interactions (*i.e.* acetate, MES and MPD), the water structure itself is relatively poorly conserved. In structure II*A*, a continuous hydrogen-bonding network comprising 133 waters can be identified (Fig. 6*a*). The cluster is delimited by residues from each of the four conserved GTPase motifs I–IV; of the 54 residues that contribute polar atoms to the hydrogen-bonding network, 16 are evolutionarily conserved in SRP GTPases. An MES molecule bound at the motif I P-loop, near the center of the cluster, and an MPD molecule bound at one end near motif IV necessarily introduce discontinuities (Fig. 6*a*) and because the network is sensitive to solute interactions, its representation in pairwise comparisons is limited: in structure I monomer *A* ~62% of the 133 network *a* water positions identified in structure IIA are present, in 1ls1 ~52% are present and in the *B* monomers of structures I and II ~47% and ~32%, respectively, are present. Of the 133 waters in the network in II*A*, only 26 are present in all monomers of the structures at 1.1 Å; nevertheless, of these eight are conserved in all apo Ffh NG structures and two are completely conserved among all binding states of *T. aquaticus* Ffh NG.

The second network, which we which term network *b*, occupies a hydrophobic pocket that contributes to the 'latch' interface of the Ffh NG domain and links conserved GTPase motifs I and III (Fig. 6*b*). It comprises 15 water molecules in structure II*A*, of which five are conserved in all monomers of the 1.1 Å structures and four are conserved in all monomeric binding states. These conserved waters contribute to three

continguous water-ring structures in tetrameric, pentameric and hexagonal (with the carbonyl of Leu192) hydrogen-bonding arrangements that are positioned above a hydrophobic surface (Fig. 6*b*). The network remains well represented across the different monomers, as in structure I*A*, 1ls1 and structure II*B* ~76% of the water positions in network *b* are occupied and in structure I*B* ~65% are occupied. Four of the seven residues that donate polar atoms to the hydrogen-bonding network are evolutionarily conserved in SRP GTPases; interestingly, the peptide flip at Gln107 (Fig. 1*b*), which brings the carbonyl O atom of this residue from an interaction with network *a* to an interaction with network *b*, can perhaps be considered to couple them. The significance of the cluster is that it is located between helices α2, α3 and the DAGQ motif (Fig. 6*b*; Focia, Shepotinovskaya *et al.*, 2004), which together undergo a concerted conformational change during assembly of the heterodimeric complex with the receptor FtsY: these waters, then, must be released to allow the reconfiguration of the motif at the heterodimer interface.

Finally, network *c* spans the interface of the N and G domains (Fig. 6*c*). It includes three waters (W426, W429, W440) identified previously as contributing to mobility of the interface between the N and G domains of Ffh (Ramirez *et al.*, 2002) and of the 42 water molecules that can be identified in the network, ~23% are buried at the interface. 16 are conserved in all monomers of the 1.1 Å structures and the cluster identified in II*A* is similarly present in each of the others: of the water positions that comprise network *c*, ~64% are present in structure I*A*, ~85% in structure II*B* and ~68% in 1ls1 and in structure I*B*. 28 protein residues contribute polar atoms to the hydrogen-bonding network, of which eight are evolutionarily conserved in SRP GTPases, but strikingly, network *c* has distinct boundaries drawn by two large hydrophobic surfaces, the first comprised of 12 hydrophobic side chains contributed primarily by αN4 and the C-terminal helix that packs against it and the other formed by six hydrophobic side chains contributed by the G domain. This interface rearranges during assembly of the Ffh–FtsY heterodimer (Focia, Shepotinovskaya *et al.*, 2004), such that the latter surface, the underlying β-sheet core of the G-domain, slides over the former, the N-domain and C-terminal helix of the GTPase (Gawronski-Salerno *et al.*, 2006).

Although none of the networks can be mapped completely in a simple way from one structure to the other, the positions of conserved waters that contribute to each, particularly for networks *b* and *c*, identify a localized 'core' or 'spine' of water positions adjacent to the protein surface that is maintained in each structure (Fig. 6). It has been shown that nucleotide binding does not cause large conformational changes in the structure of the Ffh NG domain (Freymann *et al.*, 1999; Padmanabhan & Freymann, 2001). That is, aside from the binding pocket itself (*i.e.* cluster *a*), the protein structure and therefore the water-cluster interactions of clusters *b* and *c* are likely to be largely maintained prior to interaction with the receptor. Further, although cluster *a* comprises much of the active site, the conserved elements of the cluster are adjacent to, but not within, the binding site (see Fig. 3*b*). As the nucleotide (or its proxy; here, MES) binds, the center of network *a* must be largely released but its periphery retained. These observations, though illustrative, highlight that each of these networks is associated with functionally significant conformational changes that occur during the initial stages of the assembly of the targeting complex of the SRP GTPases Ffh and FtsY. Upon assembly of the Ffh–FtsY receptor complex, extensive rearrangement of the N/G interface and the α2/α3/DAGQ packing regions that contribute to the 'latch' interface occurs and those regions are largely desolvated (Egea *et al.*, 2004; Focia, Shepotinovskaya *et al.*, 2004). The core positions of clusters *b* and *c* are conserved and both are delimited by adjacent hydrophobic surfaces that may play a role in facilitating water release and protein rebinding at the interface (Royer *et al.*, 1996).

4. Conclusions

The ultrahigh-resolution structures of Ffh NG illustrate a number of phenomena that must be occurring in all proteins in their interaction with water; they also allow us to begin to focus on particular regions of water structure that are likely to be functionally important. Firstly, we establish local adjustment (or 'sloshing') of water molecules in conjunction with movement of protein residues as they adopt alternate configurations. This behavior, which can be readily observed in structures determined at ultrahigh resolution (Yamano *et al.*, 1997), maintains hydrogen-bonding interactions with the protein and preserves much of the overall water structure. Secondly, we observe the conservation of a subset of water positions between the different high-resolution structures and between the structures at lower resolution of the apo and nucleotide-bound protein. Finally, three functionally important pockets contain networks of hydrogen-bonded water molecules that are maintained in a large part between different structures determined under different solution and crystal-packing conditions.

Previous studies suggest that different water positions observed in multiple structures of the same protein together define the complete solvent structure, although all positions are not always occupied simultaneously (Mattos & Ringe, 2001); that is, different structures sample different discrete populations of water surrounding the protein structure. This aggregate arrangement is thought to be conserved, with local changes to the arrangement of water positions resulting from minor protein conformational changes. Further, the observation of water at particular positions in this arrangement is primarily dictated by the resolution and quality of the data, the relative level of order of nearby protein atoms and the distance between the water position and ordered protein atoms. Here, in structure II, we identify ~10% of the water in alternate positions; these can perhaps be considered to be sampling the aggregate distribution of water structure. Studies of conserved water positions in multiple structures of a protein have also suggested that only relatively few water positions are conserved, irrespective of crystal-packing environment and crystallization conditions (Loris *et al.*, 1999, 1994). We

identify from 17 to 28% of observed waters as conserved between structures determined under different crystallization conditions.

The arrangement of water positions contributing to three networks is to some extent conserved among multiple structures of apo Ffh NG at 1.1 Å resolution which had different crystallization conditions and crystal packing. The clustered water structure can be considered to be anchored by 'spines' of five or six highly conserved water molecules that contribute to the first-shell hydration of the protein (see Fig. 6) and that presumably can adjust to small shifts in protein backbone and side-chain configuration (see Fig. 2). We find that these clusters are similarly present in each of the structures compared. However, while over half of the waters identified as conserved between structures contribute to the three clusters we discuss (the others contribute to other regions that are not considered), with respect to the water networks themselves as few as 19% and maximally 38% are present in similar positions in the different structures. This is consistent with and may arise from longer distance 'sloshing' of the water structure (and so rearrangement, not addressed here). Certainly contributing to the difficulty in analyzing this in these structures is the intrinsic disorder of protein and solvent-atom configurations described above.

Also, importantly, the promiscuity of binding, particularly at the nucleotide-binding site, such that solute interactions interfere with intrinsic 'protein-directed' water structure, interferes with any attempt to image a 'pristine' water structure. Although the conservation of the water structure with and without an acetate molecule in the Ffh NG active site suggests that small molecules that continue this hydrogen-bonding pattern can be readily bound without major perturbation of the local water structure, the behavior of MES binding suggests that a simple analysis is necessarily incomplete. Ligand atoms must displace bound water molecules from the hydrogen-bonding network as they approach the binding site before energetically favourable rearrangements that make the ligand binding more favourable than water binding can occur. These networks are presumably exploited by nature to guide the binding of ligands or proteins at a protein's surface. The presence of the highly conserved and extensive hydrogen-bonding networks around the protein's surface may also be exploited by nature to communicate large conformational changes of the protein (Eisenmesser *et al.*, 2005; Volkman *et al.*, 2001). However, whether or not the distinct configurations of the water networks we identify here serve to couple functionally important alternate configurations of residues of Ffh that are distant in space remains to be established.

This work can also be taken as a caution. Although further studies of the water structure of multiple binding states of the same protein at ultrahigh resolution will allow investigation into the role of water structure in binding and conformational change, such studies must be carefully considered in the context of their crystal-packing and solution environments. The former significantly constrict that part of the structure that can be considered to 'interact' with solvent; the latter introduce solutes specific to the experiment that may or may not interfere with understanding the overall behavior of binding. Nevertheless, the three ultrahigh-resolution structures of the apo Ffh NG domain begin to allow us to focus on particular regions of the protein and define subsets of solvent interactions that are likely to play a key role in the function of these proteins. They should contribute to the detailed understanding of the mechanism by which binding GTP and interactions during the encounter of the two SRP GTPases at the membrane trigger the conformational changes that assemble the latched state of the SRP protein-targeting complex.

We thank J. S. Coon V, K. Dietrich, P. J. Focia, J. Gawronski-Salerno, I. Kruk and E. McCabe for contributions to the figures and critical reading of the manuscript. R. Ramirez contributed to the development of the scripts used in this work. T. Lu and E. Alam carried out preliminary studies of the water structure in Ffh NG. This work was supported by grant GM58500 from the NIH and by support from the R. H. Lurie Comprehensive Cancer Center to the Structural Biology Facility at Northwestern University. SSRL is supported by the DOE Office of Basic Energy Sciences. The SSRL Biotechnology Program is supported by the NIH, National Center for Research Resources, Biomedical Technology Program and by the DOE Office of Biological and Environmental Research. Portions of this work were performed at the DuPont–Northwestern–Dow Collaborative Access Team (DND-CAT) Synchrotron Research Center, Sector 5 and BioCARS, Sector 14 of the Advanced Photon Source (APS) at Argonne National Laboratory. Use of the APS was supported by the US Department of Energy, Basic Energy Sciences, Office of Science under Contract No. W-31-109-Eng-38. DND-CAT was supported by DuPont, Dow, the state of Illinois and the NSF. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources under grant No. RR07707. DMF and UDR measured the diffraction data; UDR was responsible for development of computational tools and crystallographic refinement and drafted the manuscript; DMF revised the manuscript.

References

Beamer, L. J., Li, X., Bottoms, C. A. & Hannink, M. (2005). *Acta Cryst.* D**61**, 1335–1342.

Bottoms, C. A., White, T. A. & Tanner, J. J. (2006). *Proteins*, **64**, 404–421.

Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., Ferrer, J., Pire, C., Esclapez, J. & Rice, D. W. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4846–4851.

Cherfils, J. & Chardin, P. (1999). *Trends Biochem. Sci.* **24**, 306–311.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Davis, I. W., Arendall, W. B. III, Richardson, D. C. & Richardson, J. S. (2006). *Structure*, **14**, 265–274.

DePristo, M. A., de Bakker, P. I. & Blundell, T. L. (2004). *Structure*, **12**, 831–838.

Egea, P. F., Shan, S. O., Napetschnig, J., Savage, D. F., Walter, P. & Stroud, R. M. (2004). *Nature (London)*, **427**, 215–221.

Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E. & Kern, D. (2005). *Nature (London)*, **438**, 117–121.
Focia, P. J., Alam, H., Lu, T., Ramirez, U. D. & Freymann, D. M. (2004). *Proteins*, **54**, 222–230.
Focia, P. J., Shepotinovskaya, I. V., Seidler, J. A. & Freymann, D. M. (2004). *Science*, **303**, 373–377.
Freymann, D. M., Keenan, R. J., Stroud, R. M. & Walter, P. (1997). *Nature (London)*, **385**, 361–364.
Freymann, D. M., Keenan, R. J., Stroud, R. M. & Walter, P. (1999). *Nature Struct. Biol.* **6**, 793–801.
Gariani, T., Samuelsson, T. & Sauer-Eriksson, A. E. (2006). *J. Struct. Biol.* **153**, 85–96.
Gawronski-Salerno, J., Coon, J. S. V, Focia, P. J. & Freymann, D. M. (2006). In the press.
Halic, M., Becker, T., Pool, M. R., Spahn, C. M. T., Grassucci, R. A., Frank, J. & Beckmann, R. (2004). *Nature (London)*, **472**, 808–814.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Keenan, R. J., Freymann, D. M., Stroud, R. M. & Walter, P. (2001). *Annu. Rev. Biochem.* **70**, 755–775.
Kleywegt, G. J. & Jones, T. A. (1997). *Methods Enzymol.* **277**, 525–545.
Kleywegt, G. J., Zou, J.-Y., Kjeldgaard, M. & Jones, T. A. (2001). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
Knapp, J. E., Pahl, R., Srajer, V. & Royer, W. E. Jr (2006). *Proc. Natl Acad. Sci. USA*, **103**, 7649–7654.
Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D. & Thorne, R. E. (2002). *Acta Cryst.* D**58**, 459–471.
Lamzin, V. S., Perrakis, A. & Wilson, K. S. (2001). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, pp. 720–722. Dordrecht: Kluwer Academic Publishers.
Lo Conte, L., Chothia, C. & Janin, J. (1999). *J. Mol. Biol.* **285**, 2177–2198.
Loris, R., Langhorst, U., De Vos, S., Decanniere, K., Bouckaert, J., Maes, D., Transue, T. R. & Steyaert, J. (1999). *Proteins*, **36**, 117–134.
Loris, R., Stas, P. P. & Wyns, L. (1994). *J. Biol. Chem.* **269**, 26722–26733.
Lovell, S. C., Davis, I. W., Arendall, W. B. III, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins*, **50**, 437–450.
Mattos, C. (2002). *Trends Biochem. Sci.* **27**, 203–208.
Mattos, C., Bellamacina, C. R., Peisach, E., Pereira, A., Vitkup, D., Petsko, G. A. & Ringe, D. (2006). *J. Mol. Biol.* **357**, 1471–1482.
Mattos, C. & Ringe, D. (1996). *Nature Biotechnol.* **14**, 595–599.
Mattos, C. & Ringe, D. (2001). *International Tables for Crystallography*, edited by M. G. Rossmann & E. Arnold, pp. 623–640. Dordrecht: Kluwer Academic Publishers.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.
Montoya, G., Kaat, K., Moll, R., Schäfer, G. & Sinning, I. (2000). *Structure*, **8**, 515–525.
Montoya, G., Svensson, C., Luirink, J. & Sinning, I. (1997). *Nature (London)*, **385**, 365–369.
Morris, A. L., MacArthur, M. W., Hutchinson, E. G. & Thornton, J. M. (1992). *Proteins*, **12**, 345–364.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Murshudov, G. N., Vagin, A. A., Lebedev, A., Wilson, K. S. & Dodson, E. J. (1999). *Acta Cryst.* D**55**, 247–255.
Nakasako, M. (1999). *J. Mol. Biol.* **289**, 547–564.
Nicholls, P. (2000). *Cell. Mol. Life Sci.* **57**, 987–992.
Otwinowski, Z. (1993). *Proceedings of the CCP4 Study Weekend. Data Collection and Processing*, edited by L. Sawyer, N. W. Isaacs & S. Bailey, pp. 55–62. Warrington: Daresbury Laboratory.
Padmanabhan, S. & Freymann, D. M. (2001). *Structure*, **9**, 859–867.
Pardanani, A., Gambacurta, A., Ascoli, F. & Royer, W. E. Jr (1998). *J. Mol. Biol.* **284**, 729–739.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Petukhov, M., Cregut, D., Soares, C. M. & Serrano, L. (1999). *Protein Sci.* **8**, 1982–1989.
Pujadas, G. & Palau, J. (2001). *Protein Sci.* **10**, 1645–1657.
Ramirez, U. D. (2005). Thesis. Neuroscience Institute Graduate Program, Northwestern University, Chicago, USA.
Ramirez, U. D., Minasov, G., Focia, P. J., Stroud, R. M., Walter, P., Kuhn, P. & Freymann, D. M. (2002). *J. Mol. Biol.* **320**, 783–799.
Ringe, D. & Mattos, C. (1999). *Med. Res. Rev.* **19**, 321–331.
Royer, W. E. Jr, Pardanani, A., Gibson, Q. H., Peterson, E. S. & Friedman, J. M. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14526–14531.
Sanschagrin, P. C. & Kuhn, L. A. (1998). *Protein Sci.* **7**, 2054–2064.
Shepotinovskaya, I. V., Focia, P. J. & Freymann, D. M. (2003). *Acta Cryst.* D**59**, 1834–1837.
Simonovic, M. & Volz, K. (2001). *J. Biol. Chem.* **276**, 28637–28640.
Stura, E. A. & Wilson, I. A. (1990). *Methods*, **1**, 38–49.
Stura, E. A. & Wilson, I. A. (1991). *J. Cryst. Growth*, **110**, 270–282.
Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.
Teeter, M. M. (1991). *Annu. Rev. Biophys. Biophys. Chem.* **20**, 577–600.
Teeter, M. M. & Kossiakoff, A. A. (1984). *Neutrons in Biology*, edited by B. P. Schoenborn, pp. 335–348. New York: Plenum.
Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
Terzyan, S., Zhu, G., Li, G. & Zhang, X.-C. (2004). *Acta Cryst.* D**60**, 54–60.
Vetter, I. R. & Wittinghofer, A. (2001). *Science*, **294**, 1299–1304.
Volkman, B. F., Lipson, D., Wemmer, D. E. & Kern, D. (2001). *Science*, **291**, 2429–2433.
Willard, L., Ranjan, A., Zhang, H. Y., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). *Nucleic Acids Res.* **31**, 3316–3319.
Wilson, M. A. & Brunger, A. T. (2000). *J. Mol. Biol.* **301**, 1237–1256.
Word, J. M., Lovell, S. C., LaBean, T. H., Taylor, H. C., Zalis, M. E., Presley, B. K., Richardson, J. S. & Richardson, D. C. (1999). *J. Mol. Biol.* **285**, 1711–1733.
Worthylake, D. K., Rossman, K. L. & Sondek, J. (2000). *Nature (London)*, **408**, 682–688.
Yamano, A., Heo, N. H. & Teeter, M. M. (1997). *J. Biol. Chem.* **272**, 9597–9600.
Zhang, B., Zhang, Y., Wang, Z. & Zheng, Y. (2000). *J. Biol. Chem.* **275**, 25299–25307.
Zhang, X. J. & Matthews, B. W. (1994). *Protein Sci.* **3**, 1031–1039.