Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Hydrophobin HFBII in detail: ultrahigh-resolution structure at 0.75 Å

Johanna Hakanpää,[a]* Markus Linder,[b] Alexander Popov,[c] Andrea Schmidt[c] and Juha Rouvinen[a]*

[a]Department of Chemistry, University of Joensuu, PO Box 111, 80101 Joensuu, Finland, [b]VTT Biotechnology, PO Box 1500, 02044 VTT Espoo, Finland, and [c]EMBL Hamburg/DESY, Notkestrasse 85, 22603 Hamburg, Germany

Correspondence e-mail: johanna.hakanpaa@joensuu.fi, juha.rouvinen@joensuu.fi

Received 17 November 2005
Accepted 9 January 2006

PDB Reference: *T. reesei* hydrophobin HFBII, 2b97, r2b97sf.

Hydrophobins are small proteins secreted by filamentous fungi that have a unique ability to spontaneously form amphiphilic layers. Hydrophobins have only recently been structurally characterized through the first crystal structure determination of a protein of this class, *Trichoderma reesei* hydrophobin HFBII [Hakanpää, Paananen *et al.* (2004), *J. Biol. Chem.* **279**, 534–539]. The resolution of the HFBII structure has now been extended to an ultrahigh resolution of 0.75 Å. The structure was refined conventionally and multipole refinement has been initiated. The ultrahigh-resolution structure is analyzed here in detail and comparison is made to the previous atomic resolution structure of the same protein as well as to other ultrahigh-resolution structures found in the Protein Data Bank.

1. Introduction

To date, about 24 000 protein structures have been determined by X-ray crystallography and deposited in the Brookhaven Protein Data Bank (PDB; Berman *et al.*, 2000). A fraction of these structures, around 600 in number, exceed the diffraction limit of 1.2 Å and may be categorized as atomic resolution structures. In the late 1990s, the first ultrahigh-resolution structures were published. Here, ultrahigh resolution is defined as a resolution of 0.8 Å and beyond. The world record for protein structure resolution is held by crambin (Jelsch *et al.*, 2000), with an amazing 0.54 Å resolution. A total of 14 ultrahigh-resolution structures, as defined here, can be found in the PDB, nine of which are protein structures. In addition, one ultrahigh-resolution protein structure is on hold.

The low number of atomic and ultrahigh-resolution structures (2.5% of the total content of the PDB) is explained by the characteristics of protein crystals. They scatter X-rays only weakly, since the atoms are light; the unit cell is usually very large and the crystals contain a considerable amount of disordered solvent. Crystals that diffract to ultrahigh resolution are therefore usually of proteins that are quite small in size, packed tightly (with a low solvent content) and that belong to low-symmetry space groups with only one molecule in the asymmetric unit. The reason why ultrahigh-resolution structures seem to be a recent innovation in macromolecular crystallography is largely a consequence of the development of novel crystallographic tools. The availability of synchrotron sources, development of detectors, introduction of cryocooling and increased computing speed, among other things, have led to the possibility of collecting data to higher and higher resolution, closing in on the resolution previously only attainable for small molecules. However, good crystals did exist even before the handling of atomic and ultrahigh-resolution data was possible (or feasible).

When the resolution of a given data set reaches and even exceeds the bonding distance of atoms in molecules, the electron-density maps start to show features that are not visible at more moderate resolution. Features such as density peaks caused by H atoms, deformation density caused by bonding or lone-pair electrons of carbonyl O atoms become visible. This means that the structure gives a crystallographer more information, but also more features to model. The question is: is this extra information relevant enough to go through an extensive refinement?

The standard procedure for refining macromolecular X-ray data is to employ a spherical atom model (or IAM, the independent atom model) with isotropic temperature factors. The bond lengths and angles are derived from a library, based on small-molecule data, such as that according to Engh & Huber (1991). H atoms are not included in the structure. All this is performed because on average the amount of unique reflections is too low to include more parameters in the refinement. What we lose, however, is the information of the bond lengths and angles that in reality deviate from standards, the protonation states of certain amino acids and the non-spherical features describing the chemical bonding.

If the resolution of the data permits, anisotropic displacement parameters may be used and H atoms may be included in the refinement with most crystallographic refinement programs. The model may also be allowed to refine freely to avoid forcing the structure to fit the standard parameters, as has recently been performed with rubredoxin (Bönisch *et al.*, 2005). There are two methods of approaching the problem of the non-sphericity of the atoms. One is to place a dummy electron between the bonded atoms to represent the deformation part of the density. This DBE approach (dummy bond electron) is incorporated in the *Phenix* suite (Afonine *et al.*, 2004). The other option is to directly model the valence electrons of the atoms in the structure. This can be performed with the multipolar refinement program *MOPRO* (Jelsch *et al.*, 2000). The multipolar model is a very precise description of the structure, whereas DBE is more approximate but compensates accuracy with speed.

Despite the availability of crystals diffracting to high resolution and appropriate tools with which to process the data, a question still remains regarding the motivation for the collection of atomic and ultrahigh-resolution data. The aim of X-ray structure determination is to learn about the biological function of the target protein. In this respect, the atomic or ultrahigh-resolution data would be expected to bring out something new and biologically relevant to justify the extensive refinement instead of simply cutting the data at 1.2 Å. The visualization of H atoms, identification of atoms by atom type, description of bonding features, accurate distances and modelling the mobility through the anisotropic temperature factors (Harata *et al.*, 1998; Merritt, 1999; Schmidt & Lamzin, 2005) justify pushing the diffraction to atomic resolution and beyond (Dauter *et al.*, 1997; Schmidt & Lamzin, 2002). What makes ultrahigh-resolution structures stand out especially is the possibility of studying the electronic properties and charge distribution of the molecule directly from the experimental data (Longhi *et al.*, 1998; Lamzin *et al.*, 1999; Jelsch *et al.*, 2000). This could be of extreme importance in studying the interactions of an enzyme with its substrate or the oxidation states of metal ions in redox enzymes. Even if the protein is not an enzyme, ultrahigh-resolution structures can serve as reference tools for validation and refinement of structures of more modest resolution, since so few ultrahigh-resolution structures are currently available.

Here, we present an ultrahigh-resolution structure of *Trichoderma reesei* hydrophobin HFBII to a resolution of 0.75 Å. The previously published atomic resolution (1.0 Å) structure of HFBII (PDB code 1r2m; Hakanpää, Paananen *et al.*, 2004) was subjected to similar refinement and analysis for comparison. Structures were refined with *SHELX* and the multipole refinement of the ultrahigh-resolution data was initiated with *MOPRO*. The ultrahigh-resolution structure was carefully analyzed for the quality of electron-density map, disorder, residual density, observed H atoms, intramolecular and intermolecular interactions, interactions of the hydrophobic patch, solvent, *B* factors and anisotropy as well as bond lengths and angles. The atomic and ultrahigh-resolution structures have been compared with each other to distinguish the benefits of ultrahigh-resolution data. The ultrahigh-resolution structure of HFBII was also compared with other ultrahigh-resolution structures found in the Protein Data Bank in order to highlight the common features of these structures.

Hydrophobins are small secreted proteins of filamentous fungi that are characterized by their ability to spontaneously self-assemble to form an amphiphilic layer on a hydrophobic–hydrophilic (*e.g.* air–water) interface. Hydrophobins are divided into two classes according to the solubility of the assembled layers and their hydropathy patterns (Wessels, 1994). *T. reesei* hydrophobin HFBII is a class II hydrophobin with a molecular weight of about 7 kDa. Hydrophobins show low sequence similarity in general and even within a given class. However, a common feature is the presence of eight conserved cysteine residues in the sequence (Fig. 1).

Their tendency to aggregate remained an obstacle for the structure determination of hydrophobins and it was only in 2004 that the first crystal structure of a hydrophobin was solved (Hakanpää, Paananen *et al.*, 2004). To date, this is still the only crystal structure available, but crystallization conditions and preliminary X-ray characterization have been reported for another hydrophobin (Askolin *et al.*, 2004). The structure of HFBII was solved to an atomic resolution of 1.0 Å and revealed a compact rigid structure with a novel fold (Fig. 1).

Importantly, the hydrophobin structure revealed a hydrophobic patch on the protein surface (Fig. 1) comprising of 12% of the total surface area. This hydrophobic surface area gives rise to the amphiphilic nature of the hydrophobin molecule and explains on the molecular level the capability of hydrophobins to act as surfactants. In the crystal structure, there were two molecules in the asymmetric unit, forming a compact dimer. The hydrophobic patches of this dimeric pair were partly packed against each other, concealing most of the

Table 1
Data-collection statistics for ultrahigh-resolution and atomic resolution data.

Values in parentheses refer to the highest resolution shell.

	Ultrahigh-resolution data (PDB code 2b97)	Atomic resolution data (PDB code 1r2m)
Unit-cell parameters		
a (Å)	78.93	78.66
b (Å)	46.49	46.31
c (Å)	34.86	34.59
β (°)	112.20	112.16
Space group	*C*2	*C*2
Source	BW7A, EMBL Hamburg	X11, EMBL Hamburg
Wavelength (Å)	0.7747	0.8126
Resolution range (Å)	20–0.75 (0.76–0.75)	25–1.0 (1.02–1.00)
No. of observations	433130	216723
No. of unique reflections	132282 (4133)	61978 (3006)
Completeness	98.7 (91.3)	99.8 (99.6)
R_{sym}	7.2 (58.7)	3.2 (23.5)
$I/\sigma(I)$	18.01 (1.91)	7.6 (3.2)

hydrophobic areas from the solvent. This finding suggested that hydrophobin is in its functional conformation in solution and agreed very well with the fact that hydrophobins, despite their name, are readily soluble in water. In solution, hydrophobin HFBII has been found to form tetramers (at a concentration of 10 mg ml^{-1}) and monomers (at a concentration of 0.5 mg ml^{-1}) (Torkkeli *et al.*, 2002). The concentration of HFBII in the crystallization drop was between these values, *i.e.* 4 mg ml^{-1}.

Structural studies of hydrophobins are at an extremely interesting stage at the moment. Only one structure has been determined and the need to produce more is justified to enable analysis of the universality of the hydrophobin HFBII fold. In particular, as the two hydrophobin classes differ in their properties, the structure determination of a class I hydrophobin would be of the utmost importance. However, the current HFBII structure, determined to atomic resolution and now extended to ultrahigh resolution, provides very reliable and precise information about hydrophobin structures in general.

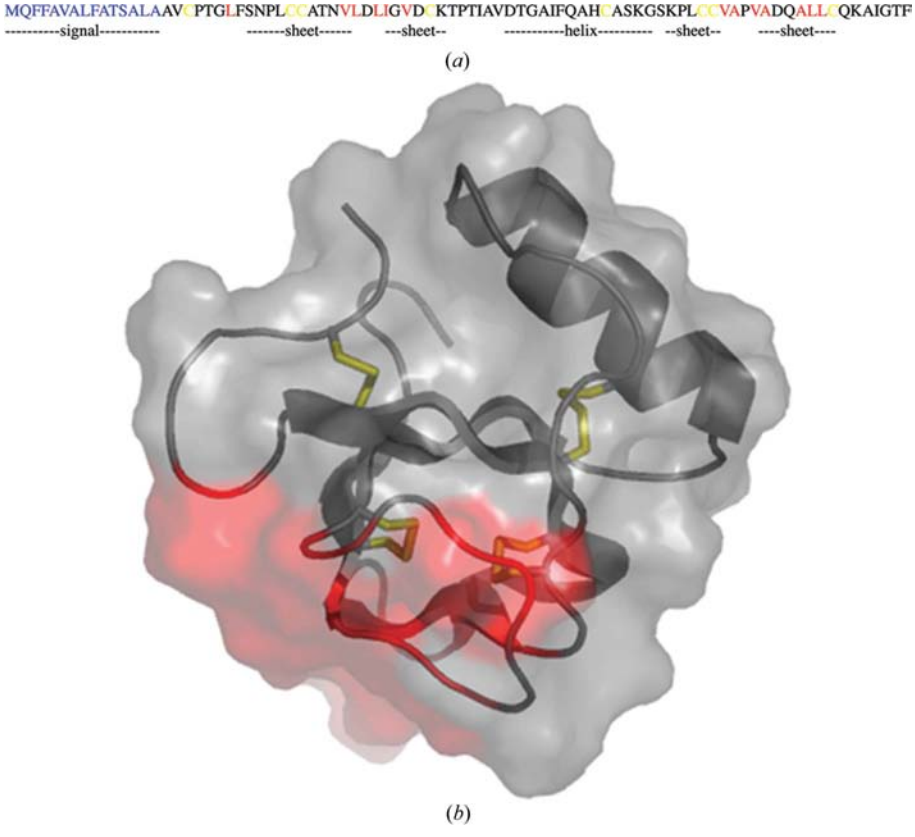

Figure 1
(*a*) Sequence of *T. reesei* hydrophobin HFBII, with yellow colour indicating the conserved cysteine residues and red colour highlighting the residues of the hydrophobic patch. The signal sequence is shown in blue. (*b*) The structure of *T. reeesei* hydrophobin HFBII; disulfide bridges are highlighted in yellow and the hydrophobic patch in red. The figure was produced using *PyMOL* (DeLano, 2002).

2. Experimental

2.1. Production, purification and crystallization

T. reesei hydrophobin HFBII was produced, purified and crystallized as described previously (Linder *et al.*, 2001; Hakanpaa, Parkkinen *et al.*, 2004). In brief, the hanging-drop vapour-diffusion method was used at 293 K. The crystallization solution contained 0.2 *M* lithium sulfate, 15% polyethylene glycol 2000 and 0.1 *M* sodium HEPES pH 7.5 as buffer. $MnCl_2$ was added directly to the crystallization drop to a final concentration of 50 m*M*. Lyophilized protein was dissolved in pure water; the protein concentration in the crystallization drop was 4 mg ml^{-1}. Streak-seeding was used to improve the quality of the crystals. The dimensions of the crystals reached dimensions of 0.3 × 0.15 × 0.15 mm.

2.2. Data collection

Prior to data collection, crystals were soaked in cryoprotectant solution (crystallization solution with the PEG 2000 concentration elevated to 30%) and flash-frozen in liquid nitrogen; measurements took place at 100 K. The data were collected on a MAR CCD 165 mm detector; low-resolution and high-resolution passes were collected separately. The program *BEST* was used to estimate the most optimal data-collection parameters (Popov & Bourenkov, 2003). Low-resolution data were collected with 1.5° oscillation width and a crystal-to-detector distance of 120 mm. High-resolution data were collected with 0.5° oscillation increments and a crystal-to-detector distance of 40 mm. The total angular range

Table 2
Outline of the refinement of the ultrahigh-resolution structure of HFBII to 0.75 Å resolution.

Refinement cycle	Model	GooF	R (at 4σ/ all data)	R_{free} (at 4σ/ all data)
1	Starting model, resolution 1.0 Å	4.70	18.4/18.9	19.9/20.35
2	Resolution 10–0.75 Å, bulk-solvent parameters added	3.81	19.2/21.1	20.9/22.6
3	Introduction of double conformers	3.78	19.1/20.9	20.6/22.4
5	Anisotropic temperature factors	2.58	13.8/15.4	15.4/16.9
6	Addition of riding H atoms	2.38	12.4/13.9	14.1/15.5
11	Rearrangement of waters and side chains	2.26	11.4/13.0	13.5/14.9
20	Resolution ∞–0.75 Å	2.27	11.5/13.0	13.4/14.8
21	Weighting scheme changed	1.50	11.7/13.2	13.6/15.0
22	Unit-cell parameters refined	1.49	11.7/13.1	13.6/14.9
32	Minor adjustments	1.48	11.7/13.1	13.6/14.9
35	ω-restraints strengthened	1.48	11.6/13.0	13.4/14.8
40	Unrestrained refinement	1.47	11.6/13.0	13.5/14.8
41	Free R removed	1.48	11.6/13.0	

covered was 171°. Data were processed, scaled and merged with *DENZO*/*SCALEPACK* (Otwinowski & Minor, 1997). Details of the data-collection statistics are presented in Table 1. For comparison, the statistics of previously collected atomic resolution data are also presented. It is noteworthy to observe that while resolution improved from 1.0 to 0.75 Å, the amount of information (*i.e.* the number of unique reflections) was doubled.

2.3. Refinement

The atomic resolution structure of HFBII was used as a starting model for refinement against the ultrahigh-resolution data and rigid-body refinement by *REFMAC* (Murshudov *et al.*, 1997) from the *CCP*4 suite (Collaborative Computation Project, Number 4, 1994) was used to fit the structure to the new data. This was followed by one round of positional refinement by *REFMAC* and the addition of 250 water molecules using *ARP*/*wARP* (Perrakis *et al.*, 1999). The reflection data as diffraction intensities were input to *SHELX* (Sheldrick & Schneider, 1997) using the program *MTZ2VARIOUS* from the *CCP*4 suite to ensure that the R_{free} flag remained the same. Multiple conformations, anisotropic displacement parameters and H atoms were added during the refinement. H atoms were added using the so-called riding model, in which the coordinates of H atoms are restrained by the coordinates of the heavier atom that they are attached to. An outline of the refinement is presented in Table 2. The program *O* was used to inspect the maps (Jones *et al.*, 1991).

At the end of the refinement, the unit-cell parameters had to be refined again in order to overcome a unit-cell scaling problem with about 0.5% difference to the existing unit-cell parameters. The corrected unit-cell parameters were $a = 78.54$, $b = 46.25$, $c = 34.69$ Å, $\beta = 111.64°$. Once the refinement had converged, the restraints were gradually released further to allow the structure to refine freely. The restraints were removed starting with the disulfide bridges, proceeding with the restraints on the main-chain geometry and eventually with the restraints on the side chains. Some side chains (residues 16, 20, 40, 46, 49, 62 and 66 in molecule *A* and residues 12, 16, 20, 40, 49, 59, 60, 62 and 66 in molecule *B*) remained restrained because they were fitted in dual conformations or had weak density for the side chain on the protein surface. Also, the flexible termini (residues 1, 2, 68, 69 and 70 in both molecules) of the protein remained restrained for the main chain and side chain. At the end of the unrestrained refinement, the free R set was removed and all data were used for refinement. Finally, one round of full-matrix least-squares refinement was performed in *SHELX* to calculate the individual e.s.d. values for atomic positions.

To evaluate the number of detected H atoms, an omit map was calculated in which the H atoms were retained but did not contribute to the map calculation. In this way, any non-zero density at the hydrogen position would indicate that the hydrogen is actually there. The ultrahigh-resolution structure of hydrophobin HFBII has been deposited in the PDB with code 2b97. Multipole refinement using the program *MOPRO* (Jelsch *et al.*, 2000) was initiated for the ultrahigh-resolution structure. The refinement is in progress and will be described separately.

The previously published atomic resolution structure had been refined with *REFMAC*, as described in Hakanpää, Paananen *et al.* (2004). In this model, the anisotropic temperature factors had been included but no H atoms had been added. The R factors at the end of the *REFMAC* refinement were 13.8 and 15.5% for R and R_{free}, respectively. These data were now subjected to similar refinement with *SHELX,* as were the ultrahigh-resolution data, to allow comparison.

The same set of reflections was used for estimating R_{free} as was used in *REFMAC* refinement. SWAT terms were introduced and the model was refined to full 0.98 Å resolution. Anisotropic temperature factors and riding H atoms were added. Multiple conformations of side chains were reconsidered as well as the positions of water molecules. With the weighting term adjusted, the final refinement of the atomic resolution model yielded R factors of 11.33 and 13.30% for R and R_{free}, respectively, and a goodness of fit (GooF) value of 1.194.

Unrestrained refinement was also gradually applied to the atomic resolution structure. R factors were 11.18 and 13.33% for R and R_{free} and the GooF was 1.210. Similarly to the ultrahigh-resolution data, the free R set was removed and final refinement was performed against all data. The R value was 11.23% and the GooF was 1.212. An omit map for hydrogen identification was calculated in similar fashion to that for the ultrahigh-resolution data and the individual e.s.d.s were calculated.

3. Results and discussion

3.1. Protein structure and electron density

The overall structure of hydrophobin HFBII was essentially the same as previously described (Hakanpaa, Parkkinen *et al.*,

2004). The structure of a HFBII molecule consisted of four antiparallel β-strands and an α-helix. There were two HFBII molecules in the asymmetric unit, corresponding to 1023 non-H protein atoms, one manganese ion and 256 water molecules.

The protein folds into a small β-barrel, which is tightly held together by two disulfide bridges formed by four of the conserved cysteine residues. All of the eight conserved cysteines are involved in disufide-bridge formation: two bridges lay within the β-barrel, one bridge connects the α-helix to the β-barrel and one bridge attaches the N-teminal loop to the barrel.

The protein was mainly very well ordered, with the exception of the C- and N-termini and a couple of flexible side chains. The atom types could be reasonably well distinguished by their electron density as shown in Fig. 2 (Sevcik *et al.*, 1996). The separation by atom type was evident in both the atomic and ultrahigh-resolution structures. For some side-chain atoms, the electron density was rather weak and ambiguous. These residues were Asp20, Asp40, Lys46, Lys49 and Lys66 in molecule *A*, and Asp40, Lys49 and Lys66 in molecule *B*. The weak density always occurred in protein side chains that were located on the surface of the protein and was most likely to be a consequence of flexibility in the structure.

The N-termini of both molecules in the asymmetric unit were slightly disordered and possibly had alternative conformations, but unambiguous fitting of the second conformations was not successful. The C-termini were also disordered starting from Ile68 and the electron density was considerably weaker than for the other parts of the protein, indicating either that the C-termini were ordered in more than one position in the structure or partial cleavage of the C-terminal part.

The C-terminal residue, Phe71, was not located in the electron density and was therefore not included in the model. We have verified by ESI–FTICR mass spectrometry (data not shown) that in the protein sample prior to crystallization the protein chain is intact for half of the material and half has been cleaved between Thr70 and Phe71. It is therefore possible that the protein molecules in the crystal are of the form that lacks the C-terminal phenylalanine, since the absence of this large hydrophobic residue on the protein surface might assist crystallization.

On the protein surface, there is a slight groove close to the C-terminal Thr70, a possible site for incorporation of the side chain of Phe71. The possible function of the C-terminal Phe71 could thus be to anchor the C-terminus to the protein core and lack of the terminal Phe71 could then cause disorder in the C-terminus. Also, the disorder of residues Asp20, Gln40, Lys49 and Lys66 could be explained by the absence of Phe71, since they are located close to the C-terminus and may be allowed more mobility in the absence of Phe71.

After the conventional refinement of the ultrahigh-resolution data, some residual density ($F_o - F_c$, 2σ) was still visible on the protein main chain at the peptide bonds and carbonyl O atoms as well as around the sulfur positions in the disulfide bridges. Residual density could also be found around the manganese ion. Residual density ($F_o - F_c$, 2σ) was rarely present around the peptide bonds or around the disulfide bridges of the atomic resolution structure, but was observed around the manganese ion.

Main- and side-chain atoms of 12 residues (8.6% of the total number of residues in the asymmetric unit) were fitted into two conformations. Most of the disorder was observed in the side-chain atoms, but in the case of residue Leu62 in both molecules the main chain also showed two conformations. The residues with double conformers varied a little between the two molecules in the asymmetric unit. In molecule *A* residues Thr16, Gln40, Lys46, Leu62 and Lys66 were fitted in two conformations, while in molecule *B* residues Leu12, Thr16, Asp20, Asp59, Gln60, Leu62 and Lys66 were modelled in two positions. The disordered residues are all located on the surface of the protein, except for Leu12, which is located in the β-barrel and partially buried by a loop region. Residues Leu12, Thr16, Asp59, Gln60 and Leu62 were rather distinctly

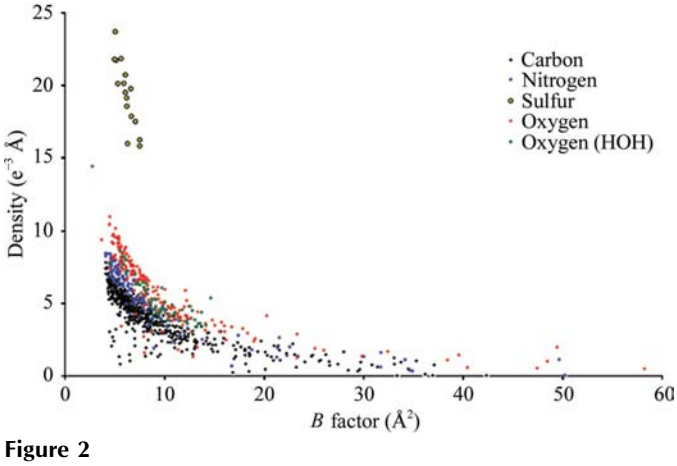

Figure 2
The electron densities at atomic positions in the ultrahigh-resolution structure of HFBII plotted as function of the *B* factors. Atom types can be distinguished. The figure is by courtesy of Dr Victor Lamzin.

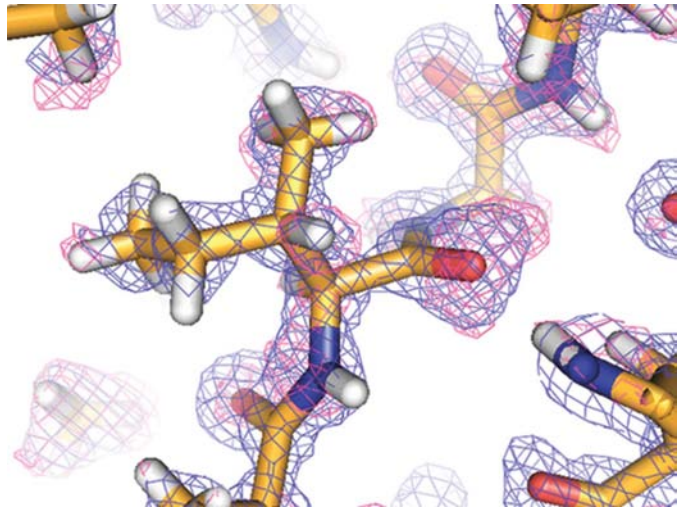

Figure 3
Omit map of HFBII ultrahigh-resolution structure showing density at the hydrogen positions of Ile22, chain *B*. The figure was produced using *PyMOL* (DeLano, 2002). $2F_o - F_c$ maps at 1.0σ are shown in blue and $F_o - F_c$ maps at 2.0σ in pink.

observed in two positions (discrete disorder), while the density was weak and ambiguous for Asp20, Gln40, Lys46 and Lys66 and even the dual conformation was inadequate to describe the disorder in these parts (unresolved disorder). Some double conformers lay close in sequence (Asp59, Gln60, Leu62) and some close in space (Asp20 and Lys66, Leu12 and Leu62). Disorder occurred mostly as isolated cases and could not be related to any functionality, although the lack of the C-terminal Phe71 might have contributed to disorder of certain residues as discussed above.

3.2. H atoms

Of the 990 H atoms present in the two protein molecules, 591 (59.7%) presented non-zero density at the hydrogen position in the omit map of the ultrahigh-resolution structure (Fig. 3) and were therefore declared directly observed. In the atomic resolution structure, 520 (52.5%) H atoms were observed. Low *B* factors usually indicated that H atoms bonded to that atom can be observed, while disordered residues were much more unlikely to show density for H atoms. In Fig. 4, a histogram is presented with the percentage of observed H atoms as a function of *B* factor of the atom the hydrogen is bonded to. The average *B* factor for the observed H atoms in the ultrahigh-resolution structure was 10.7 Å^2, while the average for non-observed H atoms was 18.9 Å^2. The average *B* factors for the H atoms of the atomic resolution structure were 15.0 and 24.4 Å^2 for observed and non-observed H atoms, respectively. The *B* factors of H atoms were not refined freely, but were restrained to be 1.2 times the value of the *B* factor of the atom the hydrogen is attached to.

There are three aspartic acid residues, one histidine residue and no glutamic acid residues in the HFBII molecule. The protonation states of histidine, aspartic acid and glutamic acid residues as well as that of the acidic C-terminus can be visually estimated. In the refinement, the histidine was assigned to be protonated at the ring nitrogen ND1, while the acidic groups were assumed not to be protonated. The omit map calculated after refinement indicated zero density for the hydrogen in the ND1 position in the ultrahigh-resolution structure. For the NE2 position, residual density exists for the hydrogen, indicating protonation of NE2 instead of ND1. However, since only half of the H atoms can be detected visually, this estimation of H-atom positions is not very reliable.

For the acidic groups, indication of protonation could be detected if the two CO distances in the acid group differed considerably in the unrestrained refinement. Since many of the aspartic acid residues and the C-termini could not be allowed to refine freely owing to disorder, only residues Asp20 and Asp59 from molecule *A* and Asp25 from molecule *B* could be investigated in this way. In these residues the CO distances in the carboxyl group were equal, suggesting no protonation. This could be expected, since the measured pH of the crystallization solution was close to neutral, with a value of 7.39.

3.3. Internal interactions

As mentioned above, four disulfide bridges were found in each HFBII molecule. Disulfide bridges were formed between residues Cys3 and Cys52, Cys13 and Cys43, Cys14 and Cys26, and Cys53 and Cys64. Two of the disulfide bridges were formed by cysteines that were consecutive in the sequence. In addition to the disulfide bridges, several hydrogen bonds hold together the HFBII molecule. A total of 85 internal hydrogen bonds were detected in the two molecules of the asymmetric unit, 45 hydrogen bonds in molecule *A* and 40 in molecule *B*, when studied with the program *XtalView* (McRee, 1999). 52 of these 85 hydrogen bonds are formed between main-chain atoms in a manner characteristic of the secondary-structure elements in which they occur. In addition, there were nine hydrogen bonds between main-chain atoms outside the secondary-structure array, 21 hydrogen bonds between a main-chain and a side-chain atom and five hydrogen bonds between two side-chain atoms. The hydrogen-bond network was quite similar in both molecules of the asymmetric unit and only minor variations were found. However, during validation by the program *WHATIF* (Vriend, 1990), several unsatisfied hydrogen-bond donors and acceptors were found, which

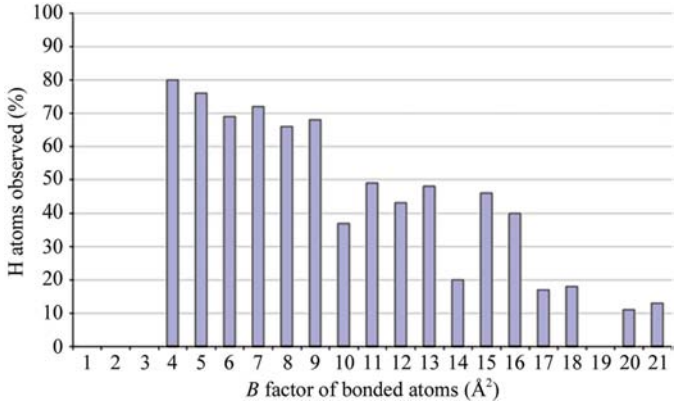

Figure 4
Histogram of the frequency of observed H atoms as a function of the *B* factor of the bonded atom.

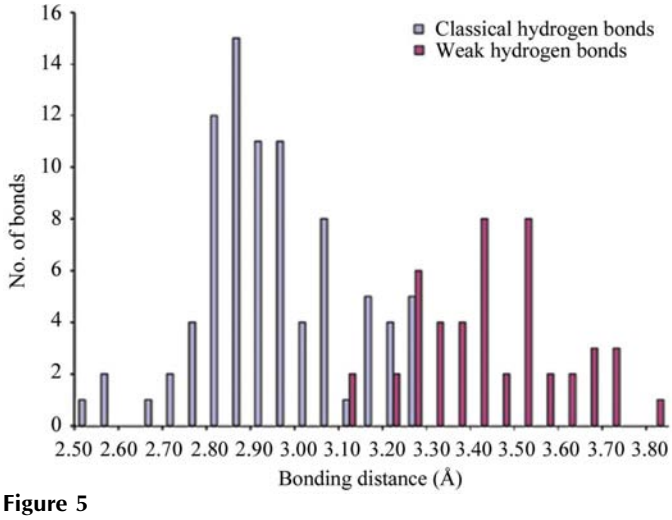

Figure 5
The distance distribution of the hydrogen-bonding distances for weak and classical hydrogen bonds.

Table 3
Intermolecular contacts between symmetry-related molecules.

Symmetry relation to (x, y, z)	Atom in the asymmetric unit	Atom in the symmetry-related molecule	Distance (Å)	Contact type
($-x$, y, $-z$) plus (0 0 1)	*A*7 CD2	*A*7 CD2	3.79	van der Waals
($-x$, y, $-z$) plus (0 0 1)	*A*55 CB	*B*54 O	3.57	van der Waals
($-x$, y, $-z$) plus (0 0 1)	*B*65 NE2	*B*65 OE1	3.00	Hydrogen bond
($-x$, y, $-z$) plus (0 0 1)	*B*8 CE2	*B*8 CE2	3.44	van der Waals
($-x$, y, $-z$)	*A*25 OD1	*B*25 OD2	3.26	Coordinated to Mn
($-x$, y, $-z$)	*A*60 OE1	*B*25 OD2	3.18	Coordinated to Mn
($-x$, y, $-z$)	*A*25 OD2	*B*60a NE2	2.87	Hydrogen bond
($-x$, y, $-z$)	*A*60 CD	*B*25 OD2	3.40	van der Waals
($-x$, y, $-z$)	*B*16*b* OG1	*B*19 N	3.08	Hydrogen bond
($-x$, y, $-z$)	*B*15 CB	*B*19 CD2	3.65	van der Waals
(x + 1/2, y + 1/2, z) plus (-1 -1 0)	*A*27 NZ	*B*34 OD1	2.71	Salt bridge
(x + 1/2, y + 1/2, z) plus (-1 -1 0)	*A*46 NZ	*B*9 O	2.74	Hydrogen bond
(x + 1/2, y + 1/2, z) plus (-1 -1 0)	*A*45 O	*B*11 CD	3.47	van der Waals
($-x$ + 1/2, y + 1/2, $-z$) plus (0 -1 1)	*A*34 OD2	*A*25 OD1	2.98	Coordinated to Mn
($-x$ + 1/2, y + 1/2, $-z$) plus (0 -1 1)	*A*34 OD2	*A*60 OE1	3.10	Coordinated to Mn
($-x$ + 1/2, y + 1/2, $-z$) plus (0 -1 1)	*A*41 CB	*A*57 CG2	3.83	van der Waals
($-x$ + 1/2, y + 1/2, $-z$) plus (0 0 1)	*A*58 O	*A*41 CB	3.62	van der Waals
(x + 1/2, y + 1/2, z) plus (-1 -1 -1)	*A*34 OD1	*B*27 NZ	2.81	Salt bridge
(x + 1/2, y + 1/2, z) plus (-1 -1 -1)	*A*34 OD1	*B*27 CD	3.56	van der Waals
($-x$, y, $-z$) plus (0 -1 1)	*A*35 CG2	*B*34 CB	3.87	van der Waals

hinted that the structure should be inspected for weak hydrogen bonds.

While classical hydrogen bonds are formed between a donor and an acceptor that are both electronegative atoms, such as oxygen or nitrogen, weak hydrogen bonds may form even with a carbon as a donor or a π-electron system (*i.e.* an aromatic residue) as an acceptor. The *NCI* server (Madan Babu, 2003) was used to search for weak hydrogen bonds in the HFBII ultrahigh-resolution structure. The interactions searched for were main-chain NH–π, C$^{\alpha}$H–π and C$^{\alpha}$H–OC and side-chain (Lys) NH–π and (Ser, Thr, Tyr) OH–π. The server found a total of 47 C$^{\alpha}$H–OC interactions in the two molecules of the asymmetric unit. The criteria for detection of this type of weak hydrogen bond was that the distances C$^{\alpha}$–CO and H$^{\alpha}$—CO are less than or equal to 3.8 and 3.3 Å, respectively, and the angles C$^{\alpha}$—H$^{\alpha}$—C and H$^{\alpha}$—C—O are greater than or equal to 120 and 90°, respectively.

Almost all of these weak hydrogen bonds were present in both molecules; only two bonds were unique to molecule *A* and one bond to molecule *B*. When the omit maps calculated for hydrogen identification were inspected, 41 of the H atoms in the donors of weak hydrogen bonds had non-zero density and were thus observed. Of these, 26 weak hydrogen bonds could be identified in the β-sheet array. The distance distribution of the bonding distances in classical and weak hydrogen bonds detected in the ultrahigh-resolution structure are presented in Fig. 5. However, the distinction between weak hydrogen bonds and van der Waals interactions is not clear, especially when the bonding distance is remarkably larger than that of classical hydrogen bonds.

As shown in Fig. 5, the shortest classical hydrogen-bonding distance observed was 2.54 Å. However, this hydrogen bond is formed by a donor in a disordered residue and is thus not reliable for hydrogen-bonding distance. Altogether, there are 12 hydrogen bonds involving an atom from a disordered residue. Of these, three are abnormally short, as described above. The shortest 'true' hydrogen-bonding distance is observed between residues His42 and Ser45 in molecule *A*. The distance between His42 O and Ser45 OG is 2.67 Å. In molecule *B*, the corresponding hydrogen bond exists with a bonding distance of 2.73 Å.

No *X*H–π interactions involving aromatic residues were detected in the *NCI* search. In one HFBII molecule there are only two aromatic residues, Phe8 and Phe39. When the structure was visually inspected, the lack of *X*H–π interactions in the *NCI* search proved logical, since the server only looks for specific donors (main-chain NH and C$^{\alpha}$H, lysine side-chain NH and OH of the serine, threonine and tyrosine side chains). None of these donors were available in the vicinity of the phenylalanine residues.

The phenylalanine residues seem to play an important role in the internal interactions. Phe39, a well conserved residue, is located in the helical part of the protein and the side chain of the residue points towards the β-barrel. The phenylalanine forms a hydrophobic islet and nearby hydrophobic residues (Pro11, Pro29 Pro50, Ala67, Val2 and Val33) pack around it. This hydrophobic interaction connects the α-helix to the β-barrel from the opposite end of the helix compared with the disulfide bridge. Similarly, Phe8 makes interactions with the nearby Leu7, Leu63 and Val54. Phe8 is just about located on the protein surface and is adjacent to the hydrophobic patch, also making interactions with the neighbouring symmetry-related molecules.

3.4. Hydrophobic patch and crystal packing

In the crystal structure, the hydrophobic patches on the protein surface of the two protein molecules in the asymmetric unit are mostly packed against each other and therefore concealed from the solvent. The hydrophobic patch consists of the side chains of Val7, Val18, Leu19, Leu21, Ile22, Val24, Val54, Ala55, Val57, Ala58, Ala61, Leu62 and Leu63 (Fig. 1). Only residue 7 comes from outside the β-barrel structure, from the N-terminal loop, and the neighbouring Phe8 is also partially involved. The closest contacts between the two molecules of the asymmetric unit are through the hydrophobic patch; the distance is at its shortest between the side chains of residues Val24 (molecule *A*) and Leu19 (molecule *B*), equalling 3.58 Å. Several side-chain atoms on the hydrophobic surface are about 3.8 Å from the side-chain atoms of the hydrophobic patch of the other molecule in the asymmetric unit.

Each molecule *A* makes crystal contacts with eight symmetry-related molecules, whereas each molecule *B* makes crystal contact with seven symmetry-related molecules. Interactions occur between two neighbouring *A* molecules and two neighbouring *B* molecules, as well as between molecule *A* and molecule *B* of adjacent units. Crystal contacts involve interactions between the hydrophobic surface areas of neighbouring molecules, contacts through the manganese ion, salt bridges, hydrogen bonds and van der Waals contacts. Symmetry relations and examples of symmetry contacts are presented in Table 3. Symmetry relations and contacts between residues were calculated with *EDPDB* (Zhang & Matthews, 1995).

The manganese ion coordinated to the protein structure appears to be very important for the lattice contacts. Hydrophobin HFBII is not a metalloprotein and the manganese ions originate from the crystallization solution, where their introduction had dramatic effects on the crystal diffracting power, as described previously (Hakanpaa, Parkkinen *et al.*, 2004). There is one manganese ion per two HFBII molecules in the asymmetric unit. The binding site is located on the side of the β-barrel, close to the hydrophobic patch.

The manganese ion is hexacoordinated, in octahedral fashion, to two water molecules and four charged protein side chains: three aspartates and a glutamine, located on the protein surface. Asp25 and Gln60 are those of molecule *A*. Asp25 is from molecule *B* from the neighbouring asymmetric unit $(-x, y, -z)$ and Asp34 is from molecule *A* from another neighbouring asymmetric unit $[(-x + 1/2, y + 1/2, -z)$ plus $(0\ -1\ 1)]$. The coordination of manganese and the distances to the ligands are presented in Fig. 6.

Additional salt bridges are formed between Lys27 of molecule *A* and Asp34 of the symmetry-related $(x + 1/2, y + 1/2, z)$ plus $(-1\ -1\ 0)$ molecule *B* and between Asp34 of molecule *A* and Lys27 of the symmetry-related $(x + 1/2, y + 1/2, z)$ plus $(-1\ -1\ -1)$ molecule *B*.

A hydrophobic contact exists between the asymmetric unit and a symmetry-related asymmetric unit $(-x, y, -z)$ plus $(0\ 0\ 1)$ through the hydrophobic surface areas of neighbouring molecules. Even though the hydrophobic surface area is mostly buried in the dimer interface of the asymmetric unit, a small portion is still accessible to solvent. In the crystal structure this area is partly concealed by the neighbouring asymmetric unit. Here, the contacts are not quite as tight as between the molecules within one asymmetric unit.

The solvent-accessible areas of both molecules in the asymmetric were calculated with *EDPDB* (Zhang & Matthews, 1995). The solvent-accessible area of an isolated HFBII dimer was calculated as 7338 Å^2. The buried solvent-accessible area of molecule *A* was calculated as 304 Å^2 and that for molecule *B* as 315 Å^2, whereas the size of the hydrophobic patch on the protein surface is about 400 Å^2.

Water molecules located within 5 Å from the side chains of the residues of the hydrophobic patch were analyzed. 46 water molecules were found within this distance with an average *B* factor of 15.42 Å^2 and a mean distance of 3.93 Å from a hydrophobic side chain. Most of these waters were hydrogen bonded to the protein through main- or side-chain O or N atoms, but ten water molecules formed hydrogen bonds with other water molecules only. No waters were found between the hydrophobic patches of the two molecules of the asymmetric unit that pack close together. Residue Ile22 in molecule *A* and residues Leu7, Val18 and Ala55 in molecule *B* had no water molecules within 5 Å distance of the side chain. The closest contact between an atom of a hydrophobic side chain and water molecule was that of CB of Ala55 in molecule *A*, with a distance of 2.95 Å.

3.5. Solvent

256 water molecules were located in the asymmetric unit. 20 water molecules had occupancies less than 1.0 and two of the water molecules were located in special positions. During the refinement, water molecules were added to the model only if corresponding density could be found in the $2F_o - F_c$ map at 1.0σ and the isotropic temperature factor was less than 60 Å^2. No H atoms were added to water molecules. The Matthews coefficient (Matthews, 1968) was estimated as 1.92 Å^3 Da^{-1}, corresponding to a solvent content of 37%. If the volume of a water molecule is assumed to be about 30 Å^3, the solvent content corresponds to about 1460 water molecules in the unit cell. The amount of water molecules located in the entire unit cell is 1024, which corresponds to the identification of about 70% of possible water molecules in the structure. 167 water molecules formed direct hydrogen bonds to protein atoms, forming the first hydration shell. 60 water molecules were present in the second hydration shell and 14 in the third hydration shell.

3.6. Bond lengths and angles

The r.m.s.d. values for bond lengths and bond-angle distances were 0.017 and 0.037 Å, respectively, for the ultra-high-resolution structure after restrained refinement. The corresponding values for the atomic resolution structure were 0.017 and 0.035 Å for bond lengths and bond-angle distances, respectively.

The ultrahigh-resolution structure was validated with *WHATIF* (Vriend, 1990) and *PROCHECK* (Laskowski *et al.*, 1993) after restrained and unrestrained refinement. *WHATIF* detected 16 unusual bond angles and one unusual bond length in the restrained model. For the unrestrained model, 23 unusual bond angles and 15 unusual bond lengths were detected. In the restrained model, the deviating angles and distances were for residues with dual conformations, residues on the protein surface or residues of the flexible termini. The additional angles and distances found in the unrestrained data were mostly of well ordered main- and side-chain atoms, but the deviations were only minor. For both refinements bond-angle variability was found to be normal, with an r.m.s.d. of 2.545° for unrestrained refinement and 2.335° for restrained refinement. The variability was to be found normal for bond lengths, with r.m.s.d.s of 0.027 Å for unrestrained refinement and 0.017 Å for restrained refinement. The Ramachandran plot was excellent, with 88.8% of the residues in the most

Table 4
B factors (Å^2) of ultrahigh-resolution and atomic resolution structures after restrained and unrestrained refinement.

	Atomic, restrained	Atomic, unrestrained	Ultrahigh, restrained	Ultrahigh, unrestrained	No. of atoms atomic/ ultrahigh
B (main chain)	12.40	12.41	8.32	8.31	570/570
B (side chain)	15.07	15.12	10.78	10.81	455/453
B (protein)	13.59	13.61	9.41	9.42	1025/1023
B (solvent)	21.54	21.75	19.64	19.77	217/256
B (Mn)	8.65	8.67	4.44	4.45	1/1
B (all atoms)	14.97	15.03	11.45	11.49	1243/1280

favoured regions and 11.2% in additionally allowed regions for both the restrained and the unrestrained ultrahigh-resolution structures.

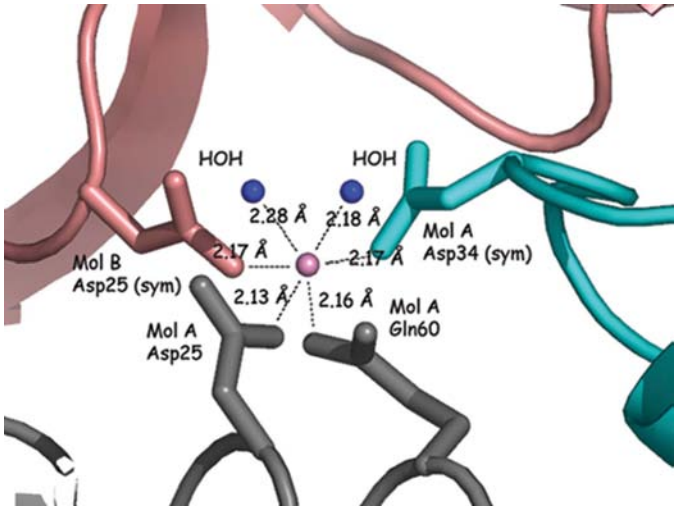

Figure 6
Coordination of the manganese ion. The figure was produced using *PyMOL* (DeLano, 2002).

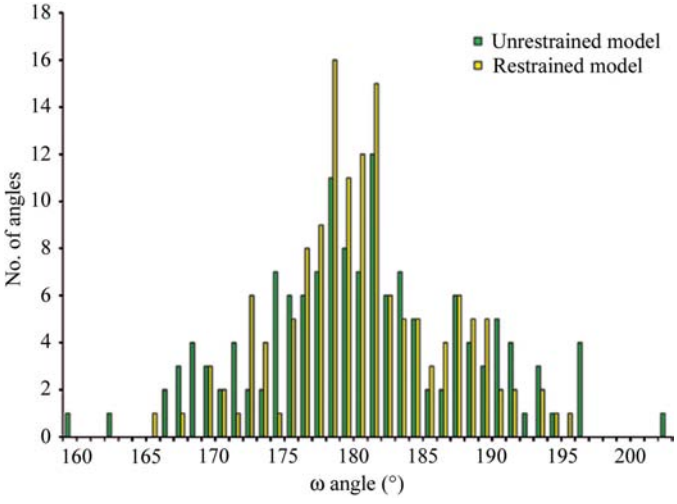

Figure 7
Distribution of the ω angles in the the restrained and unrestrained ultrahigh-resolution model.

To estimate the true error in bond lengths, instead of the variability (the r.m.s.d.), a bond very unlikely to change, namely the $C^{\alpha}-C^{\beta}$ bond, was inspected more closely. There were five $C^{\alpha}-C^{\beta}$ bond distances that deviate from small-molecule values by more than 0.05 Å in the unrestrained ultrahigh-resolution structure, as listed by *PROCHECK*. The deviating distances were 1.464 Å (*A* Leu19), 1.471 Å (*B* Asn17), 1.466 Å (*B* Leu19), 1.663 Å (*B* Ala37) and 1.603 (*B* Ala44). Only one deviating bond length was found in the restrained ultrahigh-resolution structure, with a distance of 1.478 Å (*B* Asn17). The small-molecule values for these bond lengths are 1.530 Å for leucines, and asparagine and 1.521 Å for alanines.

Torsion-angle evaluations showed unusual values for Pro50 and Pro56 in molecule *A* and Pro56 in molecule *B*. All these residues are well ordered with clear unambiguous density. Pro50 is located on the β-strand that follows the α-helix. Pro50 packs against the conserved Phe39 and Pro56 is located in a β-hairpin loop.

Backbone torsion-angle evaluation revealed unusual conformations with poor φ–ψ combinations in eight residues (the same for restrained and unrestrained refinements). This accounts for 5.7% of residues and is not exceptional. For the ω angles, the default restraint in *SHELX* produced a standard deviation of 7.218 and thus appeared to be too weak as judged by the validation programs. The restraint was tightened at the end of the restrained refinement and finally yielded a standard deviation of 5.714, which is in the normal range. However, the actual standard deviation of ω values for the unrestrained refinement was 7.502. The median of the ω-angle distribution was 179.6° (average 180.1°) for the restrained refinement and 179.9° (average 180.1°) for unrestrained refinement (Fig. 7). Even though the distribution was broader for the unrestrained structure, the mean value was slightly closer to the ideal value of 180°. The most deviating ω angle is between residues Ile31 and Val32 in molecule *A*. This angle has an ω value of 164.5° in restrained structure and 159.1° in the unrestrained. Residues Ile31 and Val32 are located on the loop preceding the α-helix.

3.7. *B* factors and anisotropy

After the restrained refinement, the average isotropic *B* factor was 11.45 Å^2 for all atoms and 9.41 Å^2 for protein atoms. Table 4 presents the average *B* factors for both the atomic and the ultrahigh-resolution structure after restrained and unrestrained refinement, calculated with the program *MOLEMAN* (Kleywegt *et al.*, 2001*a*,*b*). The temperature factors are considerably lower for the ultrahigh-resolution structure in comparison with the atomic resolution structure. Since low *B* factors (Afonine *et al.*, 2004) are one of the criteria for the selection of atoms for multipole refinement with *MOPRO*, this further supported (as did the residual density in the ultrahigh-resolution maps) the application of this refinement method for the ultrahigh-resolution structure. Also, no significant increase in *B* factors could be detected when the restraints were removed, indicating that there is no overfitting of the model, even at atomic resolution.

Table 5
Mean anisotropies of the ultrahigh-resolution and atomic resolution structures.

Values in parentheses are the number of atoms used to calculate the mean anisotropy.

	Atomic	Ultrahigh
A (protein)	0.559 (1025)	0.437 (1023)
A (solvent)	0.447 (217)	0.362 (256)
A (Mn)	0.832 (1)	0.692 (1)
A (all atoms)	0.539 (1243)	0.422 (1280)
A (N atoms)	0.586 (168)	0.457 (169)
A (C atoms)	0.556 (646)	0.437 (642)
A (O atoms)	0.535 (195)	0.411 (196)
A (S atoms)	0.654 (16)	0.511 (16)

Table 5 presents the mean atomic anisotropies calculated with the program *PARVATI* (Merritt, 1999). No protein atoms were found to be perfectly isotropic in *PARVATI* mean anisotropy evaluation. The most extreme anisotropy in the ultrahigh-resolution structure was found for the carbonyl O atom of Ala61 in molecule *B*. In molecule *B*, there are consecutive residues in the sequence that are nicely ordered in two conformations. These residues are Asp59, Gln60 and Leu62. Ala61 lies in the middle of this region of dual conformations, which suggests that the large anisotropy originates from the carbonyl O atom being disordered into two positions but left unmodelled. The difference in anisotropy between atomic and ultrahigh-resolution structures could stem from the fact that on going to lower resolution, the isotropic *B* factors appear to artificially increase, thereby smearing out details of atomic shape (Afonine *et al.*, 2004). This behaviour is the reason for the *B*-factor cutoff usually applied in multipole refinement.

3.8. Comparison of HFBII models

Molecules *A* and *B* of the ultrahigh-resolution structure were least-squares fitted in *SHELXPRO* (Sheldrick & Schneider, 1997). Fitting of 974 atoms produced an overall r.m.s.d. of 1.50 Å, while fitting the C^{α} atoms in *XtalView* (McRee, 1999) produced an r.m.s.d. of 0.80 Å. The main difference is in the C-termini, where the terminal residue Thr70 comes much closer to the helical part of the protein in molecule *B* than in molecule *A*. This also has an effect on the side-chain conformation of residue Gln40, which is located in the α-helix. More subtle differences can be ascribed to the N-terminus and some side-chain conformations. A few examples of the side chains are Ile22 in the hydrophobic patch and residues Asp34 and Lys27. The isoleucines are located in the hydrophobic patch and pack against each other on the surface of molecules *A* and *B*. The side-chain atom CD1 of Ile22 in molecule *B* bends away from molecule *A* to avoid clashing with CG1 of Ile22 of molecule *A*. Asp34 of molecule *A* is coordinated to the manganese ion, whereas the Asp34 of molecule *B* is not. Instead, Asp34 of molecule *B* forms a salt bridge with the Lys27 of a symmetry-related molecule *A*.

The ultrahigh-resolution structure was also least-squares fitted with the atomic resolution structure for both chains. With 1953 atoms fitted, the overall r.m.s.d. was 0.8143 Å and the structures were just about identical. For fitting the C^{α} atoms, the r.m.s.d. was 0.42 Å. Slight diversity could be detected in the termini and some side-chain conformations.

Throughout the refinement process, the *R* factors remained higher for the ultrahigh-resolution structure compared with the atomic resolution structure: *e.g.* at the end of the unrestrained refinement the *R* factor was 11.2% for the atomic resolution structure and 13.0% for the ultrahigh-resolution structure. This could be a consequence of differences in the crystals, since atomic and ultrahigh-resolution data were collected from separate crystals in different beamlines at different times and using different criteria for data collection. Also, the difference in the *R* factors could reflect the fact that the refinement of atomic resolution data is much more complete at this stage, since there are fewer features to model.

No indications of specific structural radiation damage were detected in the electron-density maps of either structure, in spite of the long data collection at a synchrotron beam. The disulfide bridges were all intact and even though some aspartate side chains had weak densities, this is more likely to arise from the flexibility of the structure than actual radiation damage. However, long exposure to a synchrotron beam also causes non-specific damage that can be seen as an increase in the temperature factors. The data-collection time for the ultrahigh-resolution structure (about 48 h) was significantly longer than that for atomic resolution structure (about 12 h). Therefore, the higher *R* factors for the ultrahigh-resolution data may also be explained by non-specific radiation damage.

Table 6
Current ultrahigh-resolution protein structures.

Protein	PDB code	Resolution (Å)	Space group	No. of residues	Solvent (%)
Crambin	1ejg	0.54	$P2_1$	46	30
Antifreeze	1ucs	0.62	$P2_12_12_1$	64	41
AR	1us0	0.66	$P2_1$	316	35
Rubredoxin	1yk4	0.69	$P2_12_12_1$	52	27
Syntenin	1r6j	0.73	$P2_1$	82	31
HFBII	2b97	0.75	$C2$	70 × 2	37
Subtilisin	1gci	0.78	$P2_12_12_1$	269	45
PAK pilin	1x6z	0.78	$P4_12_12_1$	123	nd
HiPIP	1iua	0.80	$P2_12_12_1$	83	34
Trypsin	1pq7	0.80	$P1$	224	32
CBM36	1w0n	0.80	$P2_12_12_1$	131	30

3.9. Comparison of all ultrahigh-resolution structures

In the PDB, nine protein structures qualify as ultrahigh-resolution structures. In addition, one ultrahigh-resolution protein structure is on hold. These proteins are presented in Table 6. Currently available ultrahigh-resolution protein structures are those of crambin (Jelsch *et al.*, 2000), antifreeze protein RD1 (Ko *et al.*, 2003), human aldose reductase–inhibitor complex (Howard *et al.*, 2004), rubredoxin mutant W4L/R5S (Bönisch *et al.*, 2005), the PDZ2 domain of syntenin (Kang *et al.*, 2004), serine protease subtilisin (Kuhn *et al.*, 1998), PAK pilin (Dunlop *et al.*, 2005), high-potential iron–sulfur protein HiPIP (Liu *et al.*, 2002), trypsin (Schmidt *et al.*,

Table 7
Parameters of refinement for ultrahigh-resolution structures.

Protein	R (%)	R_{free} (%)	Program	R.m.s.d. bond (Å)	R.m.s.d. angle (°)
Crambin	9.0	9.4	*MOPRO*	0.023	2.7
Antifreeze	13.3	15.5	*SHELX*	0.012	2.1
AR	9.4	10.3	*SHELX*	0.016	2.1
Rubredoxin	10.03	10.83	*SHELX*	0.026	nd
Syntenin	7.5	8.7	*SHELX*	0.019	1.7
HFBII	13.0	14.8	*SHELX*	0.027	2.5
Subtilisin	9.9	10.3	*SHELX*	0.012	1.68
PAK pilin	14.2	15.4	*SHELX*	0.019	1.99
HiPIP	10.1	11.4	*SHELX*	0.016	2.0
Trypsin	10.7		*MOPRO*	nd	nd
CBM36	11	13	*REFMAC*	0.024	1.6

2003) and carbohydrate-binding module CBM36 of a xylanase (Jamal-Talabani *et al.*, 2004). In Table 6, the proteins are presented starting with the structure of highest resolution data. Hydrophobin HFBII is positioned in the middle of the table with 0.75 Å resolution. The record of highest resolution is still held by crambin.

All these structures have been determined rather recently; the first one determined was serine protease subtilisin, published in 1998 (Kuhn *et al.*, 1998). Most of the proteins are small in size, but large molecules such as aldose reductase and subtilisin are also present. The space group is usually of low symmetry, often monoclinic or at the most orthorhombic, with the recently published PAK pilin being the only exception, belonging to a tetragonal space group. Hydrophobin is unique in the sense that it is the only structure with more than one protein molecule in the asymmetric unit. However, some structures include in addition moieties other than protein, *e.g.* the aldose reductase structure, which has an inhibitor molecule bound. The solvent content is rather low for all the structures, but not exceptionally low.

Some parameters of the refinement of ultrahigh-resolution structures are presented in Table 7 along with the method of refinement. Most structures were refined with *SHELX* (Sheldrick & Schneider, 1997) and only two structures have undergone multipolar refinement, even though the resolution would make it possible for several proteins. The methods employed during the refinement vary from protein to protein (such as selection of restraints, criteria for solvent molecules), which makes the structures not so easily comparable. Syntenin, even though not the structure of the highest resolution, is the structure of the highest precision, with R factors that are usually only found for small molecules.

To conclude, ultrahigh-resolution structures can provide very precise and detailed information relevant to the function of the protein. Ultrahigh-resolution structures allow direct visualization of chemical properties and are also important reference tools for structures of more modest resolution and therefore worth the time-consuming refinement. With current crystallographic tools, the number of ultrahigh-resolution structures is expected to increase.

Most of the practical work presented here was conducted during a Marie Curie Fellowship at EMBL Hamburg (Contract No. HPMT-CT-2000-00174). The entire staff of EMBL Hamburg is thanked for creating a pleasant working environment during the Marie Curie traineeship. Dr Victor Lamzin at EMBL Hamburg is gratefully acknowledged for the means of producing the *B*-factor/density plots. The Marie Curie Fellowship Grant and the ISB National Graduate School are gratefully acknowledged for funding. Dr Christian Jelsch and the *MOPRO* group in Nancy are thanked for providing us with the program *MOPRO* and constant correspondence while we were learning how to use it. Dr Janne Jänis at the University of Joensuu is thanked for recording the mass spectra of HFBII. Mrs Liisa Saharinen and Mrs Reetta Kallio-Ratilainen are thanked for skilled technical assistance.

References

Afonine, P. V., Lunin, V. Y., Muzet, N. & Urzhumtsev, A. (2004). *Acta Cryst.* D**60**, 260–274.
Askolin, S., Turkenburg, J. P., Tenkanen, M., Uotila, S., Wilson, K. S., Penttilä, M. & Visuri, K. (2004). *Acta Cryst.* D**60**, 1903–1905.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Bönisch, H., Schmidt, C. L., Bianco, P. & Ladenstein, R. (2005). *Acta Cryst.* D**61**, 990–1004.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1997). *Curr. Opin. Struct. Biol.* **7**, 681–688.
DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* DeLano Scientific, CA, USA. http://www.pymol.org.
Dunlop, K. V., Irvin, R. T. & Hazes, B. (2005). *Acta Cryst.* D**61**, 80–87.
Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A**47**, 392–400.
Hakanpää, J., Paananen, A., Askolin, S., Nakari-Setälä, T., Parkkinen, T., Penttilä, M., Linder, M. B. & Rouvinen, J. (2004). *J. Biol. Chem.* **279**, 534–539.
Hakanpaa, J., Parkkinen, T., Hakulinen, N., Linder, M. B. & Rouvinen, J. (2004). *Acta Cryst.* D**60**, 163–165.
Harata, K., Abe, Y. & Muraki, M. (1998). *Proteins*, **30**, 232–243.
Howard, E. I., Sanishvili, R., Cachau, R. E., Mitschler, A., Chevrier, B., Barth, P., Lamour, V., Van Zandt, M., Sibley, E., Bon, C., Moras, D., Schneider, T. R., Joachimiak, A. & Podjarny, A. (2004). *Proteins*, **55**, 792–804.
Jamal-Talabani, S., Boraston, A. B., Turkenburg, J. P., Tarbouriech, N., Ducros, V. M.-A. & Davies, G. J. (2004). *Structure*, **12**, 1177–1187.
Jelsch, C., Teeter, M. M., Lamzin, V., Pichon-Pesme, V., Blessing, R. H. & Lecomte, C. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 3171–3176.
Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kang, B. S., Devedjiev, Y., Derewenda, U. & Derewenda, Z. S. (2004). *J. Mol. Biol.* **338**, 483–493.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001*a*). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 353–356. Dordrecht: Kluwer Academic Publishers.
Kleywegt, G. J., Zou, J. Y., Kjeldgaard, M. & Jones, T. A. (2001*b*). *International Tables for Crystallography*, Vol. *F*, edited by M. G. Rossmann & E. Arnold, pp. 366–367. Dordrecht: Kluwer Academic Publishers.
Ko, T.-P., Robinson, H., Gao, Y.-G., Cheng, C.-H. G., Devries, A. L. & Wang, A. H.-J. (2003). *Biophys. J.* **84**, 1228–1237.
Kuhn, P., Knapp, M., Soltis, M., Ganshaw, G., Thoene, M. & Bott, R. (1998). *Biochemistry*, **37**, 13446–13452.
Lamzin, V. S., Morris, R. J., Dauter, Z., Wilson, K. S. & Teeter, M. M. (1999). *J. Biol. Chem.* **274**, 20753–20755.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Linder, M., Selber, K., Nakari-Setälä, T., Qiao, M., Kula, M.-R. & Penttilä, M. (2001). *Biomacromolecules*, **2**, 511–517.
Liu, L., Nogi, T., Kobayashi, M., Nozawa, T. & Miki, K. (2002). *Acta Cryst.* D**58**, 1085–1091.
Longhi, S., Czjzek, M. & Cambillau, C. (1998). *Curr. Opin. Struct. Biol.* **8**, 730–737.
McRee, D. E. (1999). *J. Struct. Biol.* **125**, 156–165.
Madan Babu, M. (2003). *Nucleic Acids Res.* **31**, 3345–3348.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Merritt, E. A. (1999). *Acta Cryst.* D**55**, 1109–1117.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. J. H. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Popov, A. & Bourenkov, G. P. (2003). *Acta Cryst.* D**59**, 1145–1153.
Schmidt, A., Jelsch, C., Ostergaard, P., Rypniewski, W. & Lamzin, V. S. (2003). *J. Biol. Chem.* **278**, 43357–43362.
Schmidt, A. & Lamzin, V. S. (2002). *Curr. Opin. Struct. Biol.* **12**, 698–703.
Schmidt, A. & Lamzin, V. S. (2005). *Acta Cryst.* D**61**, 1132–1139.
Sevcik, J., Dauter, Z., Lamzin, V. S. & Wilson, K. S. (1996). *Acta Cryst.* D**52**, 327–344.
Sheldrick, G. M. & Schneider, T. (1997). *Methods Enzymol.* **276**, 319–343.
Torkkeli, M., Serimaa, R., Ikkala, O. & Linder, M. (2002). *Biophys. J.* **83**, 2240–2247.
Vriend, G. A. (1990). *J. Mol. Graph.* **8**, 52–56.
Wessels, J. G. H. (1994). *Annu. Rev. Phytopathol.* **32**, 413–437.
Zhang, X.-J. & Matthews, B. W. (1995). *J. Appl. Cryst.* **28**, 624–630.