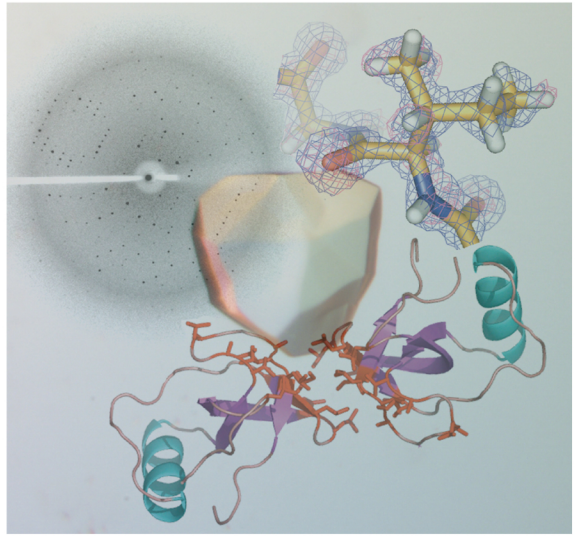
Dissertations
Department of Chemistry
University of Joensuu

78/2006

Johanna Hakanpää

Structural studies of *Trichoderma reesei* hydrophobins HFBI and HFBII - the molecular basis for the function of fungal amphiphiles



Structural studies of *Trichoderma reesei* hydrophobins HFBI and HFBII - the molecular basis for the function of fungal amphiphiles

Johanna Hakanpää

Department of Chemistry
University of Joensuu
Finland

Joensuu 2006

Johanna Hakanpää
Department of Chemistry, University of Joensuu
P.O. Box 111, 80101 Joensuu, Finland

Supervisor
Prof. Juha Rouvinen, University of Joensuu

Referees
Prof. Adrian Goldman, University of Helsinki
Prof. Jari Ylänne, University of Jyväskylä

Opponent
Prof. T. Alwyn Jones, University of Uppsala

To be presented, with the permission of the Faculty of Science of the University of Joensuu, for public criticism in Auditorium K1, Yliopistokatu 7, Joensuu, on May 12th, 2006, at 12 o'clock noon.

ISBN: 952-458-831-5 (PDF)
ISSN: 1456-3975
Dissertations, Department of Chemistry ; no78

Joensuun yliopistopaino
Joensuu 2006

ABSTRACT

Hydrophobins are small proteins secreted by filamentous fungi. They are able to spontaneously self-assemble onto a hydrophobic-hydrophilic interface and turn a hydrophobic surface hydrophilic and vice versa. Hydrophobins are also very surface active and lower the surface tension of water considerably. Hydrophobins have important functions in fungal growth, during which they are abundantly expressed. However, the origin of the unique properties of hydrophobins has remained largely unknown due to a lack of three-dimensional structural information.

In this study, the crystal structures of two hydrophobins, HFBI and HFBII, and a structure of a variant of HFBI have been determined. The hydrophobins originate from a rot fungus, *Trichoderma reesei*, and are class II hydrophobins. The structure of HFBII was determined at an ultra-high resolution of 0.75 Å and its multimerization state was found to be dimeric. The structure of the native HFBI was solved from pseudo-merohedrally twinned data and it was tetrameric. The structure of the N-Cys HFBI-variant revealed a unique crystal packing arrangement with an exceptionally high solvent content; the quaternary structure of the N-Cys HFBI was a detergent-associated octamer.

The overall structure of all the hydrophobins consisted of a small β-barrel tightly held together by four disulphide bridges formed by eight conserved cysteines in the hydrophobin sequence. An α-helix resided on one side of the β-barrel. A patch consisting of aliphatic side-chains of hydrophobic residues was identified in the protein surface and it was concluded that this patch is the source of the amphiphilicity of these molecules. The patch covered about 18% of the total surface area of each hydrophobin molecule and its shape was relatively flat.

The multimers in the crystal structures were formed by the hydrophobic patches of the adjacent molecules packing close together. The formation of a multimer concealed a considerable part of the hydrophobic patch from the solvent. Therefore, it was concluded that multimerization in solution is driven by the formation of an energetically favorable state through concealment of the hydrophobic surface areas. Furthermore, when self-assembled into a monolayer the entire hydrophobic surface area presumably faces the hydrophobic counterpart, thus, self-assembly is favored over multimerization.

A structural change was identified in the area of the hydrophobic patch in half of the molecules in the native HFBI-structure. This and the relatively high root-mean-square distance indicated that in the hydrophobin structures there is intrinsic plasticity, which might also be a requirement for these molecules that interact with surfaces that are often coarse.

LIST OF ORIGINAL PUBLICATIONS

This thesis is a summary of the following original publications I-III and the submitted manuscript IV, referred to in the text by their Roman numerals.

I Hakanpää J., Parkkinen T., Hakulinen N., Linder M., Rouvinen J. (2004) Crystallization and preliminary X-ray characterization of *Trichoderma reesei* hydrophobin HFBII, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60** 163-165.

II Hakanpää J., Linder M., Askolin S., Nakari-Setälä T., Parkkinen T., Penttilä M., Rouvinen J. (2004) Atomic resolution structure of the HFBII hydrophobin: a self-assembling amphiphile, *J. Biol. Chem.* **279** 527-533.

III Hakanpää J., Linder M., Popov A., Schmidt A., Rouvinen J. (2006) Hydrophobin HFBII in detail: ultrahigh-resolution structure at 0.75 Å, *Acta Crystallogr., Sect D: Biol. Crystallogr.* **62** 356-367.

IV Hakanpää J., Szilvay G., Kaljunen H., Maksimainen M., Linder M., Rouvinen J. Two crystal structures of *Trichoderma reesei* hydrophobin HFBI - structure of a protein amphiphile with and without detergent interaction, submitted.

CONTENTS

ABBREVIATIONS

DBE	dummy bond electron
EMBL	European Molecular Biology Laboratory
ESI-FTICR	electrospray ionization Fourier transform ion cyclotron resonance
HEPES	N-(2-hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid)
kDa	kilodalton = 10^3 g/mol
LDAO	lauryldimethylamine oxide
MS	mass spectrometry
MW	molecular weight
NCI	non-canonical interactions
NMR	nuclear magnetic resonance
OSG	1-S-octyl-β-D-thioglucoside
PDB	the Brookhaven Protein Data Bank
RMSD	root-mean-square distance
SAA	solvent accessible area
Å	ångström = 1×10^{-10} m

1. INTRODUCTION

In the late 1980s, a gene that encodes a small, secreted protein was discovered in filamentous fungi. Proteins of this type were found to cover the hyphal cell walls of the fungal aerial structures and were thus named hydrophobins[1]. In spite of its name, the protein was only moderately hydrophobic and water-soluble. As more hydrophobins were identified, it became clear that the sequence similarity among this protein family is low and yet the sequence is characterized by the presence of eight conserved cysteine residues (Figure 1).

Another characteristic feature of hydrophobins was found to be their unique capability to self-assemble into an amphiphilic protein layer at a hydrophobic-hydrophilic interface, e.g. at the boundary between air and water or between a hydrophobic solid and water[2,3]. Hydrophobins were found to be abundantly expressed during fungal growth and important for the development of fungi[4]. To date, there are 271 entries for hydrophobins in the Entrez protein database of the NCBI (National Center for Biotechnology Information) and hydrophobins have been found in at least 30 species of filamentous fungi[4]. A fungus may contain several hydrophobin genes that are expressed at different times of the fungal life-cycle and targeted for specialized functions[5,6].

Based on the differences in the assembled hydrophobin monolayers and also on the clustering of the hydrophobic and hydrophilic amino acid residues in the protein sequence, hydrophobins were grouped into two classes by Wessels[7]. Class I hydrophobins form amphiphilic layers with an appearance of amyloid-like rodlets and the layers may be dissolved only by treatment with trifluoro acetic acid, while class II

Figure 1. Sequence alignment of hydrophobins. Hydrophobins ABH2[8], ABH3[9], EAS[10] and SC3[6] are class I hydrophobins while the remaining sequences represent class II hydrophobins[5,11-18]. The conserved cysteines are highlighted in orange. The secondary structure elements of EAS[10] are indicated above and those of HFBII are below. The residues of the hydrophobic patch in the HFBII structure and the corresponding residues in class II sequences are indicated in red. The hydrophobic, aliphatic residues in class I sequences between the second and the third, as well as the seventh and the eighth cysteines, are indicated in blue.

hydrophobin layers lack the amyloid-fibril pattern and are more easily broken down, e.g. with 60% ethanol. By sequence, class II hydrophobins are more invariant in length, composition, and the spacing of the conserved cysteines, in comparison to the class I hydrophobins (Figure 1).

The size of a hydrophobin molecule varies, typically from 70 to 130 amino acid residues and the protein contains a signal sequence for secretion. Hydrophobins may also be glycosylated[19] and several hydrophobin domains can be joined via regions of glycine-asparagine repeats, e.g. in the tripartite hydrophobin CFTH1 from *Claviceps fusiformis*[16] and in the pentahydrophobin CPPH1 from *Claviceps purpurea*. A hydrophobin from *Agrocybe aegerita* contains a leucine-zipper domain in the middle of the molecule[20]. The role of this domain is thought to be in homo- or heterodimer formation.

A milestone in hydrophobin research was the discovery of the dual role the hydrophobins play in fungal growth[21] : 1) they lower the surface tension of water and 2) they provide a protective coating of fungal surface (Figure 2). When a fungus grows in a moist environment, hydrophobin is secreted from the tip of the growing hypha. The protein migrates to the air-water interface and forms an amphiphilic layer, which considerably lowers the surface tension of water. The *Schizophyllum commune* hydrophobin SC3[21] can lower the surface tension of water from 72 mJ/m^2 to 24 mJ/m^2 and the *Trichoderma reesei* hydrophobin HFBII[4] to 28 mJ/m^2.

The decreased surface tension assists the growth of the aerial hypha by lowering the mechanical stress of hypha penetrating the water surface[21]. However, once it grows into the air, the hydrophilic cell wall of the growing hypha becomes exposed to a hydrophobic air environment. To avoid this unfavorable situation, the secretion of hydrophobin continues and the protein now assembles at the interface between air and the hydrophilic cell wall, creating a protective coating on the hyphal surface. Hydrophobins also cover fungal spores, the surface and air cavities of fruiting bodies and mediate attachment to hydrophobic surfaces, e.g. in pathogenic interactions[22]. Hydrophobins are specific to filamentous fungi and have not been found in species from other kingdoms, but surface-active proteins with similar functions have been identified in filamentous bacteria[23,24].

The unique properties of hydrophobins make them interesting candidates for industrial applications. To add to the attraction, it has been shown that gram scale production of hydrophobins is possible for *T. reesei* HFBI and HFBII[25,26]. These hydrophobins are also highly soluble in water (up to 100 mg/ml) and tolerant of high temperatures[4]. The adhesion of hydrophobins to hydrophobic solids is very strong e.g. 1.5 mg of SC3 is enough to cover about 1 m^2 of teflon[27]. In addition, hydrophobins have been proven non-toxic in every-day life, since hydrophobins are ingested on the surfaces of mushrooms. Some examples of potential applications are tissue engineering, i.e. an increase in the biocompatibility of medical implants and devices with hydrophobin coating, immobilization of antibodies in biosensors, drug delivery by using vesicles

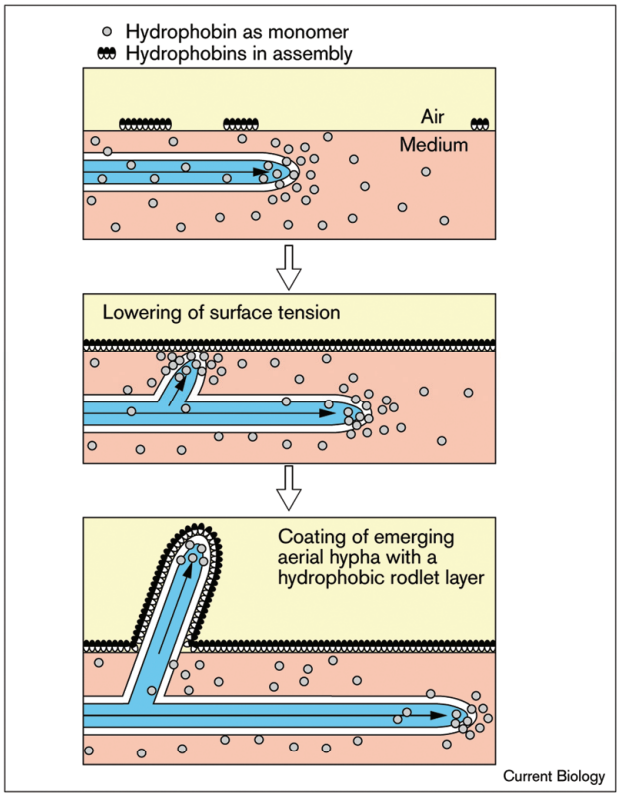


Figure 2. The model for the formation of aerial hyphae in *S. commune*. Reprinted[21] with permission from Elsevier © Elsevier Science Ltd ISSN 0960-98222

stabilized by hydrophobin, nanotechnology, i.e. patterning surfaces with nanometer accuracy[28], usage in separation techniques[29], usage as emulsifiers in food processing and indicators for beer gushing[4], anti-fouling applications and usage as additive for hair-care products in the cosmetics industry[30].

The molecular origin of the unique properties of hydrophobins has, however, remained largely unknown, due to the lack of structural information. Previous to this study, no three-dimensional structures existed for hydrophobins. The eight conserved cysteine residues had earlier been shown to form disulphide bridges in consecutive order[31,32]. This finding led to the conclusion that the three-dimensional structure of hydrophobins consists of four loop areas, two of which are predominantly composed of hydrophobic residues[28,33].

The formation of internal disulphide bridges also ruled out the possibility of intermolecular disulphide bridges being involved in the self-assembly process and their role was determined to be the stabilization of soluble hydrophobin molecules, thus preventing premature self-assembly in an aqueous environment[27]. By Fourier transform infrared (FTIR) and circular dichroism (CD) spectroscopy, it was shown for *S.*

commune hydrophobin SC3 that the secondary structure content of the protein changes during the self-assembly procedure. The water-soluble form was found to be rich in β-sheet structure and the β-sheet content increased after self-assembly at the air-water interface. An α-helical structure was induced specifically upon assembly on a hydrophobic solid[34].

The structural studies of hydrophobins were initiated with an NMR (nuclear magnetic resonance) -study of a class I hydrophobin EAS from *Neurospora crassa*[10]. This study concluded that hydrophobins are largely unstructured in solution, except for a small area of β-sheets stabilized by disulphide bridges. It was suggested that by encountering the hydrophobic-hydrophilic interface the protein undergoes transition from a disordered to an ordered state. Several examples of this type of intrinsically unfolded proteins exist; the protein or its segment is disordered in solution but folds in the presence of a substrate. However, the disordered regions are typically of a low sequence complexity with compositional bias and are predicted to be highly flexible[35].

2. AIMS OF THE STUDY

T. reesei is a rot fungus that produces three hydrophobins, called HFBI, HFBII and HFBIII. These are all class II hydrophobins. The aim of this study was to use X-ray crystallography to relate the characteristics of hydrohobins to their structure and thus improve our understanding of how these amphiphiles function. The specific aims were:

1) To crystallize and determine the structure of HFBI and HFBII.

2) To determine the structure of a variant form of HFBI, a covalent dimer named N-Cys HFBI.

3) To identify the amino acid residues responsible for the amphiphilic nature.

4) To determine the multimerization state in the crystal structure.

5) To identify the areas in the structure susceptible to conformational changes during self-assembly.

6) To observe any interactions that might assist in determining the functional mechanism of hydrophobins and explain how the amphiphilic layers are formed.

3. MATERIALS AND METHODS

The materials and methods used are all described in detail in the original publications. HFBI, HFBII and N-Cys HFBI were obtained from VTT Biotechnology as lyophilizised powder. The N-Cys HFBI-variant was constructed by adding an extension of 13 residues to the N-terminus of an HFBI molecule. The sequence of the extension was SCPATTTGSSPGP. The second residue in the extension was a cysteine with a free sulphhydryl group and two variant molecules were linked together through a covalent S-S-bond between the cysteines in their N-terminal extensions forming a covalent dimer. The molecular weights of the proteins were 7.5 kDa, 7.2 kDa, and 17.3 kDa for HFBI, HFBII and the N-Cys HFBI-variant, respectively. HFBI, HFBII, and the N-Cys HFBI were dissolved in pure water for crystallization to concentrations of 16 mg/ml, 8 mg/ml, and 6 mg/ml, respectively. The hanging-drop vapor-diffusion method was applied for crystallization at 293 K.

All the data were collected at the EMBL (European Molecular Biology Laboratory) Hamburg, located at the DORIS storage ring, DESY (Deutsches Elektronen-Synchrotron). The beamlines used were X11, BW7A, and BW7B. X11 and BW7A are equipped with a MAR CCD (charge coupled device) 165 mm detector and BW7B is equipped with a MAR IP (image plate) 345mm detector. BW7A is a tunable beamline, whereas X11 and BW7B operate at fixed wavelengths. The data collection temperature was 100 K. Data were either processed with the DENZO/SCALEPACK[36] programs or with the XDS[37] program.

The Xtalview[38] program was used to calculate the anomalous difference Patterson maps. The *ab initio* program ACORN[39] was used to solve the phases for the first crystal structure (HFBII) by direct methods and refined phases were then used in ARP/wARP[40] to build in the main chain of the protein molecules. The subsequent hydrophobin structures were solved by the molecular replacement method using the Molrep[41] program in the CCP4-suite[42]. The solvent content of the crystals and the Matthews coefficient were calculated according to Matthews[43]. A protein fold similarity search with the PRIDE program was used to find proteins with a similar fold[44].

The structures were refined using the CNS[45], Refmac 5[46] and SHELXL[47] programs. The O[48] program was used to inspect the electron density maps. The WHAT IF[49] and PROCHECK[50] programs were used to validate the structures. The Moleman[51] program was used to calculate the average B-factors. Hydrogen atoms, when present in the structure, were excluded from these calculations. The NCI (non-canonical interactions) -server[52] was used to search for 'weak' hydrogen bonds in the ultra-high resolution structure of HFBII. The crystal contacts were calculated with the EdPDB[53] program and the Contact program in the CCP4-suite. The solvent accessible areas (SAA) were calculated with the Areaimol[54] program in the CCP4-suite for each molecule in each structure as 1) an isolated molecule 2) a molecule in contact with the other molecules in the asymmetric unit and 3) a molecule in the crystal, i.e. in contact with the other

molecules in the asymmetric unit and the symmetry related molecules. Similarly, the SAAs of the hydrophobic patches were calculated as a summation of the SAAs of the side-chain atoms of the patch residues. The similarity of the hydrophobin molecules was analyzed by superimposing them pair-wise by C_α in the Shelxpro[47] program. This superimposition included residues 7-73 in the HFBI-sequence and 2-68 in the HFBII-sequence, to insure that the length of the match remains comparable despite different protein chain lengths.

4. RESULTS

4.1. CRYSTALLIZATION[I, II, III, IV]

Crystals of HFBII grew from two conditions: (1) 25 % polyethylene glycol (MW 2000), 0.4 M lithium sulphate and 0.1 M Tris at pH 8.5; and (2) 15% polyethylene glycol (MW 2000,) 0.2 M lithium sulphate, 0.1 M HEPES at pH 7.5. In the latter case, the crystallization drop contained manganese dichloride at a concentration of 50 mM and streak seeding was applied to improve the quality of the crystals. Crystals of HFBI grew from (3) a solution of 0.1 M zinc sulphate and 0.1 M sodium cacodylate pH 6.5. The detergent 1-S-octyl-β-D-thioglucoside (OSG) was used as an additive; its concentration in the drop was 9 mM. The crystals of the N-Cys HFBI-variant grew from (4) a crystallization solution of 0.1 M zinc acetate and 0.1 M HEPES, pH 7.0. The detergent lauryldimethylamine oxide (LDAO) was added to the crystallization drop; its concentration was 2 mM. The crystals used for data collection are shown in Figure 3. The structures of the detergents used in the crystallization are presented in Figure 4.

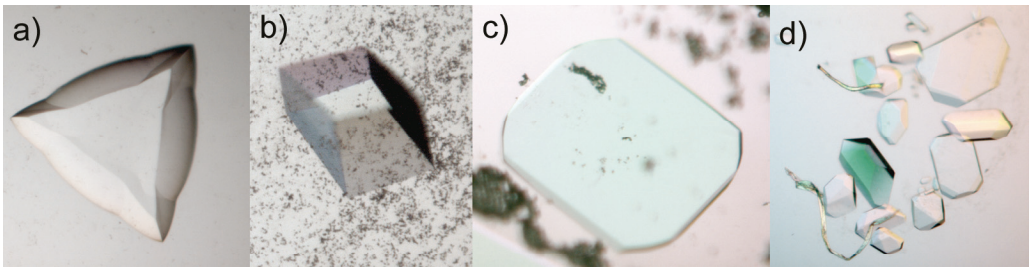


Figure 3. Crystals used for data collection: a) HFBII, b) HFBII in presence of manganese ions, c) HFBI and d) N-Cys HFBI-variant

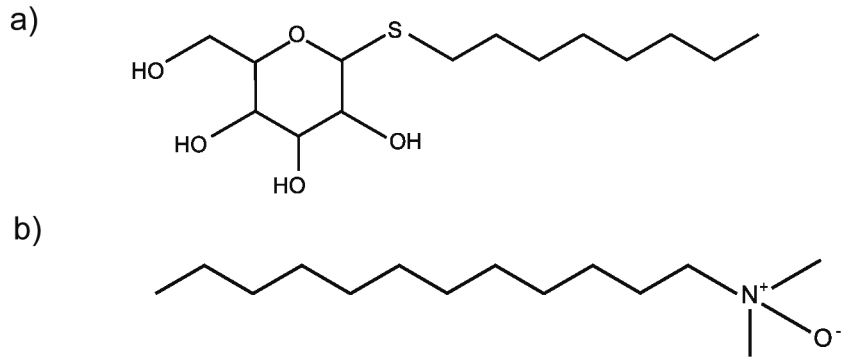


Figure 4. The detergents used in the crystallization. a) OSG b) LDAO

4.2. DATA COLLECTION[I, II, III, IV]

A summary of the data collection statistics is presented in Table 1. For HFBII, crystallized in the presence of manganese ions, the data were collected with the home source at 1.5 Å resolution, at the EMBL beamline X11 at 1.0 Å resolution and at the EMBL beamline BW7A at 0.75 Å resolution. Of these, the statistics of the highest resolution data set are presented in Table 1.

4.3. STRUCTURE SOLUTION[I, II, III, IV]

The monoclinic structure of HFBII at 1.0 Å was the first hydrophobin crystal structure determined. No structure for a homologous protein was available, due to the low sequence similarity and attempts to produce a heavy atom derivative for an isomorphous replacement failed. An atomic resolution data set of the monoclinic form of HFBII had been collected at a synchrotron source at wavelength 0.81 Å, and a 1.5 Å resolution data set using the copper rotating anode in the home laboratory. The anomalous difference Patterson maps, calculated from the home laboratory data showed a clear signal caused by one manganese ion coordinated to the two protein molecules of the asymmetric unit. The signal of the manganese ion was not present in the Patterson maps of the synchrotron data, due to the wavelength used; the absorption edge of manganese is at 1.8961 Å. The atomic resolution data together with the coordinates of manganese were the inputs into the ACORN program, which uses a direct *ab initio* -method to calculate the phases. Two HFBII molecules and one manganese ion were located in the unit cell. The protein main chain was built into the density automatically and the side-chains were mutated manually to match the protein sequence. The structures of the cubic crystal form of HFBII and the native HFBI were solved by molecular replacement, by using a molecule of the monoclinic HFBII structure as a search model. A molecule from the native HFBI structure was used as a search model to solve the structure of the N-Cys HFBI-variant.

Table 1. The data collection statistics. Numbers in parenthesis refer to the highest resolution shell.

Protein	HFBII	HFBII	HFBI	N-Cys
Cryst. condition	1	2	3	4
Unit cell parameters				
a (Å)	72.2	78.5	108.9	92.4
b (Å)	72.2	46.3	49.6	120.2
c (Å)	72.2	34.7	85.7	119.7
$\alpha = \gamma$ (°)	90.0	90.0	90.0	90.0
β (°)	90.0	111.6	129.4	90.0
Space group	I23	C2	C2	$C222_1$
Beamline	X11	BW7A	BW7B	BW7B
Wavelength (Å)	0.8122	0.7747	0.8430	0.8423
Processing program	DENZO	DENZO	XDS	XDS
Resolution range (Å)	25-3.25 (3.31-3.25)	20-0.75 (0.76-0.75)	20-2.1 (2.3-2.1)	20-2.9 (3.0-2.9)
No. observations	22261	433130	78121	55098
No. unique reflections	1074 (48)	132282 (4133)	20790 (4909)	14303 (1374)
Completeness (%)	99.6 (100)	98.7 (91.3)	99.0 (98.7)	94.6 (96.5)
R_{sym} (%)	4.6 (29.7)	7.2 (58.7)	4.6 (27.5)	7.7 (35.5)
$I / \sigma(I)$	13.1 (8.9)	18.0 (1.9)	22.8 (6.4)	15.4 (4.4)
No. molecules/au	1	2	4	4
V_M ($Å^3$/Da)	2.03	1.92	2.89	4.80
Solvent content (%)	39	37	59	74

4.4. REFINEMENT[II, III, IV]

A summary of the refinement statistics for each data set is presented in Table 2. A test set, including 5% of the reflections, was used to calculate R_{free}. Due to the modest resolution of data, the data of the cubic HFBII crystal form was not refined. The refinement statistics for the unrestrained refinements of atomic (1.0 Å) and ultra-high (0.75 Å) resolution data of the monoclinic crystal form of HFBII are both presented, for comparison. The atomic resolution structure of HFBII deposited at the Brookhaven Protein Data Bank (PDB) with code 1R2M was refined with Refmac 5[II]. This atomic resolution structure was later refined with SHELXL to facilitate the comparison of atomic and ultra-high resolution structures[III]. The values presented here refer to those after the SHELXL-refinement. In the SHELXL-refinements for HFBII, the anisotropic displacement parameters were refined and the hydrogen atoms were included in the models. Towards the end of the refinements, the restraints were gradually released. For HFBI and the N-Cys HFBI-variant, a conventional isotropic refinement was used. In addition, the twin operator -h,-k,h+l was included in the refinement of the native HFBI data and the twin fraction was also refined by using the BASF instruction.

Table 2. The refinement statistics.

Protein	HFBII	HFBII	HFBI	N-Cys
Space group	C2	C2	C2	$C222_1$
Max. resolution (Å)	1.0	0.75	2.1	2.9
Phases	Acorn	Rigid refinement	Molrep	Molrep
Refinement	SHELXL	SHELXL	SHELXL	CNS
R (%)	11.2	13.0	22.4	23.2
R_{free} (%)	13.3	14.8	27.4	25.4
RMSD bond (Å)	0.029	0.027	0.006	0.007
RMSD angle (°)	2.5	2.5	2.1	1.4
No. atoms	1243	1280	2084	2123
No. of protein atoms	1025	1023	1972	1995
No. of water molecules	217	256	108	14
PDB-code	1R2M	2B97	2FZ6	2FZ7
Residues in most favored region (%)	89.7	88.8	76.5	84.1
Average B-factor, all atoms ($Å^2$)	15.0	11.5	44.9	43.0

4.5. THE CRYSTAL STRUCTURE OF HFBII[II, III]

The structure of a hydrophobin HFBII molecule consists of four anti-parallel β-strands and an α-helix (Figure 5). Strands 1 and 2 constitute a β-hairpin motif, as do the strands 3 and 4. The two β-hairpin motifs arrange to form a small β-barrel and the α-helix resides on one side of the barrel. The fold of HFBII appears to be novel, as was also shown by a protein fold similarity search. The structure consists of a single domain with four disulphide bridges. The bridges form between Cys3 and Cys53, Cys13 and Cys43, Cys14 and Cys26 and Cys52 and Cys64. The first bridge connects the N-terminal loop to the β-barrel, the second bridge connects the α-helix to the β-barrel and the remaining two bridges hold the strands in the β-hairpin structures together. The disulphide bridges in the β-hairpins are located inside the barrel structure in a symmetrical manner.

Most importantly, a hydrophobic patch is located on the surface of HFBII molecules, consisting of side-chains of hydrophobic, aliphatic amino acid residues and giving rise to the amphiphilic character of the molecule. The patch mostly consists of the residues in the loop areas of the two β-hairpin motifs. The residues involved are Leu7, Val18, Leu19, Leu21, Ile22, Val24, Val54, Ala55, Val57, Ala58, Ala61, Leu62 and Leu63. Leu7 comes from the N-terminal loop, Val18-Val24 belong to the first β-hairpin motif and Val54-Leu63 to the second β-hairpin motif. Val18 is the last residue of the first β-strand, Ile22 and Val24 belong to the second β-strand, and Ala61, Leu62, and Leu63 are residues of the fourth β-strand. The size of the hydrophobic patch is about 18% of the total surface area and its shape is relatively flat.

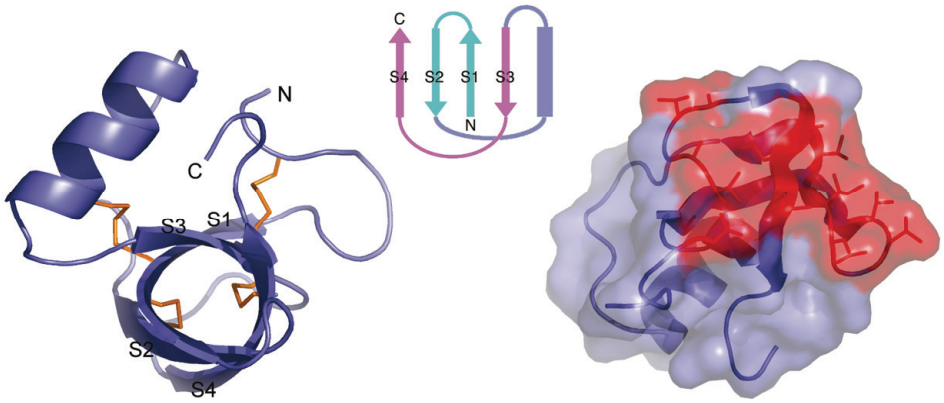


Figure 5. The structure of an HFBII molecule. Disulphide bridges are indicated in orange. In the topology diagram, the β-strands are indicated by arrows and the α-helix by a rectangle. In the surface model, the residues of the hydrophobic patch are indicated in red.[55]

In the asymmetric unit of the HFBII structure, at ultra-high resolution, there were two HFBII molecules (designated A and B), one manganese ion and 256 molecules of water. The amount of detected water molecules is about 70% of the water molecules possible in the asymmetric unit. In the atomic resolution structure, 217 water molecules were detected in the asymmetric unit, which is about 59% of the possible water molecules. The hydrophobic patches of the two HFBII molecules in the asymmetric unit pack towards each other and the molecules form a dimer that partially conceals the hydrophobic patches of each molecule from the solvent (Figure 6). The manganese ion was coordinated to aspartic acid residues (Asp25 in molecule A, as well as Asp34 and

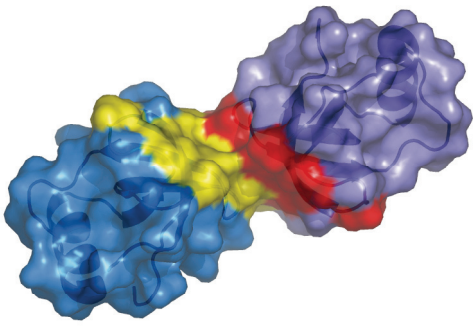


Figure 6. The hydrophobin HFBII dimer, hydrophobic patches in red and yellow.[55]

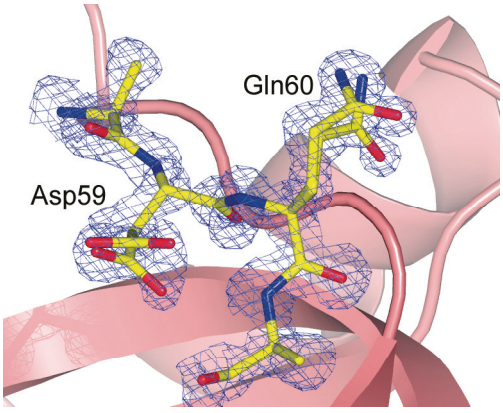


Figure 7. Residues Asp59 and Gln60 in molecule B fitted in two conformations. The ($2F_O$-F_C) electron density map is contoured at 1.0σ.[55]

Asp25 in symmetry related molecules A and B, respectively) and a glutamine residue (Gln60 in molecule A) on the protein surface. The hexacoordinated metal ion was in contact with three HFBII molecules and two water molecules.The electron density maps were of excellent quality, owing to the high resolution of the data. Both the C- and the N-termini were, however, found to be flexible and the electron density for the C-terminal Phe71 was missing entirely. A lack of Phe71 was also observed for about one half of the protein material prior to crystallization in high resolution ESI-FTICR MS (electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry) measurements (data not shown). This opens way to the possibility that the crystallized form of HFBII did not contain Phe71 in the C-terminal to begin with.

Both a discrete and unresolved disorder was present in the structure, with respect to amino acid side-chains on the protein surface. Residues Thr16, Gln40, Lys46, Leu62, and Lys66 in molecule A, and residues Leu12, Thr16, Asp20, Asp59, Gln60, Leu62, and Lys66 in molecule B were fitted into two conformations in the ultra-high resolution structure. Seven of these residues were rather distinctly observed in two positions (Figure 7). For the residues with unresolved disorder, namely Lys66 in molecule A and Asp20, Gln40, Lys46 and Lys66 in molecule B, even the dual conformation was inadequate to describe the situation in the crystal. Most of these residues were located in the vicinity of the C-terminus and might have been allowed more degrees of freedom in the absence of the C-terminal Phe71. In the atomic resolution structure, residues Thr16, Leu19, Asp20, Val33, Lys49, Leu62, and Lys66 in molecule A and residues Thr16, Asp20, Val57, Gln60, Leu62, and Lys66 in molecule B were fitted in two conformations. For residues Asp20, Lys49 and Lys66 in molecule A and Asp20, Leu62 and Lys66 in molecule B, the electron density was ambiguous, indicating unresolved disorder. None of the disorder observed in these structures could be directly related to functionality, even though Leu19, Val57, and Leu62 are residues of the hydrophobic patch. It is surprising that more residues could be fitted in dual conformation in the atomic resolution structure than in the ultra-high resolution structure. However, the two data sets of HFBII were collected from two different crystals and so the difference in dual conformations is likely due to small variations in the crystals.

The solvent content of the HFBII crystals was rather low. Therefore, there are many contacts between adjacent molecules in the crystal: each molecule A made symmetry contacts with eight symmetry-related molecules and each molecule B made contacts with seven symmetry related molecules. The crystal contacts involved interactions such as salt bridges, hydrogen bonds, coordination through the manganese ion, hydrophobic interactions, and van der Waals contacts. The resolution of the data was excellent, therefore, the determined structures were very accurate and reliable. Even with the restraints released, the R-factors and the geometric parameters remained very good. The small deviations in these structures, in comparison to values from small molecules, are thus likely to be real deviations due to the environment of a specific residue. In addition, especially in the ultra-high resolution data, the B-factors were very low.

4.6. THE CRYSTAL STRUCTURE OF NATIVE HFBI[IV]

The overall structure of an HFBI molecule is very similar to HFBII (Figures 5,8). There are four molecules (designated A-D) present in the asymmetric unit of the crystal structure, two of which, molecules A and C, have a similar conformation to that of HFBII. However, molecules B and D possess a structural change in the region of the second β-hairpin motif. There, an area of seven residues, from Ala60 to Ala66, has moved to a more extended conformation and bent away from the β-barrel structure. The difference in the position of the main chain C_α-atom was at most about 10 Å, when comparing the locations of Ala63 in the superimposed molecules C and D. The change in the conformation also affected the secondary structure content of molecules B and D, by shortening the fourth β-strand from eight to three residues. Thus, β-sheet content of these molecules decreased from 37% (as observed in A and C) to 31%. Two disulphide bridges (Cys58-Cys69 and Cys8-Cys57) delimited the area of the structural change and so the core structure of the molecules remained unaltered. The disulphide bridge array for all the HFBI molecules was also the same as in the HFBII structure.

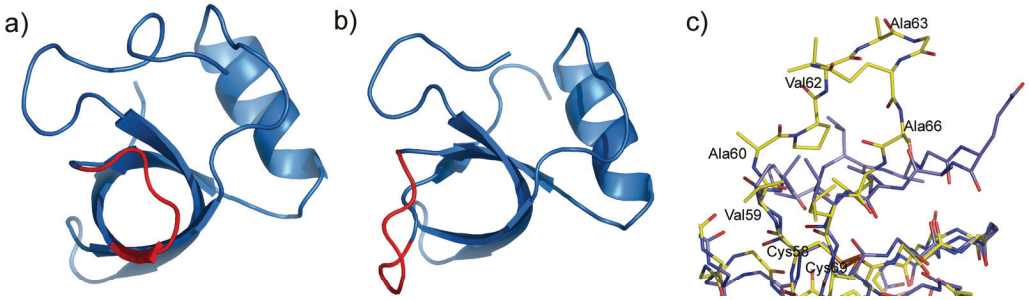


Figure 8. The structures of HFBI, a) molecule A and b) molecule B. The area of the structural change is highlighted in red. c) Magnification of the area of the structural change in molecule B (yellow) with molecule A superimposed (blue).[55]

The hydrophobic patch area on the protein surface, as recognized for HFBII, was also found on the surface of the HFBI molecules. The residues forming the hydrophobic patch were Leu12, Val23, Leu24, Leu26, Ile27, Leu29, Val59, Ala60, Val62, Ala63, Ala66, Leu67, and Leu68. When the conserved cysteines were aligned, the residues of the hydrophobic patch were exactly the same in these two hydrophobin structures, excluding Leu29 in the HFBI structure, which corresponded to a valine (Val24) in the HFBII structure (Figure 9). The structural change in the second β-hairpin motif in molecules B and D of the HFBI structure also affected the shape of the hydrophobic patches of these molecules, since some of the patch residues were located in the moving area, i.e. Ala60, Val62, Ala63 and Ala66. The size of the hydrophobic patch remained about the same, but its shape was altered in that it was slightly curved.

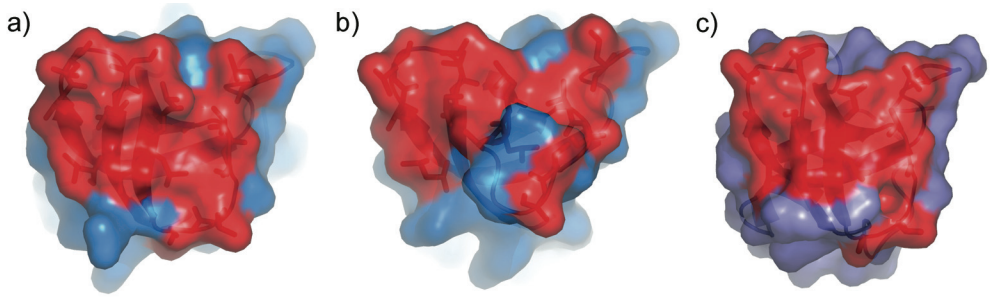


Figure 9. The surface models with areas of the hydrophobic patch highlighted in red. a) HFBI, molecule A b) HFBI, molecule B c) HFBII.[55]

In addition to the four protein molecules, the asymmetric unit contained four zinc ions and 108 water molecules. No detergent molecules could be located from the electron density maps, even though the addition of the OSG detergent was essential in order to produce crystals of an adequate diffraction quality. The four protein molecules formed a tetrameric structure (Figure 10) and two of the zinc ions were coordinated to the site identical to that of the manganese in the HFBII structure. The residues involved in coordination were aspartic acid residues (Asp30); one coordination site was located between molecules A and B and an equivalent site between molecules C and D. The two additional zinc coordination sites were in the helical regions, between molecules A and C_{sym} and between molecules B and D_{sym}. These ions were also coordinated to aspartic acid residues (Asp40 and Asp43). The corresponding metal ion coordination is not possible for HFBII molecules, since the corresponding residues are a threonine and an isoleucine.

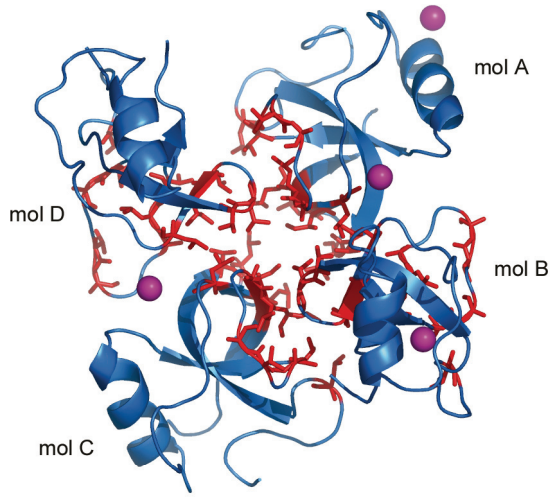


Figure 10. The tetrameric structure of HFBI. The zinc ions are indicated as magenta spheres and the residues of the hydrophobic patch with a stick-model in red.[55]

The electron density maps were of good quality, excluding small areas in the helical regions of molecules B and D. Here, the electron density was weak and ambiguous for residues Lys32, Pro34, Ans46, Val46 and Cys48 in molecule B and for residues Lys32, Ser35, Asp40, Phe44 and Arg45 in molecule D. In addition, in the HFBI structure, the termini for all molecules were flexible and disordered. The sequence of HFBI is five residues longer from the N-terminus than that of HFBII when the conserved cysteines are aligned (Figure 1). The N-teminal of HFBI has also been previously shown to be susceptible to deamination and cleavage[25]. In the crystal structure, a different number of residues for each molecule could be fitted into the electron density at the termini. Therefore, molecule A consisted of residues Asn2-Val73, molecule B of residues Val7-Ala75, molecule C of residues Gly5-Ala75 and molecule D of residues Asn6-Gly74.

Crystal contacts were identified between molecules A and C and between molecules B and D. In addition, molecules B and D had contacts to symmetry related molecules B and D, respectively. The packing interactions mostly occurred through the helical regions of the molecules. The interaction types were hydrogen bonds, van der Waals interactions, and coordination through the zinc ions. The R-factors and the B-factors remained relatively high throughout the refinement, likely due to the pseudo-merohedral twinning present in the crystal. The usage of twinned data is problematic, especially when the twinning is perfect, as was the case with the HFBI data. However, the electron density maps were of good quality, excluding some small areas, and the structure was proved plausible by the validation programs.

4.7. THE CRYSTAL STRUCTURE OF N-CYS HFBI-VARIANT[IV]

In the crystal structure of the N-Cys HFBI-variant, four HFBI molecules could be located in the asymmetric unit. The general structure of these molecules was similar to that of the HFBII molecules and molecules A and C of the HFBI structure. No signs of structural change, as seen in molecules B and D of the HFBI structure, were seen here. The N-terminal extensions that connect two molecules into a covalent dimer could not be located in the electron density maps. In fact, the N-termini of this structure were flexible and the protein chain of each molecule starts at the Asn6 residue of the HFBI sequence. It is possible that the long and flexible extension is so disordered that it cannot be distinguished in the electron density maps. It is more likely, however, that the absence of the extension and some N-terminal residues is due to the crystallized form being fragmented N-Cys HFBI, lacking the covalent bond and being cleaved from the N-terminus. The presence of fragments of varying length and also the presence of native HFBI in the protein material was determined by high resolution ESI-FTICR MS measurements prior to crystallization (data not shown). However, the estimation of the quantity of each component was difficult.

The hydrophobic patch was located at the protein surface, identical to that of the native HFBI. In addition to the four HFBI molecules (designated A-D), there were seven LDAO detergent molecules, two zinc ions and 14 water molecules in the asymmetric

unit. The four molecules of HFBI were all organized in the same plane. Two of the LDAO molecules were closed in to the space in the midst of the HFBI molecules. The rest of the detergent molecules were packed close together on one side of the HFBI-bundle (Figure 11). The hydrophobic tails of the detergent molecules were packed towards the hydrophobic surface patches of the hydrophobin molecules. The zinc ions were coordinated to aspartic acid residues, identical to the position described for the helical region of the HFBI structure. The additional coordination site identified in the native HFBI structure was occupied by a water molecule in the N-Cys HFBI structure. Due to the modest resolution of the data, only a few water molecules could be included in the refinement.

The electron density maps were of good quality throughout, in spite of the modest resolution of the data and the high solvent content. For the detergent molecules, the density was slightly ambiguous; the extended carbon chain could be easily observed for each molecule but the density for the charged head group was not always clearly visible. This could be due to some of the LDAO molecules having two conformations,

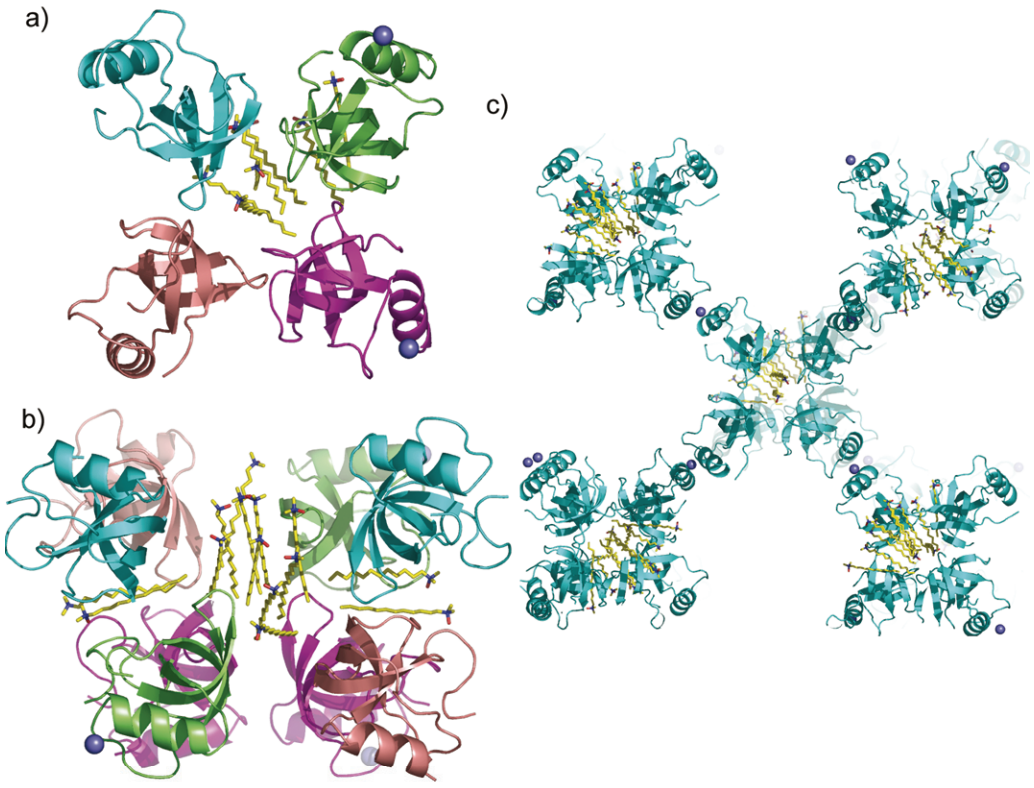


Figure 11. The structure of the N-Cys HFBI-variant. The zinc ions are indicated as blue spheres and the detergents in yellow. a) Contents of the asymmetric unit. b) The quaternary structure. c) The arrangement of the molecules in the crystal.[55]

i.e. the position of the molecule remains the same but the chain runs in the opposite direction. Slight disorder was also observed at the C-termini of all protein molecules but, nevertheless, all the C-terminal residues could be fitted into the density. Some additional density was visible at the N-termini but not sufficiently so to add residues preceding Ans6 to the protein chain. This indicated that the crystal might be composed of molecules with a varying number of residues in the N-terminus. For a couple of residues, e.g. Gln54 and Gln65, on the protein surface, hints of dual conformations were detected but left unmodelled because of the modest resolution of the data. Due to this modest resolution, determining the correct conformation for some of the side-chains was challenging.

The organization of the molecules in this crystal with exceptionally high solvent content was unique. If the interactions between the HFBI molecules only were considered, the protein was in a monomeric state. However, the quaternary structure was clearly a detergent-mediated octamer, where a cluster of 14 LDAO detergent molecules was buried between eight HFBI molecules, four on each side (Figure 11b). Additional crystal packing interactions were formed through the zinc ions, located between molecules A and D_{sym} and molecules B and C_{sym}(Figure 11c). The clever organization of the molecules managed to bury the hydrophobic surface areas of the protein and still leave room for large solvent pores. In spite of the modest resolution and the exceptionally high solvent content of the crystals, the R-factors and the B-factors were quite good and the electron density maps clear and unambiguous. The restraints used were quite strong because of the modest resolution, and so the validation reports presented no problems.

5. DISCUSSION

5.1. CRYSTALLIZATION OF HYDROPHOBINS[I, II, III, IV]

The presence of metal ions has been crucial for the crystallization of all the current hydrophobin structures. In spite of the aggregation tendency of hydrophobins, HFBII crystallized from various conditions with variant pHs and precipitants. However, the crystals diffracted X-rays extremely poorly. With the addition of manganese ions to the crystallization solution, the diffraction power of the crystals was dramatically enhanced. HFBI crystallized using either isopropanol or zinc sulphate as a precipitant. The crystals from isopropanol grew for months, whereas the crystals from zinc sulphate formed immediately after pipeting. However, both crystal forms diffracted X-rays very poorly. The addition of the OSG detergent slowed down the growth rate of the crystals produced with zinc sulphate and improved the diffraction. The use of detergents also proved important, since the N-Cys HFBI-variant crystallized only in the presence of both zinc ions and detergent. Crystals could be produced with several different detergents but only with LDAO was the diffraction quality satisfactory. HFBII has also been crystallized in the presence of a detergent (our unpublished data).

Judging by the structures, the importance of the metal cations is evident: the ions mediate contacts between the residues on the protein surface, which in the absence of the metal would have repulsive interactions. Also, in the case of the N-Cys HFBI structure, the role of the detergent molecules is clear; the octameric structure could not have formed without the detergent molecules acting as mediators for the hydrophobic interaction. The high affinity of HFBII and especially of HFBI towards non-ionic surfactants has been previously described [29].

A different crystallization condition and preliminary X-ray characterization[56] was reported for HFBI during this study. The precipitants and the buffer, as well as its pH, differed from those used in this study. The symmetry of the crystals reported by Askolin and co-workers was hexagonal and they had used a detergent, CYMAL-5 (cyclohexyl-pentyl-β-D-maltoside), for crystallization. The structure appeared to contain four to six molecules in the asymmetric unit; with four molecules in the asymmetric unit the solvent content would be about 60%, which would resemble the situation in the monoclinic crystals presented in this study.

5.2. PSEUDO-MEROHEDRAL TWINNING IN HFBI CRYSTALS[IV]

By definition, twinning occurs when a crystalline specimen consists of multiple domains that are mutually reoriented according to a specific transformation that does not belong to the symmetry operations of the crystal point group but is related in some way to the crystal lattice[57]. The specific symmetry operation that relates the domains of a twinned crystal is called the twin law. The diffraction pattern of each domain may be related to the next domain by inversion, a mirror plane or rotation, of which only rotation is possible in protein crystallography.

Each domain in a twinned crystal contributes to the diffraction of X-rays and the diffraction pattern is a superimposition of all these contributions. The reflections in a twinned diffraction pattern should also be weighted according to the volume of each domain; this is termed the twin fraction. Twins that are composed of two twin domains are referred to as hemihedral twins. When the twin fraction of hemihedral twins is 0.5 or very close, the twinning is termed 'perfect'.

Sometimes twinning can be observed by visual inspection of the crystal under a microscope or as additional spots in the diffraction pattern, where certain zones can be distinguished by extra reflections. In this case, the twinning is termed non-merohedral and the reflections either ovelap completely, overlap partially or do not overlap at all. The point group of the crystal lattices $\bar{1}$ in triclinic, 2/*m* in monoclinic, *mmm* in orthorhombic, 4/*mmm* in tetragonal, $\bar{3}m$ in rhombohedral, 6/*mmm* in hexagonal and $m\bar{3}m$ in cubic are referred to as the holohedral point groups[58]. A point group, which belongs to the same crystal system with the holohedral point group and is a subgroup of a holohedral point group is termed a merohedral point group. When the twin law is a symmetry operation of the crystal lattice but not of the point group symmetry, the

twinning is called twinning by merohedry or merohedral twinning. This is possible in tetragonal, trigonal, hexagonal and cubic crystal systems[59]. Merohedral twinning cannot be identified from the diffraction pattern, since the reflections of the twin domains overlap completely. Two special cases exist related to merohedral twinning: pseudo-merohedral twinning and reticular merohedral twinning. In pseudo-merohedral twinning, the parameters of the crystal have some special relationship to those of a crystal of a higher symmetry crystal system. As with merohedral twinning, perfect pseudo-merohedral twinning cannot be identified by inspecting the diffraction pattern. Reticular merohedry is a special case of twinning in rhombohedral crystals, where twinning through twofold axes parallel to threefold axis exists, even though the rhombohedral lattice holohedry is $\bar{3}m$.

Twinned data can be treated and there are several examples of solved structures for each type of twinning[59]. Warning signals of twinning are listed, for example by Herbst-Irmer and Sheldrick[60] and there are several ways to detect twinning, even when it is merohedral[59,61]. Estimation of twin fraction and determination of the twin law are the key parameters after detecting that the data is twinned. Data can be detwinned if the twin factor is not too close to 0.5; the twin fraction can be estimated, for example, by using the Detwin[62] program in the CCP4-suite given that the twin law has been determined. Twinned data, detwinned or not, may sometimes be successfully used for phase determination by molecular replacement (MR), multiple isomorphous replacement (MIR), multiple wavelength anomalous dispersion (MAD) and even single wavelength anomalous dispersion (SAD)[59]. For refinement, twinning can be accounted for, e.g. in CNS[45]and SHELXL[47].

In the case of the native HFBI data, no clear indications of non-merohedral twinning could be found from the diffraction images. The data processed very well in space group F222 with unit cell parameters of a = 49.60 Å, b = 108.94 Å and c = 132.55 Å (R_{meas} and I/σ(I) in the last resolution shell were 28.7% and 8.22, respectively). However, after the structure determination with two molecules in the asymmetric unit, the R-factors remained high (R ~ 30%, R_{free} ~ 40%) after refinement and the electron density about the helical region of one molecule in the asymmetric unit was ambiguous and poor. The closest space group solution next to F222 was C2, which suggested that pseudo-merohedral twinning should be inspected.

Pseudo-merohedral twinning where a monoclinic crystal mimics an orthorhombic one may arise when the β angle in a monoclinic cell is close to 90°. The same happens when the two unit cells of the twin domains share the same a and b axes, but in opposite directions[63], i.e. they satisfy the condition c cosβ = -a / 2. This rule is true for HFBI and the twinning was also confirmed by a twin test with Detwin in the CCP4 suite. Detwin also estimated the twin fraction of HFBI data to be close to 0.5 (Figure 12). A similar kind of pseudo-merohedral twinning in monoclinic space group had been previously detected for γδ T-cell ligand T10[63] (true space group $P2_1$, apparent C222) and for acetyl coenzyme A synthetase[64] (true space group C2, apparent F222).

Thus, the twin law for HFBI data was assumed to be *-h*, *-k*, *h+l*, as had earlier been determined for acetyl coenzyme A synthetase.

The introduction of the twin law to the refinement lowered the R-factors considerably (R = 23%, R_{free} = 28%) and improved the electron density maps significantly. The twin fraction was refined using the BASF-instruction and its value at the end of the refinement was 0.49, and so the twinning in the HFBI data could be termed perfect. Perfect twinning was expected, since the data processed so well in the higher symmetry space group. Due to perfect twinning, detwinning of the data was not possible, so it had to be used as it was, together with the twin operator. At the end of the refinement, the R-factors and the B-factors were slightly higher than would be expected for a data set of this resolution. In addition, some of the electron density remained ambiguous. Most likely, this imprecision in the data was caused by twinning but the twinned data could, nevertheless, be used for successful and informative structure determination.

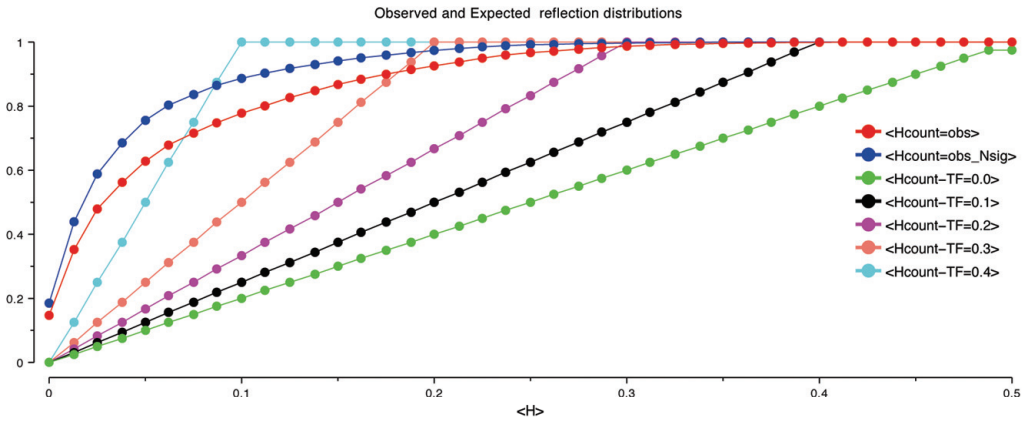


Figure 12. Estimation of the twin fraction with the Detwin program. Theoretically expected values for twin fractions of 0.0 – 0.4 indicated in green, black, purple, pink, and light blue; the values observed for HFBI data in red and blue.

5.3. THE ULTRA-HIGH RESOLUTION STRUCTURE OF HFBII[III]

Most of the deposited protein structures in the Brookhaven Protein Data Bank[65] (PDB) have been determined to a crystallographic resolution of 1.5-2.5 Å. To date, there are altogether approximately 27000 protein structures determined by X-ray crystallography deposited in the PDB. About 570 structures in the PDB qualify as atomic resolution structures, i.e. they have been determined to a resolution of 1.2 Å or better. However, ultra-high resolution structures that exceed the limit of 0.80 Å resolution are even rarer. Altogether, there are 16 crystal structures in the PDB that meet this condition, 11 of which are protein structures[66-75]. One of these protein structures at ultra-high resolution is the structure of HFBII, described in this study.

The world record for protein structure resolution is held by crambin[66] at 0.54 Å resolution. Usually, proteins that diffract to ultra-high resolution are small in size; the crystals have low solvent content (27-45% in currently determined ultra-high resolution structures) and they are of a low-symmetry space group with only one molecule in the asymmetric unit. However, exceptions to these criteria can be found, e.g. aldose reductase[68] is a fairly large protein of 316 amino acid residues and HFBII is a dimer with two molecules in the asymmetric unit.

When the resolution of X-ray data is close to the bonding distance of atoms in the target molecule, the electron density maps start to show features that cannot be seen in the maps of protein structures calculated at a more modest resolution range. These additional features are density peaks caused by hydrogen atoms and deformation density caused by the expansion or contraction of the electron cloud due to bonding. These features can be exploited to give new, biologically relevant information on the target structure. Some advantages of atomic resolution data include visualization of hydrogen atoms, identifications of atoms by atom type, description of bonding features, accurate distances, and the modeling of the mobility of atoms through anisotropic displacement parameters[76,77]. Furthermore, ultra-high resolution data enables the study of the electronic properties and charge distribution directly from experimental data[67,78].

To make good use of the special features of ultra-high and atomic resolution data, special treatments in the refinement of the data are required. Conventionally, protein structures are refined under the assumption that all the atoms are spherical in shape, the bond angles and the bond distances always agree with those derived from small molecules and the hydrogen atoms are not included in the structure. This is done because the number of unique reflections in the data is too low to include more parameters to the refinement and the results are satisfactory in spite of the approximations. If the resolution of the data permits, however, hydrogen atoms may be included and anisotropic displacement parameters used in the refinement. The structural model may also be allowed to refine freely, i.e. with the restraints released, which insures that characteristic deviations from standards are not forced to fit the small molecule values. Finally, the non-sphericity of the atoms can be approached by the DBE (dummy bond electron) -method or by modeling the valence electrons. The DBE-approach[79] is included in the Phenix-suite and literally places a dummy electron in between the bonded atoms to compensate for the deformation part of the density. In order to accurately model the valence shell structure, the multipolar refinement program MOPRO[80] may be used.

For the HFBII structure, the ultra-high resolution data allowed very detailed and precise analysis of the structure. The distinction of atoms by atom type according to density was evident for both atomic and ultra-high resolution structures of HFBII, as shown for the ultra-high resolution structure in Figure 13. At ultra-high resolution, about 60% of the hydrogen atoms present in the protein structure could be experimentally identified. In the atomic resolution structure, about 53% of the hydrogens were identified. The lower the B-factor of the atom the hydrogen was

attached to the more likely it was to be experimentally detected (Figure 14). Experimental identification of hydrogen atoms assists in the determination of the protonation states of histidines, aspartic acids and glutamic acids, as well as the C-termini. Furthermore, the existence of a hydrogen bond may be verified. In addition to classical hydrogen bonds, a total of 47 weak hydrogen bonds between the $C_{\alpha}H$ and OC groups were identified in the two molecules of the HFBII structure, when the structure was searched with the NCI-server. In 41 cases out of 47, the hydrogen atom was experimentally observed in the donor group.

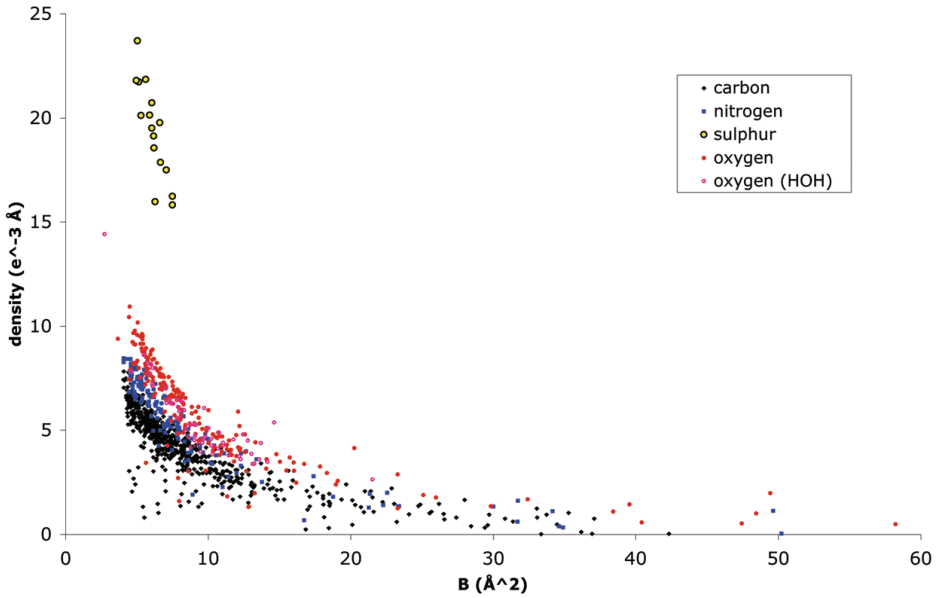


Figure 13. The electron densities at the atomic positions in the ultra-high resolution structure of HFBII plotted as a function of the B-factors. Figure courtesy of Dr Victor Lamzin.[81]

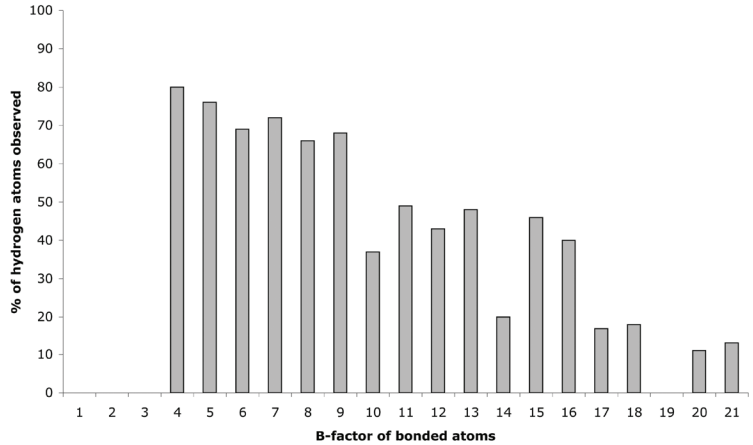


Figure 14. The frequency of the observed hydrogen atoms as a function of the B-factor of the atom that the hydrogen is bonded to.

The most important findings in the analysis of the ultra-high resolution structure of HFBII were that: 1) residual density in the electron density maps was present around the peptide bonds, the carbonyl oxygens, the manganese ion, and the sulphur atoms in the disulphide bridges in the ultra-high resolution structure (Figure 15). In the structure determined to atomic resolution, residual density in the peptide bond could only be observed occasionally. 2) The B-factors were considerably lower for the ultra-high resolution structure as compared to the atomic resolution structure (Table 3). These two findings indicated[79] that the ultra-high resolution data should be refined with the multipole refinement program MOPRO, whereas the atomic resolution data should not. This refinement could give direct experimental information, for example about the oxidation state of the manganese ion coordinated to the HFBII structure. MOPRO refinement of HFBII against the ultra-high resolution data was initiated during this study.

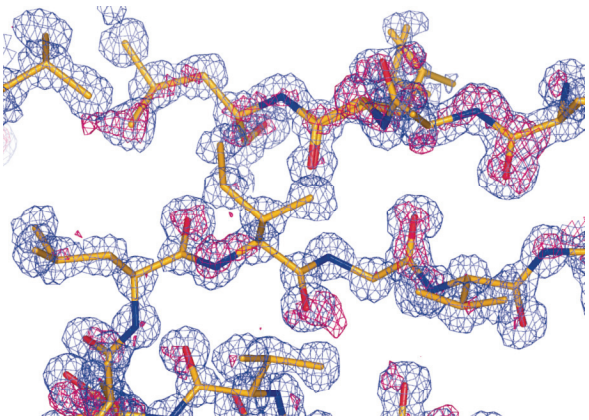


Figure 15. Residual density around the peptide bonds and carbonyl oxygens in the electron density maps of the ultra-high resolution structure of HFBII. The $(2F_O\text{-}F_C)$ map is contoured at 1.5σ (blue) and the $(F_O\text{-}F_C)$ map at 2.5σ (pink).[55]

Table 3. B-factors (Å^2) of the ultra-high resolution and the atomic resolution structures of HFBII after unrestrained refinement.

	Atomic	Ultra-high
B (main chain)	12.4	8.3
B (side chain)	15.1	10.8
B (all protein)	13.6	9.4
B (solvent)	21.8	19.8
B (Mn)	8.7	4.5
B (all atoms)	15.0	11.5

5.4. THE HYDROPHOBIN FOLD

All the hydrophobin structures determined in this study bore a remarkable similarity to each other. The only divergence was in the second β-hairpin motif of molecules B and D in the native HFBI structure. The movement here was most likely driven by tetramer formation, e.g. it seemed that if molecules B and D did retain the same conformation as molecules A and C, Gln65 would have collided with a residue of an adjacent molecule. It is unlikely that the movement in this area was affected by crystal contacts, since the area was facing a solvent channel. However, in the crystal structures of HFBII and N-Cys HFBI, this kind of movement was not possible due to crystal contacts. There is no tetrameric structure determined for HFBII, but it could be that at higher concentrations a similar tetramer could also be formed from HFBII molecules and that the corresponding movement in the second β-hairpin motif could occur.

The two molecules in the HFBII structure superimposed by C_α with an RMSD (root-mean-square distance) of 0.47 Å. Molecules A and C of the HFBI structure superimposed with an RMSD of 0.45 Å and molecule B and D with an RMSD of 1.00 Å. Superimposing any two of the HFBI molecules in the N-Cys HFBI structure produced on average an RMSD of 0.35 Å. Superimposing molecules with similar conformations from different structures produced slightly higher RMSDs, i.e. an HFBII molecule superimposed on molecules A or C from the native HFBI structure produced on average an RMSD of 0.59 Å; HFBII molecule superimposed on an HFBI molecule from the N-Cys HFBI structure produced on average an RMSD of 0.62 Å; and superimposing HFBI molecules from the N-Cys structure on molecules A or C from the native structure produced on average an RMSD of 0.47 Å. Superimposing molecules B or D from the native HFBI structure (i.e. the molecules with the structural change) on any other molecule always produced an RMSD above 2.50 Å.

The RMSD-values were relatively high, even among molecules with similar conformations. If only the areas of the β-strands in these molecules were used to calculate the RMSDs, they lowered to a range of 0.23 Å to 0.42 Å. If molecules B and D of the native HFBI structure were included in the calculations the values were above 2.2 Å, which was expected since the moving area partly overlaps with the β-sheet structure. The resolution of the data may affect the RMSD-values[82], yet the relatively high RMSDs might also indicate that plasticity of the molecule is an intrinsic property of hydrophobins. Plasticity could be a requirement for a protein that interacts with surfaces, since the interacting counterpart may not be perfectly smooth. Still, the β-barrel structure enhanced by the disulphide bridge array appears as a rigid structure, in spite of the observed 'liveliness' and large changes involving rearrangements of several residues are unlikely to occur, except for about the functional site and the loop areas. The hydrophobic surface area is the region with the highest demand for adaptability and a demonstration of such an adaptation can be seen in the HFBI structure. Some adaptability was also seen in the hydrophobic surface area of HFBII, with dual conformations for side-chains of Leu19, Val57 and Leu62.

5.5. HYDROPHOBIC PATCH[II, III, IV]

Protein molecules typically fold in such a way that the hydrophobic amino acid residues are buried in the core of the protein while charged, hydrophilic amino acid residues are located on the protein surface, interacting with solvent molecules. A hydrophobic patch was found on the surface of each hydrophobin molecule and the residues in the patch region are conserved among class II hydrophobins (Figure 1). We therefore conclude that the patch is the basis of the amphiphilicity of these molecules and the functional site of this protein. The size and composition of this patch may be involved in determining the specific properties of each hydrophobin. In the hydrophobin structures determined in this study, however, the size and shape of the hydrophobic patches are very similar (Figure 9).

The SAA of an isolated HFBII molecule was about 3960 $Å^2$ while the SAA of an isolated HFBI molecule in the native structure was about 4100 $Å^2$ and 4070 $Å^2$ in the N-Cys HFBI-variant structure. This small variation in size is likely due to the protein chains being of different length in different molecules. The SAA of HFBII dimer was about 7300 $Å^2$ and the SAA of the HFBI tetramer was about 14100 $Å^2$. The SAA of the hydrophobic patch for an isolated molecule was calculated at approximately 730 $Å^2$ for HFBII and 750 $Å^2$ and 760 $Å^2$ for HFBI in the native and the variant structures, respectively. This makes the size of the hydrophobic patch approximately 18% of the total surface area for each molecule.

When the entire content of the asymmetric unit was taken into account in calculations, the area of the hydrophobic patch exposed to the solvent was about 480 $Å^2$, 410 $Å^2$ and 280 $Å^2$ for HFBII, HFBI and HFBI in N-Cys variant, respectively. Parts of the hydrophobic patch were buried in the multimer interface in each structure, while in the N-Cys HFBI-variant structure, the detergent molecules also buried parts of the hydrophobic surface areas of the protein. In fact, in the N-Cys HFBI-variant structure, 36% of the hydrophobic patch in each molecule on average was buried by the adjacent protein molecules in the asymmetric unit and 27% of the patch area was buried by the detergent molecules. Altogether, the percentage of the patch area buried in the asymmetric unit is on average 34%, 45% and 64% for HFBII, native HFBI and variant HFBI structures, respectively.

When the symmetry related molecules were also included in the calculations of SAAs, the SAAs were approximately 2140 $Å^2$, 2780 $Å^2$ and 2910 $Å^2$ for HFBII, HFBI and N-Cys HFBI molecules, respectively. Since these figures described the actual situation in the crystals, they correlated with the solvent content of the crystals (37% for HFBII, 59% for HFBI and 74% for N-Cys HFBI). For the HFBI structure, the crystal packing, i.e. the symmetry related molecules, did not affect the solvent accessible area of the hydrophobic patch. However, in the N-Cys HFBI structure, the symmetry related molecules have an effect through the formation of the detergent-octamer structure and 70% of the patch on average is buried, while the number in absence of the symmetry related molecules is 64%. In the case of HFBII, the tight crystal packing has a

remarkable effect on the solvent accessible area, as it also does for the residues of the hydrophobic patch. 34% of the hydrophobic patch is buried between the dimer interface and in the presence of the symmetry related molecules, 81% of the patch is buried. If the total area of the molecule is considered, about 46% of the SAA is buried in the HFBII structure, while the figures for the more loosely packed HFBI and N-Cys HFBI structures are 32% and 29%, respectively.

5.6. MULTIMERIZATION[II, III, IV]

Torkkeli and co-workers found by solution SAXS (small angle X-ray scattering) and size exclusion chromatography that class II hydrophobins HFBI and HFBII from *T. reesei* form multimers in solution[83]. The multimer formation was found to be concentration dependent in such a way that at 10 mg/ml both proteins formed tetramers and at 0.5 mg/ml dimers and monomers were observed for HFBI and HFBII, respectively. By size-exclusion chromatography and dynamic light scattering, it was found that, in solution, the class I hydrophobin SC3 from *S. commune* is primarily dimeric with some monomers, tetramers and larger aggregates also present[84]. When studied by CD-spectroscopy, it was concluded that multimerization in solution does not significantly change the protein structure.

The multimerization states of hydrophobins HFBII, HFBI and N-Cys HFBI are dimeric, tetrameric and monomeric in the solved crystal structures, respectively. In the HFBII structure, two dimers are in contact through the hydrophobic patch areas, but the contact is not as tight as within the dimer and the multimerization state is concluded to be dimeric rather than being a loose tetramer. In the HFBI structure, molecule pairs A-B and C-D share more interactions than pairs A-C and B-D. It is possible that multimers of HFBI form first as dimers, which then combine to form a tetramer. The quaternary structure of the N-Cys HFBI is a detergent-associated octamer, but if the protein-protein interactions only are accounted for, the protein is monomeric. The multimerization states of the determined crystal structures and the protein concentration in the crystallizing drop correlated with the results of Torkkeli and co-workers: at a concentration of 8mg/ml, HFBI formed a tetramer and smaller multimers were formed at lower concentrations i.e. HFBII dimer at 4 mg/ml and N-Cys HFBI monomer at 3 mg/ml.

We suggest that in order to remain stable and water-soluble, hydrophobins form multimers in the solution in such a way that the functional hydrophobic patch is partially concealed from the solvent by multimerization. Thus, the driving force in multimerization is the need to achieve an energetically more favorable state, since when the hydrophobic surface area is partially concealed in the multimer interface it is in a low-energy state, as compared to the monomer, where the entire hydrophobic patch area is exposed to solvent. As 34% of the hydrophobic patch in each molecule is concealed in the dimer interface in the HFBII structure and 45% is concealed by tetramer formation in the HFBI structure, the formation of a larger multimer appears to

be more favorable. In the N-Cys HFBI-variant structure, the hydrophobin interacts with a hydrophobic counterpart and the hydrophobic patch is concealed even more efficiently (64%). Presumably, once the monolayer of the protein is formed on the hydrophobic-hydrophilic interface, the entire hydrophobic surface area faces the hydrophobic counterpart, producing again an energetically more favorable state in comparison to the multimeric, soluble state. Therefore, the formation of monolayers dominates the multimer formation. Figure 16 schematically describes our suggestion.

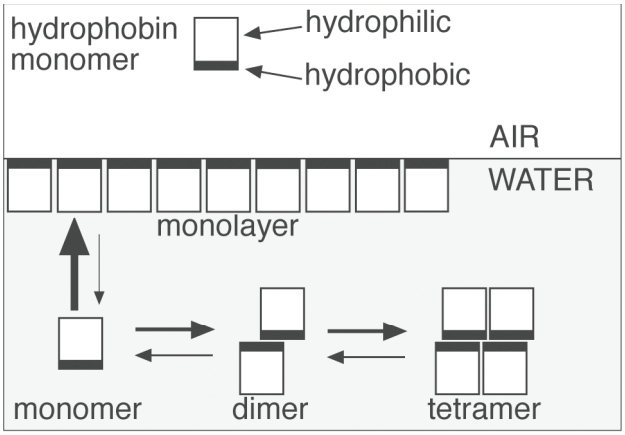


Figure 16. A proposal for the principle of multimerization in solution and the formation of a monolayer.

5.7. THE UNIVERSALITY OF THE HYDROPHOBIN FOLD

The described hydrophobin structures confirmed the existence of four internal disulphide bridges, even though the order of the bridges is not consecutive, as had been predicted for another class II hydrophobin[32]. The conservation of the functionally relevant residues, i.e. the residues of the hydrophobic patch in the sequences of class II hydrophobins strongly suggest that all the class II hydrophobins share a common fold. It has been speculated that the two hydrophobin classes have evolved independently, therefore they could have different modes of function[85]. This might also indicate that the three-dimensional structures are different. However, it had generally been recognized that the disulphide bridge array is the same for both hydrophobin classes (even though there was no direct structural evidence of this), which supports structural similarity. In addition, the few β-strands detected in the NMR-study of the class I hydrophobin EAS[10] coincide with the β-strands of the HFBII structure, as seen in Figure 1.

As judged by the sequences, the largest differences between the two hydrophobin classes are between the third and the fourth cysteine, in the area of the fifth cysteine and between the last two cysteines. These consist of, in the structures determined in this study, the two β-hairpin loops, composed of the hydrophobic residues and the α-helix. The sequences of class I hydrophobins in the area corresponding to the β-hairpin loops are also composed of several hydrophobic amino acids, which indicates that they serve a similar function. These areas are also larger; i.e. the sequences of class I hydrophobins contain more residues between the third and the fourth and between the seventh and the eighth cysteines than the class II hydrophobins. If the location of the functional site is the same in both classes, the larger region of hydrophobic residues could explain why the assembled layers of class I hydrophobins are more resistant to dissociation.

The area near the fifth cysteine is shorter in the class I sequence than in the class II sequence. The α-helix in class II structures is located in this area. In class I sequences, the stretch of residues preceding the fifth cysteine is often rich in glycine, serine and proline. These residues together with tyrosine are all 'helix breakers', the amino acid residues that are infrequently present in helical structures[86]. Some of the class II sequences also contain these residues before the fifth cysteine, but to a lesser extent. However, the area in class II sequences contains many more 'helix makers', residues alanine and glutamine, as compared to class I sequences. The 'helix maker' residues also include leucine and methionine, but these residues are not present in the discussed area in sequences of either class. The amino acid composition and the length of the segment about the fifth cysteine suggest that, in class I hydrophobins, either no α-helix is present at this position or it is considerably shorter.

The rigid, compact structures observed in this study do not support the suggestion of hydrophobins being natively unfolded (unstructured in solution) and that they only fold in the presence of hydrophobic-hydrophilic interface, as previously proposed for class I hydrophobin EAS[10]. At the end of this study, new NMR-results[87] for EAS were published, showing that EAS is indeed structured and confirming that EAS has the same β-barrel core structure and arrangement of disulphide bridges as observed for HFBII (Figure 17). Instead of the α-helix, EAS contains two short β-strands at the corresponding location. Neither the structures determined in this study nor the newly determined NMR structure support extensive refolding of hydrophobins and the consequent different secondary structural states in solution and in an assembly[32]. There is no demand for refolding, since hydrophobins are already in their functional conformation in solution, i.e. amphiphilic. The multimerization keeps the hydrophobic parts of the molecule concealed from the solvent.

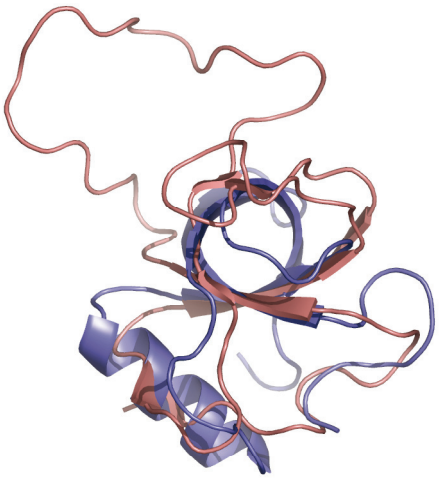


Figure 17. Superimposition of HFBII and the newly solved structure of EAS[55,87].

6. CONCLUSIONS

Two class II hydrophobins, HFBI and HFBII from *Trichoderma reesei* and a variant of HFBI have been structurally characterized by X-ray crystallography in this study. These crystal structures produced the first three-dimensional structural information of hydrophobins at a molecular level. A new fold, consisting of a small β-barrel and an α-helix and reinforced by four disulphide bridges, was discovered for this family of proteins. The functional site of hydrophobins, a patch consisting of hydrophobic residues, was identified in this study. Based on the observed multimerization states found in each crystal structure, a proposal for the principle of multimerization and monolayer formation was described. A structural change in the functional area of the native HFBI structure was discovered and it was proposed that intrinsic plasticity is an important feature for these surface-interacting proteins.

The main difficulties encountered in this study were the modest resolution of the diffraction data, even though the monoclinic crystals of HFBII did diffract remarkably well. However, for the cubic crystal form of HFBII and for the N-Cys HFBI-variant, data collection with the home source was not possible because of poor diffraction quality. Even with data collected with a high-intensity synchrotron source, the resolution remained rather low. The other main obstacle was the twinning in the native HFBI crystals, which presented some problems in refining the structure.

This study has broadened our comprehension of hydrophobins, especially for class II hydrophobins. The existence of a three-dimensional stucture enables the design of experiments to further the engineering of hydrophobins to meet the demands of industrial and technical applications. Interactions found in the detergent-associated crystal structure and the variety of multimers encountered in the crystal structures indicate that many more types of quaternary structures are yet to be discovered. However, little is still known about the details of the self-assembly process and the molecular organization of the assembled layers.

ACKNOWLEDGEMENTS

This work was carried out at the Department of Chemistry of the University of Joensuu during 2002-2006 and at the EMBL Hamburg Outstation during a three-month Marie Curie Fellowship in 2004. This work was funded by the National Graduate School in Informational and Structural Biology, which is gratefully acknowledged. I wish to thank the leader of our graduate school, Prof. Mark Johnson and Kaija Söderlund, Fredrik Karlsson and numerous other people responsible for the function and coordination of the school. Funding from the EU for the Marie Curie training period (Contract No. HPMT-CT-2000-00174) and for the data collection (Access to the research infrastructure action of the Improving Human Potential Programme) is gratefully acknowledged. Funding from the Department of Chemistry of the University of Joensuu is also gratefully acknowledged.

I owe my deepest gratitude to my supervisor, Prof. Juha Rouvinen, for all his guidance, advice, and care-taking, continuous encouragement and his broad-minded way of thinking. I wish to thank all the past and present members of the protein group, namely Nina, Jape, Carita, Tuomo, Tarja, Merja and Mikko, and all our undergraduates and summer students, especially Mirko and Heidi. My warmest thanks go out to Prof. Markku Ahlgren for his help in determining the twin operator and to Dr Janne Jänis, for recording the hydrophobin mass spectra. I would also like to thank Prof. Pirjo Vainiotalo and Prof. Juhani Syväoja for their advice as my thesis committee members, Reetta Kallio-Ratilainen and Liisa Saharinen for skilled technical assistance, and Martti Lappalainen and Urpo Ratinen, aka 'pajan pojat', for their continuous technical support. I also wish to thank the entire staff and the students of the Department of Chemistry for the friendly and enjoyable atmosphere they helped to create.

This work was carried out in collaboration with the group of Markus Linder at VTT Biotechnology. I warmly thank Markus and his group for providing us with the protein materials, letting us exploit their vast knowledge about hydrophobins and for the inspiring meetings and conversations. I also wish to thank all the people at VTT for making my stay there (Jan-Feb 2005) so enjoyable. Special thanks go to Géza Szilvay and Riitta Suihkonen for all their help.

I'm especially grateful to my supervisors during the Marie Curie fellowship in Hamburg, Dr Andrea Schmidt and Dr Alexander Popov for their guidance and their expertise, through which I learned so much. I'd like to thank our group leader, Dr Victor Lamzin for all his advice and help. I also wish to thank the fellows with whom I shared an office in Hamburg: Marc, Will, Preben and Brice and my mate Jack for making my stay in Hamburg a lot less lonely. The entire staff of EMBL Hamburg provided a warm and friendly athmosphere and the beamline staff is acknowledged for all their help during the numerous data collections trips to Hamburg. Mr. Pekka Salo from Finnair is particularly thanked for providing us the safety sheets and the means to carry our crystals onboard.

I'm very grateful for the love and support that I have received from my family and friends. I wish to dedicate this thesis to my parents, who have taught me the most important things in life. I thank my sister Katja and her family for their continuous encouragement and my brother Jouni for keeping my feet firmly on the ground. My warmest thanks also to Joi, Paul and Alek for their love and care, for making me a family member and allowing me to remain in their hearts in spite of the years that have gone by and the ocean between us. I thank Mia and Anu for the kind of friendship that nothing could replace and for the love I can always rely on.

Finally, I find myself out of words when trying to express my gratitude to and love for my fiancé Juha. I cherish the life that we share and I warmly thank you for your support, understanding and encouragement.

May, 2006

Johanna Hakanpää

REFERENCES

1. Wessels J.G.H., de Vries O.M.H., Asgeirsdottir S.A., Schuren F.H.J. (1991) Hydrophobin genes involved in formation of aerial hyphae and fruit bodies in Schizompyllum., *Plant Cell* **3** 793-799
2. Wösten H.A.B., Schuren F.H.J., Wessels J.G.H. (1994) Interfacial self-assembly of a hydrophobins into an amphiphilic protein membrane mediates fungal attachement to hydrophobic surfaces., *EMBO J*. **13** 5848-5854
3. Wessels J.G.H (1996) Fungal hydrophobins: proteins that function at an interface., *Trends Plant Sci*. **1** 9-14
4. Linder M.B., Szilvay G.R., Nakari-Setälä T., Penttilä M.E. (2005) Hydrophobins: the protein-amphiphiles of filamentous fungi., *FEMS Microbiol. Rev*. 5 877-896
5. Nakari-Setälä T., Aro N., Ilmén M., Muños G., Kalkkinen N., Penttilä M. (1997) differential expression of the vegetative and spore-bound hydrophobins of *Trichoderma reesei*. Cloning and characterization of the *hfb2* gene., *Eur. J. Biochem.* **248** 415-423
6. Schuren F.H., Wessels J.G. (1990) Two genes specifically expressed in fruiting dikaryons of Schizophyllum commune: homologies with a gene not regulated by mating-type genes., *Gene* 90 199-205
7. Wessels J.G.H. (1994) Developmental regulation of fungal cell wall formation., *Annu. Rev. Phytopathol.* **32** 413-437
8. de Groot P.W., Schaap P.J., Sonnenberg A.S., Visser J., van Griensven L.J. (1996) The *Agaricus bisporus hypA* gene encodes a hydrophobin and specifically accumulates in peel tissue of mushroom caps during fruit body development., *J. Mol. Biol.* **257** 1008-1018
9. Lugones L.G., Wösten H.A., Wessels J.G. (1998) A hydrophobin (ABH3) specifically secreted by vegetatively growing hyphae of *Agaricus bisporus* (common white button mushroom)., *Microbiology* **144** 2345-2353
10. Mackay J.P., Matthews J.M., Winefield R.D., Mackay L.G., Haverkamp R.G., Templeton M.D. (2001) The hydrophobin EAS in largely unstructured in solution and functions by forming amyloid-like structures., *Structure* 9 83-91
11. Nakari-Setälä T., Aro N., Kalkkinen N., Alatalo E., Penttilä M. (1996) Genetic and biochemical characterization of the *Trichoderma reesei* hydrophobin HFBI., *Eur. J. Biochem.* **235** 248-255
12. Penttilä M., Nakari-Setälä T., Fagerström R., Selber K., Kula M.R., Linder M., Tjerneld F. (2000) Int. Pat. Appl. PCT/FI00/00249
13. Segers G.C., Hamada W., Oliver R.P., Spanu P.D. (1999) Isolation and characterization of five different hydrophobin-encoding cDNAs from the fungal tomato pathogen *Cladosporium fulvum., Mol. Gen. Genet*. **261** 644-652
14. Munoz G., Nakari-Setälä T., Agosin E., Penttilä M. (1997) Hydrophobin gene *srh1*, expressed during sporulation of the biocontrol agent *Trichoderma harzianum., Curr. Genet*. **32** 225-230
15. Zhang L., Villalon D., Sun Y., Kazmierczak P., van Alfen N.K., (1994) Virus-associated down-regulation of the gene encoding cryparin, an abundant cell-surface protein from the chestnut blight fungus, *Cryphonectria parasitica., Gene* **139** 59-64
16. de Vries O.M.H., Moore S., Arntz C., Wessels J.G.H., Tudzynski P. (1999) Identification and characterization of a tri-partite hydrophobin form *Claviceps fusiformis*. A novel type of class II hydrophobin., *Eur. J. Biochem.* **262** 377-385
17. Bolyard M.G., Sticklen M.B. (1992) Expression of a modified Dutch elm disease toxin in *Escherichia coli., Mol. Plant-Microbe Interact*. **5** 520-524
18. Lora J.M., de la Cruz J., Benitez T., Llobell A., Pintor-Tori J.A. (1994) A putative catabolite-repressed cell wall protein from the mycoparasitic fungus *Trichoderma harzianum., Mol. Gen. Genet*. **242** 461-466
19. Wösten H.A.B., de Vries O.M.H., van der Mei H.C., Busscher H.J., Wessels J.G.H. (1994) Atomic composition of the hydrophobic and hydrophilic membrane sides of self-assembled SC3p hydrophobin., *J. Bacteriol.* **176** 7085-7086

20. Santos C., Labarère J. (1999) *Aa-Pri2*, as single-copy gene from *Agrocybe aegerita*, specifically expressed during fruiting initiation, encodes a hydrophobin with a leucine-zipper domain., *Curr. Genet.* **35** 564-570
21. Wösten H.A.B., van Wetter M.A., Lugones L.G., van der Mei H.C., Busscher H.J., Wessels J.G.H. (1999) How a fungus escapes the water to grow into the air., *Curr. Biol.* **9** 85-88
22. Ebbole D.J. (1997) Hydrophobins and fungal infection of plants and animals., *Trends Microbiol.* **5** 405-408
23. Wösten H.A.B., Willey J.M. (2000) Surface-active proteins enable microbial aerial hyphae to grow into the air *Microbiology* **146** 767-773
24. Elliot M.A., Talbot N.J., (2004) Building filaments in the air: aerial morphogenesis in bacteria and fungi., *Curr. Opin. Microbiol.* 7 1-8
25. Askolin S., Nakari-Setälä T., Tenkanen M. (2001) Overproduction, purification, and characterization of the *Trichoderma reesei* hydrophobin HFBI., *Appl. Microbiol. Biotechnol.* **57** 124-130
26. Bailey M.J., Askolin S., Hörhammer N., Tenkanen M., Linder M., Penttilä M., Nakari-Setälä T. (2002) Process technological effects of deletion and amplification of hydrohobins I and II in transformants of *Trichoderma reesei., Appl. Microbiol. Biotechnol.* **58** 721-727
27. de Vocht M.L., Reviakine I., Wösten H.A.B., Brisson A., Wessels J.G.H., Robillard G.T. (2000) Structural and functional role of the disulphide bridges in the hydrophobin SC3., *J. Biol. Chem.* **275** 28428-28432
28. Wessels J.G.H. (1997) Hydrophobins: proteins that change the nature of the fungal surface in *Adv. Microbial Physiol.* Vol 38 pp.1-45 (Ed. R.K. Poole) Academic Press Inc. San Diego
29. Linder M., Selber K., Nakari-Setälä T. Qiao M., Kula M.R., Penttilä M. (2001) The hydrophobins HFBI and HFBII from *Trichoderma reesei* showing efficient interactions with nonionic surfactants in aqueous two-phase systems., *Biomacromolecules* **2** 511-517
30. Hektor H.J., Scholtmeijer K. (2005) Hydrophobins: proteins with potential., *Curr. Opin. Biotechnol.* 16 1-6
31. de Vries O.M.H., Fekkes M.P., Wösten H.A.B., Wessels J.G.H. (1993) Insoluble hydrophobin complexes in the walls of *Schizophyllum commune* and other filamentous fungi., *Arch. Microbiol.* **159** 330-335
32. Yaguchi M., Pusztai-Carey M., Roy C., Surewicz W.K., Carey P.R., Stevenson K. J., Richards W.C., Takai S. (1993) Amino acid sequence and spectroscopic studies of Dutch elm disease toxin, cerato-umin in *Dutch Elm Disease Research, Cellular and Molecular Approaches* (Eds. M.B. Sticken and J.L. Sherald) pp. 152-170 Springer-Verlag, New York
33. Talbot N.J. (1997) Fungal biology: growing into the air., *Curr. Biol.* 7 R78-R81
34. de Vocht M.L., Scholtmeijer K., van der Vegte E.W., de Vries O.M.H., Sonveaux N., Wösten H.A.B., Ruysschaert J.M., Hadziioannou G., Wessels J.G.H., Robillard G.T. (1998) Structural characterization of the hydrophobin SC3, as a monomer and after self-assembly at hydrophobic/hydrophilic interfaces., *Biophys. J.* 74 2059-2068
35. Wright P.E., Dyson H.J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm., *J. Mol. Biol.* **293** 321-331
36. Otwinowski Z., Minor W. (1997) Processing of X-ray diffraction data collected in oscillation mode., *Methods Enzymol.* **276** 307-326
37. Kabsch W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants., *J. Appl. Crystallogr.* **26** 795-800
38. McRee D.E. (1999) XtalView/Xfit - a versatile program for manipulating atomic coordinates and electron density., *J. Struct. Biol.* **125** 156-165
39. Foadi J., Woolfson M.M., Dodson E.J., Wilson K.S., Yao J., Zheng C. (2000) A flexible and efficient procedure for the solution and phase refinement of protein structures., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56** 1137-1147

40. Lamzin V.S., Perrakis A., Wilson K.S. (2001) The ARP/wARP suite for automated construction and refinement of protein models, in *International tables for crystallography, Vol. F, Crystallography of biological macromolecules* (Rossman M.G. and Arnold E., Eds) pp. 720-722 Dordrecht, Kluwer Academic Publishers, the Netherlands. http://www.arp-warp.org
41. Vagin A., Teplyakov A. (1997) MOLREP: an automated program for molecular replacement., *J. Appl. Crystallogr.* **30** 1022-1025
42. Collaborative computational project, number 4 (1994) The CCP4 suite: programs for protein crystallography., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50** 760-763
43. Matthews B.W. (1968) Solvent content of protein crystals., *J. Mol. Biol.* **33** 491-497
44. Carugo O., Pongor S. (2002) Protein fold similarity estimated by a probabilistic approach base on C^{α}- C^{α} distance comparison., *J. Mol. Biol.* 315 887-898
45. Brünger A.T., Adams P.D., Clore G.M., DeLano W.L., Gros P., Grosse-Kunstleve R.W., Jiang J.S., Kuszewskic J.S., Nilges M., Pannu N.S., Read R.J., Rice L.M., Simonson T., Warren G.L. (1998) Crystallography & NMR System: A new software suite for macromolecular structure determination., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54** 905-921
46. Murshudov G.N., Vagin A.A., Dodson E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53** 240-255
47. Sheldrick G.M., Schneider T.R. (1997) SHELXL: High-resolution refinement., *Methods Enzymol.* **277** 319-343
48. Jones T.A., Zhou J.Y., Cowan S., Kjeldgaard M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models., *Acta Crystallogr., Sect. A: Found. Crystallogr.* **47** 110-119
49. Vriend G. (1990) WHAT IF: A molecular modeling and drug design program., *J. Mol. Graphics* **8** 52-56
50. Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures., *J. Appl. Crystallogr.* **26** 283-291
51. Kleywegt G.J., Zou J.Y., Kjeldgaard M., Jones T.A. (2001) Around O, in *Internatinal tables for crystallography, Vol. F, Crystallography of biological macromolecules* (Rossmann M.G. and Arnold E., Eds) pp. 353-356, 366-367 Dordrect, Kluwer Academic Publishers, the Netherlands
52. Madan Babu M. (2003) NCI: a server to identify non-canonical interactions in protein structures., *Nucleic Acids Res.* **31** 3345-3348
53. Zhang X.J., Matthews B.W. (1995) EDPDB: a multifunctional tool for protein structure analysis., *J. Appl. Crystallogr.* **28** 624-630
54. Lee B., Richards F.M. (1971) The interpretation of protein structures: Estimation of static accessibility., *J. Mol. Biol.* **55** 379-400
55. DeLano W.L. (2002) the PyMOL molecular graphics system, DeLano Scientific, CA, USA http://www.pymol.org
56. Askolin S., Turkenburg J.P., Tenkanen M., Uotila S., Wilson K.S., Penttilä M., Visuri K. (2004) Purification, crystallization and preliminary X-ray diffraction analysis of the *Trichoderma reesei* hydrophobin HFBI., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60** 1903-1905
57. Koch E. (1992) Application to the crystal systems, in *International Tables for Crystallography, Vol. C, Mathematical, physical and chemical tables* (Wilson A.J.C. Ed.) pp. 10-14 Dordrecht, Kluwer Academic Publishers, the Netherlands
58. Parsons S. (2003) Introduction to twinning., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **59** 1995-2003
59. Dauter Z. (2003) Twinned crystals and anomalous phasing., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **59** 2004-2016
60. Herbst-Irmer R., Sheldrick G.M. (1998) Refinement of twinned structures with SHELXL97., *Acta Crystallogr., Sect. B: Struct. Sci.* **54** 443-449
61. Yeates T.O. (1997) Detecting and overcoming crystal twinning., *Methods Enzymol.* **276** 344-358 http://nihserver.mbi.ucla.edu/Twinning/
62. Rees D.C. (1980) The influence of twinning by merohedry on intensity statistics., *Acta Crystallogr., Sect. A: Found. Crystallogr.* **36** 578-581

63. Rudolph M.G., Wingren C., Crowley M.P., Chien Y., Wilson I.A. (2004) Combined pseudo-merohedral twinning, non-crystallographic symmetry and pseudo-translation in a monoclinic crystal form of the γδ T-cell ligand T10., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60** 656-664
64. Lehtiö L., Fabrichniy I., Hansen T., Schönheit P., Goldman A. (2005) Unusual twinning in an acetyl coenzyme A synthetase (ADP-forming) from *Pyrococcus furiosus., Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61** 350-354
65. Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig I.N., Shindyalov I.N., Bourne P.E. (2000) The protein data bank., *Nucleic Acids Res.* **28** 235-242 http://www.pdb.org
66. Jelsch C., Teeter M.M., Lamzin V., Pichon-Pesme V., Blessing R.H., Lecomte C. (2000) Accurate protein crystallography at ultra-high resolution: Valence electron distribution in crambin., *Proc. Natl. Acad. Sci. U. S. A.* **97** 3171-3176
67. Ko T.P., Robinson H., Gao Y.G., Cheng C.H.C., DeVries A.L., Wang A.H.J. (2003) The refined crystal structure of an eel pout type III antifreeze protein RD1 at 0.62 Å resolution reveals structural microheterogenity of protein and solvation., *Biophys. J.* **84** 1228-1237
68. Howard E.I., Sanishvili R., Cachau R.E., Mitschler A., Chevrier B., Barth P., Lamour V., van Zandt M., Silbey E., Bon C., Moras D., Schneider T.R., Joachimiak A., Podjarny A. (2004) Ultra-high resolution drug design I: details of interactions in human aldose reductase-inhibitor complex at 0.66 Å., *Proteins* **55** 792-804
69. Bönisch H., Schmidt C.L., Bianco P., Ladenstein R. (2005) Ultrahigh-resolution study on *Pyrococcus abyssi* I. 0.69 Å x-ray stucture of mutant W4L/R5S., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61** 990-1004
70. Kang B.S., Devedjiev Y., Derewenda U., Derewenda Z.S. (2004) The PDZ2 domain of syntenin at ultra-high resolution: bridging the gap between macromolecular and small molecule crystallography., *J. Mol. Biol.* **338** 483-493
71. Kuhn P., Knapp M., Soltis S.M., Ganshaw G., Thoene M., Bott R. (1998) The 0.78 Å structure of a serine protease: *Bacillus lentus* subtilisin., *Biochemistry* **37** 13446-13452
72. Dunlop K.V., Irvin R.T., Hazes B. (2005) Pros and cons of cryocrystallography: should we also collect a room-temperature data set?, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61** 80-87
73. Liu L., Nogi T., Kobayashi M., Nozawa T., Miki K. (2002) Ultrahigh-resolution structure of high-potential iron-sulfur protein from *Thermochromatium tepidum., Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58** 1085-1091
74. Schmidt A., Jelsch C., Østergaard P., Rypniewski W., Lamzin V.S. (2003) Trypsin revisited., *J. Biol. Chem.* **278** 43357-43362
75. Jamal-Talabani S., Boraston A.B., Turkenburg J.P., Tarbouriech N., Ducros V.M.A., Davies G.J. (2004) *Ab initio* structure determination and functional characterization of CBM36: a new family of calcium-dependent carbohydrate binding molecules., *Structure* **12** 1177-1187
76. Dauter Z., Lamzin V.S., Wilson K.S. (1997) The benefits of atomic resolution., *Curr. Opin. Struct. Biol.* **7** 681-688
77. Schmidt A., Lamzin V. (2002) *Veni, vidi vici* - atomic resolution unravelling the mysteries of protein function., *Curr. Opin. Struct. Biol.* **12** 698-703
78. Longhi S., Czjzek M., Cambillau C. (1998) Messages from ultra-high resolution crystal structures., *Curr. Opin. Struct. Biol.* **8** 730-737
79. Afonine P.V., Lunin V.Y., Muzet N., Urzhumtsev A. (2004) On the possibility of the observation of valence electron density for individual bonds in proteins in conventional difference maps., *Acta Crystalogr., Sect. D: Biol. Crystallogr.* **60** 260-274
80. Guillot B., Viry L., Guillot R., Lecomte C., Jelsch C. (2001) Refinement of proteins at subatomic resolution with MOPRO., *J. Appl. Crystallogr.* **34** 214-223
81. Sevcik J., Dauter Z., Lamzin V.S., Wilson K.S. (1996) Ribonuclease from *Streptomyces aureofaciens* at atomic resolution., *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **52** 327-344
82. Carugo O. (2003) How root-mean-square distance (r.m.s.d.) values depend on the resolution of protein structures that are compared., *J. Appl. Crystallogr.* **36** 125-128

83. Torkkeli M., Serimaa R., Ikkala O., Linder M. (2002) Aggregation and self-assembly of hydrophobins from *Trichoderma reesei*: low-resolution structural models., *Biophys. J.* 83 2240-2247
84. Wang X., Graveland-Bikker J.F., de Kruif C.G., Robillard G.T. (2004) Oligomerization of hydrophobin SC3 in solution: from soluble state to self-assembly., *Protein Sci.* **13** 810-821
85. Whiteford J.R., Spanu P.D. (2001) The hydrophobin HCf-1 of *Cladosporium fulvum* is required for efficient water-mediated dispersal of conidia., *Fungal Genet. Biol.* **32** 159-168
86. Branden C., Tooze J. (1991) *Introduction to Protein Structure*., Garland Publishing Inc, New York and London, pp.13-15
87. Kwan A.H.Y., Winefield R.D., Sunde M., Matthews J.M., Haverkamp R.G. Templeton M.D., Mackay J.P. (2006) Structural basis for rodlet assembly in fungal hydrophobins., *Proc. Natl. Acad. Sci. U. S. A.* **103** 3621-3626

Earlier Dissertations from the Department of Chemistry, University of Joensuu

The complete list of earlier dissertations can be found in the dissertation 54.

54/2001 SUOMALAINEN Pekka: Structural and theoretical studies on triphenylphosphane derivatives and application as ligands in rhodium-catalyzed hydroformylation of 1-hexene

55/2001 LAITINEN Tuomo: Studies of the functional properties of endoxylanase II from *Trichoderma reesei*

56/2002 KINNUNEN Toni: Experimental and theoretical studies on the effects of ligand modifications in bipyridine complexes of ruthenium and nickel

57/2002 VALJAKKA Jarkko: Crystal structures of recombinant anti-testosterone Fab-fragments: insights into antibody engineering

58/2002 LAINE Olli: Structural characterization of synthetic polymers by mass spectrometry

59/2002 LUTZ Matthias: Tin and zirconium heterodimetallic complexes stabilized by tripodal amido ligands

60/2002 NOUSIAINEN Marjaana: Osteocalcin, calmodulin, and troponin C: metal ion and peptide binding study by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry

61/2003 HAKULINEN Nina: Structural biology of wood degrading enzymes

62/2003 KARVINEN Saila: Experimental and theoretical studies on doped and undoped rutile and anatase TiO_2 for photocatalyst and pigment use

63/2003 GOLDE Martin: Catalysts for heterogenous hydrogenation of aldonic acids

64/2003 RUUSKA Henna: *Ab initio* cluster model study on interaction of water with calcite, graphite, and copper surfaces

65/2003 MÄKINEN Marko: Conformational properties, intra-molecular H-bonds and supramolecular complex formation of tetraethyl resorcarene: an *ab initio* and ESI-FTICR mass spectrometric study

66/2003 JALKANEN Jukka-Pekka: Quantum chemical potential energy surfaces for small hydrocarbon molecules

67/2004 HUKKAMÄKI Jarkko: Preparation of micro- and mesoporous silica materials, and templating effect of MCM-41 and SBA-15 silicas on cobalt deposition

68/2004 LUHTANEN Tommi: Theoretical studies on aluminoxane co-catalysts and magnesium dichloride supports of polymerization catalysts: from monomers to nanostructures and solids

69/2004 MONONEN Kirsi: Discoloration of silver birch (*Betula pendula*) wood induced by kiln drying and hydrogen peroxide bleaching

70/2004 PORENTO Mika: Theoretical *ab initio* studies on the interaction of sulphydryl collectors with selected metal ions and sulphide minerals

71/2004 LATVA Sanna: Metal-modified activated charcoals in concentration and separation of arsenic, selenium and chromium species before EDXRF and GFAAS analyses

72/2005 JÄNIS Janne: Structural, thermodynamic, and kinetic characterization of xylanases by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry

73/2005 KONTTURI Mervi: Crystal structures and characterisation of metal complexes of dichloromethylene bisphosphonate and of its symmetrical dialkyl ester derivatives

74/2005 FENEL Fred: Mutational studies of the industrially important family 11 xylanase from *Trichoderma reesei*

75/2005 KATAJISTO Jussi: Properties of polycarbonate polymers and cyclo-olefin copolymers predicted by *ab initio* and molecular simulation techniques

76/2005 MORENO, M. Andreina: Ruthenium and iridium carbonyls. Catalytic activity and formation of ruthenium complexes containing aromatic nitrogen donor and phosphane ligands

77/2006 VENTOLA Elina: Host-guest chemistry of resorcarene derivatives studied by ESI-FTICR mass spectrometry