Structural bases for N-glycan processing by mannoside phosphorylase

Simon Ladevèze,[a,b,c] Gianluca Cioci,[a,b,c] Pierre Roblin,[d] Lionel Mourey,[e,f] Samuel Tranier[e,f]* and Gabrielle Potocki-Véronèse[a,b,c]*

[a]Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, 31077 Toulouse, France, [b]CNRS, UMR5504, 31400 Toulouse, France, [c]INRA, UMR792 Ingénierie des Systèmes Biologiques et des Procédés, 31400 Toulouse, France, [d]Synchrotron SOLEIL, L'Orme des Merisiers, BP 48, Saint Aubin, 91192 Gif-sur-Yvette CEDEX, France, [e]Institut de Pharmacologie et de Biologie Structurale (IPBS), Centre National de la Recherche Scientifique (CNRS), 205 Route de Narbonne, BP 64182, 31077 Toulouse, France, and [f]Université de Toulouse, Université Paul Sabatier, IPBS, 31077 Toulouse, France. *Correspondence e-mail: samuel.tranier@ipbs.fr, veronese@insa-toulouse.fr

The first crystal structure of Uhgb_MP, a β-1,4-mannopyranosyl-chitobiose phosphorylase belonging to the GH130 family which is involved in N-glycan degradation by human gut bacteria, was solved at 1.85 Å resolution in the apo form and in complex with mannose and N-acetylglucosamine. SAXS and crystal structure analysis revealed a hexameric structure, a specific feature of GH130 enzymes among other glycoside phosphorylases. Mapping of the −1 and +1 subsites in the presence of phosphate confirmed the conserved Asp104 as the general acid/base catalytic residue, which is in agreement with a single-step reaction mechanism involving Man O_3 assistance for proton transfer. Analysis of this structure, the first to be solved for a member of the GH130_2 subfamily, revealed Met67, Phe203 and the Gly121–Pro125 loop as the main determinants of the specificity of Uhgb_MP and its homologues towards the N-glycan core oligosaccharides and mannan, and the molecular bases of the key role played by GH130 enzymes in the catabolism of dietary fibre and host glycans.

1. Introduction

N-linked glycans are present in many living organisms, notably eukaryotes, and play a key role in major processes, including cell signalling and recognition, protein stability and activity tuning (Varki *et al.*, 2009). These oligosaccharides, which are covalently linked to the asparagine residues of glycoproteins, display relatively restricted structural diversity (Lehle *et al.*, 2006; Larkin & Imperiali, 2011). In eukaryotes, N-glycans share a common core structure composed of the β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc (Man-GlcNAc$_2$) trisaccharide, carrying decorations on the nonreducing β-linked mannosyl residue to form more complex structures (Aebi *et al.*, 2010; Nagae & Yamaguchi, 2012). Although the whole pathways of human N-glycan synthesis and maturation have been well described, little is known about their degradation, especially by bacteria or fungi, despite the fact that degradation is a key factor in microbe–host interactions (Suzuki & Harada, 2014). Until 2013, only glycoside hydrolases (GHs) had been shown to be implicated in N-glycan breakdown in the CAZy database (http://www.cazy.org/; Lombard *et al.*, 2014). Huge efforts have been made in recent years to understand exactly how this hydrolytic process takes place, particularly among gut inhabitants, as the alteration of host glycans by microbes is thought to be related to intestinal disorders, including Crohn's disease and other inflammatory
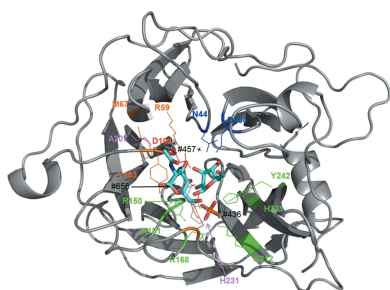
bowel diseases (IBDs; Martens *et al.*, 2008; Sheng *et al.*, 2012). The aim of these studies was to identify a broad consortium of enzymes acting on different parts of the N-glycan structure, such as exo-mannosidases and endo-mannosidases or endo-*N*-acetyl-β-D-glucosaminidases produced by gut commensals and pathogens (Roberts *et al.*, 2000; Burnaugh *et al.*, 2008; Renzi *et al.*, 2011), particularly by *Bacteroides* species (Martens *et al.*, 2009; Zhu *et al.*, 2010).

In 2013, the first evidence for N-glycan breakdown by phosphorolysis was published, which involved gut bacterial mannoside phosphorylases belonging to glycoside hydrolase family 130 (GH130; Nihira *et al.*, 2013; Ladevèze *et al.*, 2013). Only two enzymes, namely the mannoside phosphorylase (EC 2.4.1.–) Uhgb_MP, an enzyme produced by an uncultivated *Bacteroides* bacterium, and *Bt*1033, produced by *B. thetaiotaomicron* VPI-5482, are known to catalyze the conversion of β-D-Man*p*-1,4-β-D-Glc*p*NAc (Man-GlcNAc) or β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc (Man-GlcNAc$_2$) and inorganic phosphate into α-D-mannopyranose-1-phosphate and GlcNAc or GlcNAc$_2$, respectively (Nihira *et al.*, 2013; Ladevèze *et al.*, 2013). CAZy subfamilies are subgroups found within a family that share a more recent ancestor and that are usually more uniform in molecular function, reflecting a high degree of conservation in their active site (Aspeborg *et al.*, 2012). In the GH130 family, at least two enzyme subfamilies have been identified (Ladevèze *et al.*, 2013). Subfamily GH130_1 gathers enzymes that are highly specific for β-D-Man*p*-1,4-D-Glc, while GH130_2 contains enzymes that are much more flexible towards mannosides. Uhgb_MP and *Bt*1033 are classified in the GH130_2 subfamily, together with 40 other GH130 sequences, among which 15 originate from gut bacterial genomes. Integration of metagenomic and genomic data on the scale of the entire human gut microbiota revealed that GH130_2 enzymes, especially Uhgb_MP and *Bt*1033, probably play a critical role in alteration of the intestinal barrier, as their encoding genes are particularly prevalent in the human gut microbiome of patients suffering from IBDs (Ladevèze *et al.*, 2013). Based on genomic context analysis and on the *in silico* detection of signal peptides, the physiological role of Uhgb_MP and *Bt*1033 would be the intracellular phosphorolyis of β-D-Man*p*-1,4-D-GlcNAc, which can be internalized in the cell after extracellular hydrolysis of N-glycans by glycoside hydrolases belonging to the GH18, GH92 and possibly also the GH97 families (Ladevèze *et al.*, 2013). In addition to Uhgb_MP and *Bt*1033, subfamily GH130_2 contains only one other biochemically characterized enzyme, the *Ra*MP2 enzyme from the ruminal bacterium *Ruminococcus albus* 7. It has been suggested that this enzyme is involved in mannan catabolism in the bovine rumen, as it catalyzes the phosphorolysis of β-1,4-manno-oligosaccharides (Kawahara *et al.*, 2012). *In vitro*, these three enzymes present a relaxed substrate specificity compared with all other known mannoside phosphorylases. This property makes them extremely interesting biocatalytic tools for the synthesis of diverse manno-oligosaccharides by reverse phosphorolysis. In particular, Uhgb_MP is extremely efficient at producing N-glycan core oligosaccharides such as β-D-Man*p*-1,4-D-GlcNAc and

β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc, the current commercial price of which exceeds \$10 000 per milligram (Ladevèze *et al.*, 2014). Moreover, it is the only known phosphorylase to act on mannans and long manno-oligosaccharides. Uhgb_MP-based β-mannoside synthesis processes are highly attractive, thanks to its flexible specificity and because it is the only known enzyme able to produce such high added value compounds from a hemicellulose constituent as a substrate. Indeed, a one-pot reaction would allow Uhgb_MP to produce β-D-Man*p*-1,4-β-D-Glc*p*NAc directly from *N*-acetyl-glucosamine and mannan following two reaction steps: a first step of mannan phosphorolysis releasing α-D-Man-1-phosphate, and a second step of reverse phosphorolysis converting α-D-Man-1-phosphate and *N*-acetylglucosamine into β-D-Man*p*-1,4-β-D-Glc*p*NAc.

Currently, six GH130 enzyme structures are available in the RCSB Protein Data Bank, sharing a common five-bladed β-propeller fold. The crystal structure of *Bf*MGP, a *B. fragilis* NCTC 9343 enzyme classified with 78 other sequences into the GH130_1 subfamily, was recently solved in complex with the genuine substrates 4-*O*-β-D-mannosyl-D-glucose and phosphate and the product α-D-mannose-1-phosphate (Nakae *et al.*, 2013; PDB entries 3wat, 3was, 3wau and 4kmi). This enzyme, which is involved in the final steps of mannan catabolism in the human gut, exhibits a very narrow specificity towards β-D-Man*p*-1,4-D-Glc (Senoura *et al.*, 2011), like the other characterized members of the GH130_1 subfamily (*Ra*MP1 from the ruminal bacterium *R. albus* 7 and the *Rm*MGP protein from the marine bacterium *Rhodothermus marinus* DSM4252; Jaito *et al.*, 2014). This pioneering study on *Bf*MGP highlighted a probably unique reaction mechanism among known disaccharide phosphorylases, as the invariant residue Asp131, which is assumed to be the general acid/base, was not found close to the glycosidic O atom, which should be protonated in the catalytic reaction.

The five other three-dimensional structures of GH130 enzymes available to date are the apo forms of (i) four proteins belonging to the GH130_NC cluster, which gathers enzymes that are not classified into the GH130_1 and GH130_2 subfamilies, namely BACOVA_03624 and BACOVA_02161 from *B. ovatus* ATCC 8483 (PDB entries 3qc2 and 4onz; Joint Center for Structural Genomics, unpublished work), BDI_3141 from *Parabacteroides distasonis* ATCC 8503 (PDB entry 3taw; Joint Center for Structural Genomics, unpublished work) and BT_4094 from *B. thetaiotaomicron* VPI-5482 (PDB entry 3r67; Joint Center for Structural Genomics, unpublished work), and (ii) *Tm*1225 from *Thermotoga maritima* MSB8 (PDB entry 1vkd; Joint Center for Structural Genomics, unpublished work), which belongs to the GH130_2 subfamily. No function has yet been attributed to these five proteins, thus limiting our understanding of their structure–specificity relationships. Until now, nothing has been established regarding the molecular bases of the relaxed specificity of the enzymes classified into the GH130_2 family. In addition, no structural feature has been identified to explain the efficiency of Uhgb_MP and *Bt*1033 in binding and breaking down N-glycan core oligosaccharides.

Here, we present the first crystal structure of an N-glycan phosphorolytic enzyme, Uhgb_MP, solved by X-ray crystallography in complex with inorganic phosphate, mannose and N-acetylglucosamine. This study made it possible to review the previously published three-dimensional model of Uhgb_MP and provides key information to understand its catalytic mechanism. Comparative analysis of this new tertiary and quaternary structure with other GH130 structures allowed us to identify structural features specific to GH130 subfamilies that could explain their functional specificities and hence their key role in mannose foraging in the human gut. This work therefore paves the way for enzyme optimization by rational engineering to fit industrial needs as well as for the design of specific inhibitors to investigate, and potentially to control, interactions between host and gut microbes.

2. Materials and methods

2.1. Recombinant production of Uhgb_MP and enzyme purification

Uhgb_MP was produced in *Escherichia coli* BL21-AI cells (Invitrogen) after its encoding gene had been cloned into the pET-28a vector, yielding an N-terminally hexahistidine-tagged protein (detailed procedures are provided as Supporting Information). After purification by His-tag affinity chromatography and gel filtration, the enzyme was stored in 20 mM potassium phosphate pH 7.0, 150 mM NaCl (see Supporting Information).

2.2. Activity measurements

Phosphorolytic activity was assessed using two substrates, pNP-β-D-mannopyranose and β-D-mannopyranosyl-1,4-D-mannose. All reactions were carried out with 0.1 mg ml^{-1} purified enzyme at 37°C (the optimal temperature for Uhgb_MP) in 20 mM Tris–HCl pH 7.0 (the optimal pH for Uhgb_MP). For measurement of the activity in the presence of 10 mM inorganic phosphate and 1 mM pNP-β-D-mannopyranose, the pNP release rate was monitored at 405 nm using a Cary-100 UV–visible spectrophotometer (Agilent Technologies). The release rate of α-D-mannopyranose-1-phosphate from 10 mM inorganic phosphate and 10 mM β-D-mannobiose (Megazyme, Ireland) was measured by quantification of α-D-mannopyranose-1-phosphate using high-performance anion-exchange chromatography with pulsed amperometric detection (HPAEC-PAD) as described previously (Ladevèze *et al.*, 2013).

2.3. Size-exclusion chromatography multi-angle laser light scattering (SEC-MALLS) experiments

A 30 µl sample of gel-filtered Uhgb_MP at a concentration of 6 mg ml^{-1} in 20 mM potassium phosphate pH 7.0, 150 mM NaCl was loaded onto a Superdex 200 HR 10/300 column (GE Healthcare, Massy, France) using an Agilent 1260 Infinity LC chromatographic system (Agilent Technology) coupled to a multi-angle laser light scattering (MALLS) detection system. The protein was centrifuged for 5 min at 4°C at 10 000g before the sample was loaded. The column was equilibrated with a 0.1 µm filtered buffer composed of 20 mM potassium phosphate pH 7.0, 150 mM NaCl. Separation was performed at a flow rate of 0.4 ml min^{-1} at 15°C. Data were collected using a DAWN HELEOS 8+ (eight-angle) light-scattering detector and an Optilab T-rEX refractive-index detector (Wyatt Technology, Toulouse, France). The results were analyzed using the *ASTRA* v.6.0.2.9 software (Wyatt Technology).

2.4. Protein crystallization

Purified Uhgb_MP protein was concentrated using poly-ethersulfone Vivaspin concentrators (Vivascience, Sartorius, Göttingen, Germany). The concentration was determined by measuring the $A_{280\,nm}$ using a NanoDrop instrument (Wilmington, Delaware, USA). All crystallization experiments were carried out at 12°C by the sitting-drop vapour-diffusion method using MRC 96-well microplates (Molecular Dimensions, Newmarket, England) and a Nanodrop ExtY crystallization instrument (Innovadyne Technologies, Santa Rosa, USA) to prepare 400 nl droplets. The best Uhgb_MP crystals were obtained within a week with a 1:1(v:v) ratio of protein (9–12 mg ml^{-1} in 20 mM potassium phosphate pH 7.0, 150 mM NaCl supplemented with 5 mM mannose and/or 5 mM N-acetylglucosamine for the co-crystallization assays) to precipitant solution [17.5–20%(w/v) polyethylene glycol 3350, 0.175–0.2 M ammonium chloride]. Uhgb_MP crystals grew to dimensions of $0.2 \times 0.08 \times 0.02$ mm in a week. They diffracted to a maximum resolution of 1.80 Å, while those obtained by co-crystallization with mannose, N-acetylglucosamine or both diffracted to maximum resolutions of 1.94, 1.60 and 1.76 Å, respectively.

2.5. Data collection and determination of the structure

X-ray experiments were carried out at 100 K. Crystals of Uhgb_MP were soaked for a few seconds in reservoir solution supplemented with 15%(v/v) glycerol (apo) or 15%(v/v) PEG 300 (complexes) prior to flash-cooling. Apo Uhgb_MP diffraction data sets were collected on beamline ID23-1 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, while those for the complexes were collected on the XALOC beamline at the ALBA Synchrotron, Cerdanyola del Vallès, Spain (Juanhuix *et al.*, 2014). The diffraction intensities were integrated and scaled using *XDS* (Kabsch, 2010) and 5% of the scaled amplitudes were randomly selected and excluded from the refinement procedure. All crystals belonged to the orthorhombic space group $P2_12_12_1$, with six molecules per asymmetric unit, giving Matthews coefficients of 2.22 and 2.11 Å^3 Da^{-1} and solvent contents of 44 and 42% for the apo forms and the three complexes, respectively. The structures were solved by the molecular-replacement method using *Phaser* (McCoy *et al.*, 2007) from the *CCP*4 software suite (Potterton *et al.*, 2003) and chain A of the crystal structure of *Tm*1225 from *T. maritima* MSB8 (PDB entry 1vkd) as a search model for the apo form. The final translation-function Z-score was 42.8 and the R and R_{free} values of the refined structure were 0.155 and 0.190, respectively. Once solved, the apo

Table 1
Data-collection and refinement statistics for Uhgb_MP.

Values in parentheses are for the outer resolution shell.

	Native (P$_i$)	Mannose + P$_i$	N-Acetylglucosamine + P$_i$	Mannose + N-acetylglucosamine + P$_i$
Data collection				
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å, °)	$a = 84.1$, $b = 141.2$, $c = 176.2$, $\alpha = \beta = \gamma = 90$	$a = 83.9$, $b = 140.8$, $c = 168.7$, $\alpha = \beta = \gamma = 90$	$a = 83.8$, $b = 140.9$, $c = 168.6$, $\alpha = \beta = \gamma = 90$	$a = 83.7$, $b = 140.9$, $c = 168.8$, $\alpha = \beta = \gamma = 90$
No. of molecules in asymmetric unit	6	6	6	6
Matthews coefficient (Å^3 Da^{-1})	2.22	2.11	2.11	2.11
Solvent content (%)	44.66	41.82	41.74	41.75
Wavelength (Å)	0.96863	0.97949	0.97949	0.97949
Resolution range (Å)	48.16–1.80 (1.91–1.80)	75.11–1.94 (1.98–1.94)	46.68–1.60 (1.64–1.60)	45.43–1.76 (1.80–1.76)
No. of unique reflections	190253 (27574)	148300 (21410)	261650 (41727)	197652 (31069)
No. of observed reflections	927534 (101770)	1217594 (168817)	2453350 (392431)	1818257 (276166)
Completeness (%)	98.00 (88.70)	99.66 (98.85)	99.84 (98.42)	99.64 (95.24)
Multiplicity	4.88 (3.69)	8.20 (7.90)	9.37 (7.00)	9.19 (8.88)
$\langle I/\sigma(I) \rangle$	10.46 (1.63)	9.30 (3.40)	14.49 (2.07)	13.70 (2.54)
R_{merge} (%)	9.1 (70.0)	15.5 (58.9)	9.3 (103.5)	11.9 (86.5)
Refinement				
R_{work}/R_{free}	0.157/0.190	0.152/0.193	0.155/0.183	0.158/0.192
Root-mean-square deviations				
Bond lengths (Å)	0.0188	0.0181	0.0191	0.0183
Bond angles (°)	1.9218	1.8528	1.8905	1.9235
Ramachandran plot				
Favoured (%)	91.6	91.3	91.3	91.1
Allowed (%)	8.1	7.9	8.5	8.6
B factors (Å^2)				
Wilson B	24	18	22	22
Mean	35	20	22	21
Main chain	33	18	20	19
Side chain	37	21	24	23
Ligand/water	28/37	23/26	26/30	27/26
PDB code	4udi	4udj	4udg	4udk

structure was then used to solve the protein–ligand structures. The structures of Uhgb_MP in complex with mannose, with N-acetylglucosamine and with mannose and N-acetyl-glucosamine were refined to final R/R_{free} values of 0.150/0.193, 0.154/0.183 and 0.158/0.192, respectively, using *REFMAC*5 (Murshudov *et al.*, 2011). Models were built manually in α_A-weighted electron-density maps using *Coot* (Emsley & Cowtan, 2004). Water molecules were manually checked after automatic assignment and ligand molecules were manually fitted in residual maps. Refinement statistics are listed in Table 1.

2.6. SAXS measurements

Small-angle X-ray scattering (SAXS) experiments were performed on the SWING beamline at the SOLEIL synchrotron, Gif-sur-Yvette, France. The wavelength was set to 1.033 Å. A 17 × 17 cm Aviex CCD detector was positioned 1800 mm from the sample, with the direct beam off-centred. The resulting exploitable q-range was 0.006–0.6 Å^{-1}, where $q = 4\pi\sin\theta/\lambda$, considering 2θ as the scattering angle. The samples were circulated in a thermostated quartz capillary with a diameter of 1.5 mm and 10 µm wall thickness positioned inside a vacuum chamber. A 80 µl volume of sample was injected onto a size-exclusion column (Bio SEC3 300, Agilent) equilibrated in phosphate-based buffer (20 mM potassium phos-

phate pH 7.0, 150 mM NaCl) or Tris-based buffer (20 mM Tris–HCl pH 7.0, 300 mM NaCl supplemented with 10% glycerol) using an Agilent high-performance liquid-chroma-tography (HPLC) system and eluted directly into the SAXS capillary cell at a flow rate of 200 µl min^{-1} at a temperature of 10°C. Samples were separated from the pushing liquid (water) by two air volumes of 6 µl each, as described previously (David & Pérez, 2009). SAXS data were collected online throughout the elution time and a total of 149 frames, each lasting 2 s, were recorded separated by a dead time of 0.5 s between frames. The transmitted intensity was continuously measured with an accuracy of 0.1% using a diode embedded in the beam stop. For each sample, the stability of the associated radius of gyration and the global curve shape in the frames corresponding to the main elution peak were checked, and the resulting selection of curves were averaged as described previously (David & Pérez, 2009). The recorded curves were normalized to the transmitted intensity and subsequently averaged using *Foxtrot*, a dedicated in-house application. The same protocol was applied to buffer scattering. R_g values were determined by a Guinier fit of the one-dimensional curves using the *ATSAS* package (Petoukhov *et al.*, 2007). The $P(r)$ function was calculated using the *GNOM* program and the corresponding *ab initio* envelopes were calculated using the *GASBOR* program. Rigid-body SAXS modelling was performed using the *CORAL* program.

3. Results and discussion

3.1. Conformational stability optimization of Uhgb_MP

Previous work on Uhgb_MP allowed the production and purification of a recombinant form of the protein in amounts suitable for crystallization (Ladevèze *et al.*, 2013). However, owing to enzyme instability, an optimized production system was set up by subcloning the open reading frame of Uhgb_MP into the pET-28a vector (Supporting Information §S1). Subcloning into pET-28a made it possible to produce a recombinant protein with a thrombin-cleavable N-terminal hexahistidine tag and a five-residue shortened linker between the tag and the N-terminal extremity of the native enzyme. After protein purification and processing in the same buffer as previously described (Ladevèze *et al.*, 2013), the activity on *p*NP-β-D-mannopyranose was increased by 73% to 10.9 × 10^{-3} µmol min^{-1} mg^{-1}, indicating that the 16-amino-acid linker used in the initial construct negatively impacted on the Uhgb_MP activity. To avoid the use of Tween 80, which is not suitable for crystallization assays, we then screened for an optimized buffer composition by differential scanning fluorimetry (DSF; Supporting Information §S1). In-house-prepared 96 deep-well screens adapted from Ericsson *et al.* (2006) were used to assess the effect of buffer nature, pH and NaCl concentration on protein thermal stability. The denaturation

curves revealed two fusion temperatures, $T_{m1} = 65.7°$C and $T_{m2} = 70.8°$C, in the initial Tris–HCl buffer, indicating that Uhgb_MP may adopt different oligomeric states in solution. The DSF results showed unambiguously that phosphate-based buffers (sodium and potassium phosphate) largely stabilize the Uhgb_MP structure at all of the assayed pH values (5.0, 5.5, 6.0, 6.5 and 7.0) and NaCl concentrations (136, 159, 287 and 439 m*M*). The best result was observed using 100 m*M* potassium phosphate pH 6.0 with 136 m*M* NaCl, leading to an increase in T_{m1} and T_{m2} of 6.82 ± 0.07 and 8.47 ± 0.07°C, respectively. We therefore chose to use potassium phosphate buffer supplemented with 150 m*M* NaCl to purify and store the protein produced using pET-28a::Uhgb_MP. In addition, the pH was set to 7.0 to allow sufficient separation efficiency of the protein in the affinity-purification step (T_{m1} and T_{m2} increased by 4.80 ± 0.30 and 6.45 ± 0.30°C, respectively, at this pH value). Under these optimal conditions, the protein production yield reached 90 mg pure protein per litre of culture. Finally, the β-1,4-D-mannobiose phosphorolytic activity of the enzyme stored in these conditions was increased tenfold compared with that of enzyme previously expressed in pDEST17 and purified in Tris buffer (Ladevèze *et al.*, 2013).

3.2. Crystallographic structure of Uhgb_MP subunits

The overall structure of Uhgb_MP was determined by molecular replacement using the structure of *Tm*1225 from *T. maritima* MSB8 (PDB entry 1vkd), which shares 60% identity with Uhgb_MP, as a model. The crystal structure of
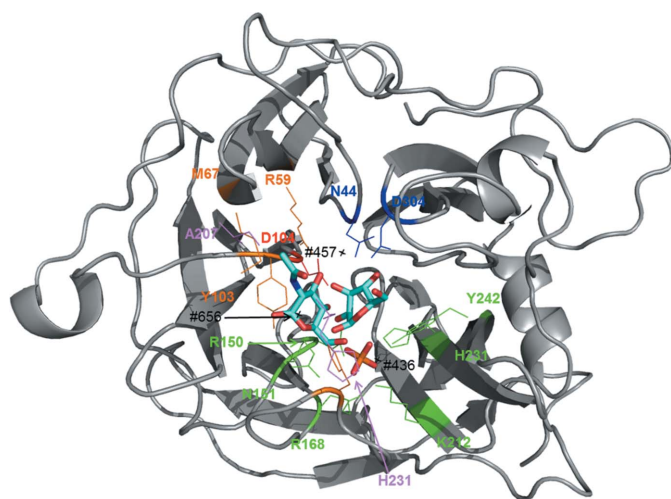

Figure 1
Monomeric Uhgb_MP fold, substrates and interacting residues. Like other GH130 enzymes, the Uhgb_MP monomer has a five-bladed β-propeller fold with a central catalytic furrow. The P$_i$, mannose and *N*-acetylglucosamine molecules present at the catalytic site are shown as sticks, while interacting residues are shown as lines. The catalytic Asp104 is shown in red, P$_i$-interacting residues in green, mannose-interacting residues in blue and *N*-acetylglucosamine-interacting residues in orange. The Asn44 and Asp104 side chains are shown in the *B* conformation, *i.e.* the conformation that is catalytically active. Water molecules 436, 457 and 656, which mediate interactions between the *N*-acetylglucosamine and residues Lys212 and Tyr242, Asp304 and His235, respectively, are shown as black crosses. Protein–P$_i$ interactions: NH$_2$ of Arg150 is contacting P$_i$ O^4 and O^2, while NH$_2$ of Asn151 contacts P$_i$ O^4. ωNH$_2$ of the Arg168 side chain binds to P$_i$ O^1 and its ω'NH$_2$ is interacting with P$_i$ O^2. The side-chain amine of Lys212 binds to P$_i$ O^2 and His231 N$^{\varepsilon 2}$ is at a hydrogen-bond distance from P$_i$ O^1. Finally, the hydroxyl of the Tyr242 side chain is in contact with P$_i$ O^3.
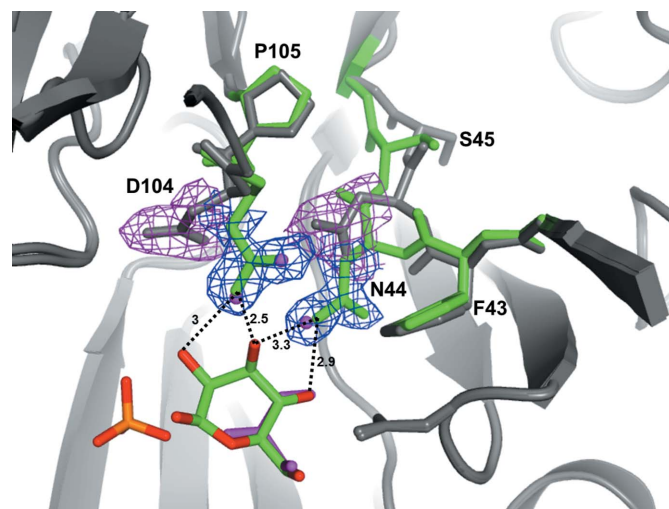

Figure 2
Alternative conformations of Asn44, Ser45 and the catalytic Asp104 upon mannose binding in the −1 subsite. Superposition of the apo structure of Uhgb_MP (PDB entry 4udi) and Uhgb_MP complexed with mannose (PDB entry 4udj), illustrating the movement of the catalytic residues when mannose is bound at the −1 subsite. The backbone of the apo form is shown in grey (*A* conformation), while the backbone of the complexed, catalytically active form (*B* conformation) is in green. Water, phosphate and glycerol molecules in the apo form are shown. $2F_o − F_c$ electron-density maps are shown (contoured at 1.0σ) for the catalytic residues in the apo and mannose-bound forms. Interatomic distances are labelled in Å.

Uhgb_MP revealed a homohexameric organization. The homohexamer consisted of a trimer of dimers with $D3$ symmetry, with six molecules per asymmetric unit. The apo structure was refined to 1.80 Å resolution, while the complexes with mannose, with *N*-acetylglucosamine and with the two combined were refined to 1.94, 1.60 and 1.76 Å, respectively (Table 1). The electron-density maps did not enable construction of the N-terminal extremity of the polypeptide chains. The N-terminal hexahistidine tag and the following 6–8 first residues have thus been omitted from the final model. The overall fold of each Uhgb_MP protomer consists of a five-bladed β-propeller (Fig. 1). The catalytic centre is located in the central cleft as previously hypothesized (Ladevèze *et al.*, 2013), as a phosphate ion (P_i) and mannose and *N*-acetylglucosamine were observed in the central furrow. P_i is deeply buried in the catalytic site, strongly stabilized by hydrogen bonds and ionic interactions with the surrounding residues (Fig. 1). Compared with other glycoside phosphorylases, P_i appeared to be quite strongly bound, since the P_i dissociation constant values previously determined for the ternary enzyme–P_i–mannobiose and enzyme–P_i–β-D-Man*p*-1,4-β-D-Glc*p*NAc-1,4-D-Glc*p*NAc complexes (0.64 and 0.13 mM, respectively; Ladevèze *et al.*, 2013) are more than 200 times lower than that determined for *Ra*MP1 (belonging to the GH130_1 subfamily), the only other GH130 enzyme for which a P_i dissociation constant has been determined. In the apo structure, a molecule of glycerol, which was used as a cryoprotectant, occupied the −1 subsite. This glycerol molecule was hydrogen-bonded to the side-chain carboxylate moiety of Asp304 ($O^{\delta1}$ to O_3 and $O^{\delta2}$ to O_1), while its O_2 atom was hydrogen-bonded to the P_i, mimicking the interactions between mannose and the surrounding amino acids that were observed in the protein–ligand complexes. Interestingly, in the Uhgb_MP structures where binding of mannose in the −1 subsite occurred (PDB entries 4udj and 4udk in Table 1), the mannose ring was found in a stressed $B_{2,5}$ boat conformation stabilized by hydrogen bonding to Asp304 ($O^{\delta1}$ to the C_6 hydroxyl and $O^{\delta2}$ to the C_4 hydroxyl); the mutation of this critical residue to asparagine abolishes 96% of the activity (Ladevèze *et al.*, 2013). This unusual conformation of mannose was present in all six chains (Supplementary Fig. S1) and was previously observed in the *Bf*MGP structures (PDB entry 3was, for which *in cristallo* activity was observed, and PDB entry 3wat). It should be noted that this $B_{2,5}$ boat conformation is less unstable in mannose compared with other monosaccharides owing to the pseudo-equatorial position of the C_2–OH, which is in an *anti* configuration to the ring O atom, thus bending the C_3–OH towards the β-glycosidic O atom, in a *syn* axial position. The binding of mannose in the −1 subsite induced a large conformational movement in the active site. Indeed, in the apo form, where a glycerol molecule was observed in place of mannose, the amino moiety of Asn44 interacts with the C_3 hydroxyl of glycerol through a water molecule (conformation *A* in Fig. 2). Upon mannose binding (conformation *B*), the peptide bond between Phe43 and Asn44 flips in order to allow direct interaction of the Asn44 side chain with the C_3 and C_4 hydroxyls of the sugar. The side

chain of the catalytic residue Asp104, mutation of which to asparagine completely abolished the activity (Ladevèze *et al.*, 2013), is also moved towards mannose in a position that is occupied by two water molecules in the apo form. In this *B* conformation, Asn44 stabilizes the catalytic Asp104 through hydrogen bonding, thereby imposing selection of the rotamer facing the mannose C_3 hydroxyl, which acts as a proton relay during catalysis (Fig. 2). This is the first time that such a concerted movement upon substrate binding in the −1 subsite has been reported for a glycoside phosphorylase. It must be noted that the *B* conformation of the catalytic Asp104 is probably the one that is active since it has also been observed for *Bf*MGP, which was demonstrated to be catalytically active *in cristallo* (Nakae *et al.*, 2013). The *A* conformation of the catalytic residue Asp104 of Uhgb_MP was observed in the structure of the GH130_2 *Tm*1225 protein in the apo form (PDB entry 1vkd). In the Uhgb_MP structures, the *B* conformation was only observed when mannose was bound in the −1 subsite, and is therefore independent of the presence of *N*-acetylglucosamine in the +1 subsite. Indeed, the *A* conformation was observed in the complex with *N*-acetylglucosamine alone (PDB entry 4udg, with glycerol in the −1 subsite and *N*-acetylglucosamine in the +1 subsite), while the *B* conformation was observed in the complex with mannose and *N*-acetylglucosamine (PDB entry 4udk, with mannose in the −1 subsite and *N*-acetylglucosamine in the +1 subsite). In the *A* and *B* conformations, *N*-acetylglucosamine bound in the +1 subsite was found in a $^4C_1$ relaxed chair conformation, stacked with Tyr103, and bound through hydrogen bonding to the C_6 hydroxyl group, the His174 imidazole ring, the Lys212 side chain and the Tyr242 hydroxyl group *via* a water molecule. The C_3 hydroxyl interacts with the Arg59 side-chain amine moiety, as well as with Asp304 $O^{\delta1}$ through a water molecule. The *N*-acetyl moiety is also involved in binding through hydrogen bonding between its NH group and the S atom of Met67 and between its carbonyl moiety and the His235 carbonyl *via* a water molecule. The hydrophobic methyl moiety faces the side chain of Ala207 and Met67, with these residues forming a hydrophobic pocket (Fig. 1).

These data enabled us to revise our previous Uhgb_MP model, which was built using the atomic coordinates of the *Tm*1225 protein from *T. maritima* MSB8 (PDB entry 1vkd) as a structural template, considering a monomeric form of the enzyme and using geometrical constraints provided by a classical inverting GH-like single-displacement mechanism. In this model, we previously hypothesized a +1 subsite formed by the Tyr103, Asp304, His174, Tyr240 and Phe283 residues, which are specifically conserved in the GH130_2 family, while the +2 subsite would be delineated by Tyr242, Pro279, Asn280 and Asp304. In this configuration, the exit of the catalytic tunnel would be orientated towards the inside of the oligomeric structure, thereby reducing access to the catalytic site. In addition, the conserved His235–Tyr240 loop from a cognate monomer would block the furrow that we have hypothesized. These new data emphasize the importance of taking into account the quaternary organization when modelling oligomeric enzymes, using SAXS data when possible to define the

low-resolution envelope and avoiding restraining the possible docking modes of substrates too much to envisage original reaction mechanisms. Here, thanks to the high-resolution crystallographic structures of hexameric Uhgb_MP, which is catalytically active (as detailed in the next section), in complex with mannose and *N*-acetylglucosamine, we propose a revised active-site topology in which the oligosaccharide chain rotates by 180°, inverting the −1 and +1 subsite positions. This new orientation allows the substrate to enter from the open side of the tunnel, and is in complete accordance with the orientation found in the crystal structure of *Bf*MGP in complex with its substrates (Nakae *et al.*, 2013).

As previously shown by kinetic analysis of Uhgb_MP, phosphorolysis of the N-glycan oligosaccharide core follows a mixed-type sequential random Bi-Bi mechanism (Ladevèze *et al.*, 2013). However, the order of substrate binding was not determined. Functional and structural data now lead us to suggest that the phosphorolytic catalytic mechanism is composed of a first step in which the inorganic phosphate is conveyed to the catalytic centre, followed by entry of the substrate to be phosphorolyzed. Indeed, the P_i binding site is located deeper in the catalytic site than the glycoside substrate, meaning that P_i could not bind after the disaccharide. Thus, mannose binding in the −1 subsite would induce a flip of the Phe43–Asn44 peptide bond from conformation *A* to conformation *B*, thus maintaining the side chain of Asp104 in a catalytically competent configuration. In reverse phosphorolysis, entry of mannose-1-phosphate would be the first step, followed by conformational changes of Phe43,

Asn44, Ser45 and Asp104. Entrance of the acceptor would lead to the reverse phosphorolysis reaction. Regarding the phosphorolytic reaction itself, our data show that when the mannosyl moiety was present in the catalytic site, no water molecule was located where it could relay the proton from the catalytic Asp104 to the interosidic O atom. In addition, comparison between the apo and complexed forms showed that the change in Asp104 from conformation *A* to conformation *B* did not allow the catalytic aspartic acid to be at a hydrogen-bonding distance from the interosidic O atom, indicating that this residue is not directly involved in proton transfer to the interosidic O atom, as seen in the *Bf*MGP structures. In the latter case, Nakae and coworkers suggested a catalytic mechanism different from that of known inverting glycoside phosphorylases (GPs), involving the assistance of C_3—OH to relay proton transfer, because, like us, they did not observe a water molecule at the catalytic site in any of their structures (Supplementary Fig. S2). The stressed $B_{2,5}$ boat conformation of mannose bound in the −1 subsite and the *B* configuration of Asp104 which is only stabilized when mannose present in the −1 subsite was compatible with such a mechanism. The first step of phosphorolysis would thus be (i) nucleophilic attack of P_i on C_1 of the mannosyl moiety bound in the −1 subsite, (ii) a two-step protonation through Asp104 $O^{\delta 2}$–Man O_3–GlcNac O_4, with the Asp104 side-chain carboxylic acid being located at 2.5 Å from the C_3–OH (Supplementary Fig. S2). However, even if the catalytic mechanism of Uhgb_MP and *Bf*MGP appears to be identical, clear differences in the substrate specificity of GH130
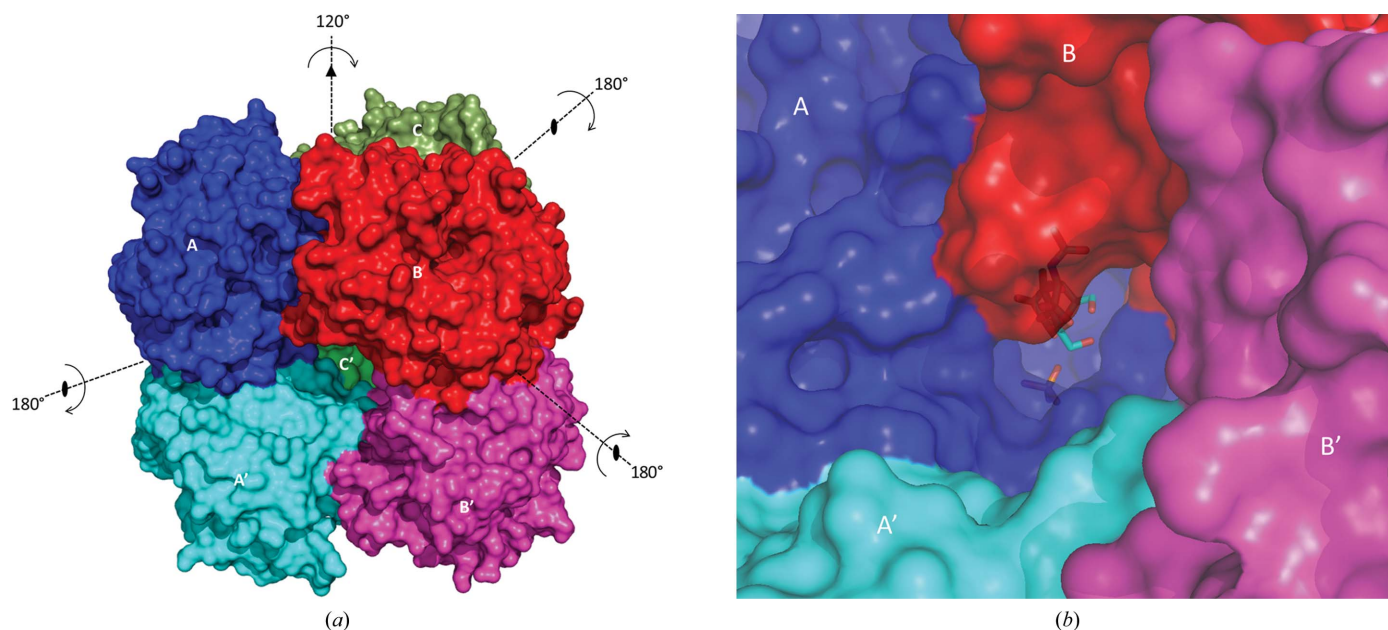
Figure 3
Uhgb_MP homohexameric structure. (*a*) Uhgb_MP is a hexameric structure formed by a trimer of dimers. Individual monomers are shown in a single colour and are labelled *A*, *B*, *C*, *A'*, *B'* and *C'* for clarity. Each pair of dimers is coloured in pale/dark blue, green and red. (*b*) Close-up of the catalytic tunnel of a single Uhgb_MP protomer in the hexameric structure. The quaternary-structure assembly imposes structural constraints on active-site accessibility. The inorganic phosphate, the mannose and the *N*-acetylglucosamine molecules present in the catalytic site of protomer *A* are shown as sticks deeply buried in the catalytic site, which is accessible by a tunnel whose sides are formed by different protomers. The extremity of the tunnel is located at the centre of the plane formed by four Uhgb_MP molecules, in this case *A*, *A'*, *B* and *B'*.

subfamilies have been identified owing to structural motifs located in other parts of the protein, as further detailed below.

3.3. Uhgb_MP quaternary structure

In the Uhgb_MP crystal structure, each subunit of the homohexamer is roughly globular in shape. There is a large cavity inside the homohexamer and large holes at the centre of each of the three lateral planes formed by the homohexamer assembly (Fig. 3a). No discretely bound solvent molecules were found in these cavities, which are 15 Å in diameter at their narrowest point, indicating that these holes are large enough to enable substrate access to the active site of each protomer. The association of the surrounding subunits in the homohexamer caps the furrow of each protomer, giving rise to a funnel whose entrance is orientated towards the lateral aperture (Fig. 3b). The catalytic site is therefore deeply buried, with the phosphate ion and the −1 subsite located at the bottom of the tunnel.

Each dimer is formed of a large buried surface area of 1300 Å^2 involving two Uhgb_MP molecules linked by twofold symmetry (Fig. 4a). The interactions promoting dimer association involve the side chains of the residues at the interface, such as His123 and His196 stacking, and several hydrogen bonds or salt bridges involving side-chain atoms, such as between Arg195 and Gln142, Glu142 and Arg193, Asp93 and Arg195, and Glu189 and Arg193. Other polar interactions involve main-chain atoms of Ala144 and His194, Tyr122 and

His196, and Trp191 and a symmetry mate. The homohexamer is formed by the association of three dimers arranged around a pseudo-threefold axis. Each dimer is related to its neighbours through symmetrical interactions involving each of the two protomers (covering an interaction surface of 840 Å^2 each; Fig. 4b). These interactions involve T-shaped stacking between the imidazole groups of His174 and His235 and hydrogen-bond interactions between side chains between Asn40 and Tyr264 and between Ser64 and Pro263. Main-chain carbonyl and amino groups are also involved in the assembly. More precisely, the Thr63 carbonyl is in contact with Tyr264 NH, while the side chains of Asn238 and Asn280 interact with the carbonyl group of Gly276 and the amide N atom of Tyr240, respectively. Finally, the side chain of Asn238 makes contact with the Pro279 carbonyl moiety.

SEC-MALLS and SAXS analysis confirmed the hexameric organization of Uhgb_MP in solution. The protein apparent molecular mass determined by SEC-MALLS was 240 kDa (n = 6.16; Supplementary Fig. S3). Guinier analysis of the SAXS data revealed that in phosphate and Tris–glycerol buffers, the radius of gyration (R_g) was considerably larger than the theoretical R_g, indicating protein aggregation. Based on data collected in Tris–glycerol buffer in the presence of 1 mM TCEP as reducing agent, an R_g value of 37.9 ± 0.08 Å was obtained, which is in good agreement with the theoretical R_g calculated from the apo crystal structure (∼37 Å). The pair distribution function $P(r)$ revealed a compact particle with a D_{max} of ∼110 Å that closely matches the largest dimension of
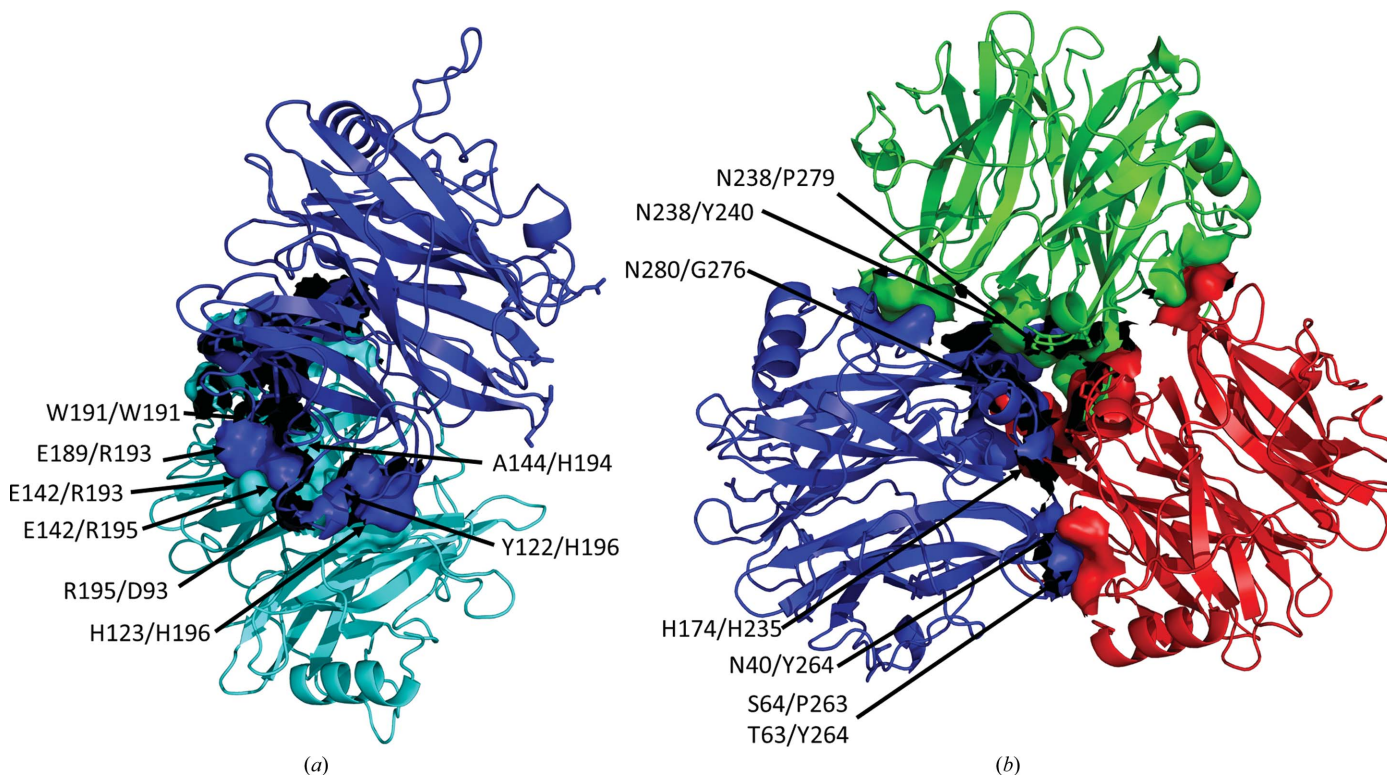
Figure 4
Interaction surfaces between the different Uhgb_MP protomers. (a) The interaction surfaces between two Uhgb_MP protomers involved in dimer formation. (b) The interaction surfaces between Uhgb_MP dimers to form the hexamer. Only the interactions involving the three upper monomers from each dimer are shown here, in order to clarify the view, as these surfaces are symmetrical in the lower monomers.

the hexamer (Supplementary Fig. S4). The *ab initio* envelope confirmed the compact shape of Uhgb_MP in solution with a trimer-of-dimers organization superimposable with the crystallographic structure. SAXS-based rigid-body modelling was attempted by taking into account the flexibility of the N-terminal residues missing from the crystal structure. The fit showed that the theoretical curve closely matched the experimental data, thus confirming the general hexameric organization of Uhgb_MP in solution (Supplementary Fig. S4).

Considering that (i) we never observed the existence of Uhgb_MP monomers, either in solution or in crystals, (ii) the quaternary structures deduced from both SAXS and crystallographic data are superimposable and (iii) the intracellular P_i concentration in bacteria is around 10 mM (Motomura *et al.*, 2011), meaning that P_i binds to Uhgb_MP *in vivo* as in the crystals, we conclude that hexamerization is required for enzyme activity and that the crystal structure presented here is certainly the most probable organization under physiological conditions.

All other known GPs belonging to the GH3, GH13, GH65, GH94, GH112, GT4 and GT35 families crystallize and act as homodimers. In contrast, it is difficult to find general features that control the oligomerization of GH130 enzymes, even for those belonging to the same subfamily. Indeed, all of the data that we gathered on functionally or structurally characterized GH130 enzymes showed that no single subfamily contained homogenous oligomerization profiles. Enzymes belonging to the GH130_NC group (the BACOVA_03624, BACOVA_02161, BT4094 and BDI_3141 proteins) crystallized as monomers, while the two functionally characterized proteins Teth514_1788 and Teth514_01789 have been shown to be dimeric and monomeric in solution, respectively (Chiku *et al.*, 2014). Various oligomeric forms (in solution or crystals) have been found in the GH130_1 (hexameric *Bf*MGP, pentameric

*Rm*MGP and dimeric *Ra*MP1) and GH130_2 (hexameric Uhgb_MP, dimeric *Tm*1225, hexameric *Ra*MP2 and tetrameric *Bt*1033) subfamilies.

Moreover, the presence of the AxxxAxxxA motif in the *Bf*MGP N-terminal helix, which was thought to mediate oligomerization (Nakae *et al.*, 2013), was not found in *Ra*MP1 or *Rm*MGP, demonstrating that this particular motif is not the only element that is able to promote the formation of oligomers in enzymes belonging to the same GH130 subfamily. Finally, no particular secondary-structure element appeared to mediate interactions between the different Uhgb_MP protomers, which were only associated by surface interactions, without any involvement of secondary-structure elements, in contrast to what was observed for *Bf*MGP (the only GP of known structure with a similar homohexameric conformation). Indeed, the *Bf*MGP loop Thr42–Met68, which was found in contact with the N-terminal tab helix and is thought to help contact the cognate protomer, is completely absent in Uhgb_MP, and more generally in all GH130_2 sequences. Taken together, these data highlight the structural originality of GH130 enzymes among glycoside phosphorylases. Nevertheless, the solutions of many other GH130 structures will be required to be able to highlight possible structural markers of oligomerization.

3.4. Molecular bases of specificity towards mannosides

Uhgb_MP is the first member of the GH130_2 subfamily to be characterized both functionally and structurally. *Tm*1225 has only been structurally characterized, and no structural data is available for the functionally characterized GH130_2 members *Ra*MP2 and *Bt*1033. In contrast to the known enzymes classified into the GH130_1 subfamily (including *Bf*MGP, the only crystallized member of this subfamily; Senoura *et al.*, 2011; Nakae *et al.*, 2013), which exhibit a narrow
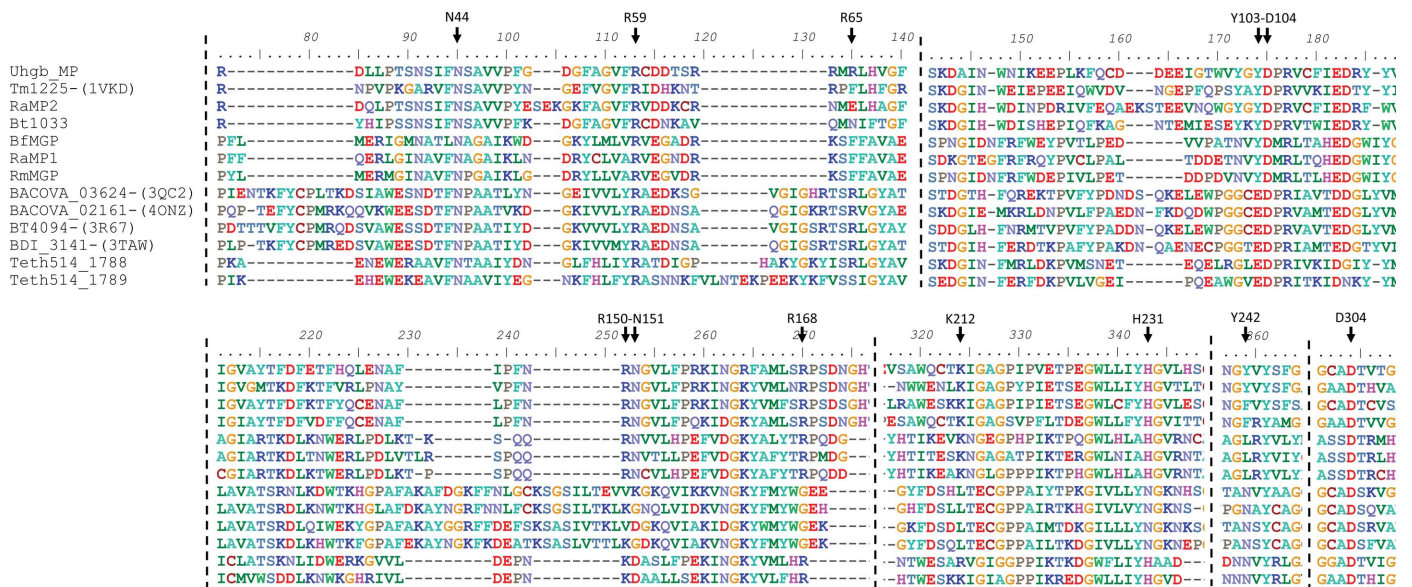
Figure 5
Sequence alignment of characterized GH130 enzymes. This alignment highlights the conservation of the catalytic and substrate-interacting residues among characterized GH130 enzymes, as well as family-specific loops.

specificity towards β-ᴅ-Manp-1,4-ᴅ-Glc, GH130_2 enzymes present a highly relaxed substrate specificity. Moreover, the Uhgb_MP structure is the first structure of an inverting glycoside phosphorylase that is active on a polysaccharide. Indeed, the only enzymes of known structure that are able to phosphorolyze polysaccharides are the retaining α-maltosyl phosphate:α-1,4-ᴅ-glucan-4-α-ᴅ-maltosyltransferases belonging to GH13 and glycogen or starch phosphorylases belonging to the GT35 family (Egloff *et al.*, 2001; Mirza *et al.*, 2006).

Uhgb_MP structures were compared with the six structures available for GH130 enzymes in order to identify any structural features that could explain the enzyme specificity, in particular towards N-glycan oligosaccharides, long manno-oligosaccharides and mannans. The residues involved in the catalytic machinery and in substrate binding are highly conserved in both the sequences and the three-dimensional structures, with the notable exception of the side chains of Asp104, Asn44 and Ser45 in the *A* configuration (Fig. 5 and Fig. 6), with overall r.m.s.d. values after C^α superposition of 2.1, 2.1, 2.1, 2.0, 2.0 and 1.0 Å for *Bf*MGP (25% identity with Uhgb_MP), BACOVA_03624 (23% identity), BACOVA_ 02161 (22% identity), *Bt*4094 (23% identity), BDI_3141 (25% identity) and *Tm*1225 (61% identity), respectively (Fig. 6). In addition, the P_i and glycosyl moieties in the −1 and +1 subsites superimposed perfectly with those present in the structure of *Bf*MGP in complex with inorganic phosphate, mannose and glucose. The electron density of mannose was separated from that of *N*-acetylglucosamine, in contrast to what would have been observed for the disaccharide β-ᴅ-Manp-1,4-ᴅ-Glc*p*NAc, because of the impossible superimposition of mannose O_1 and *N*-acetylglucosamine O_4. However, some structural features that are conserved in the subfamily explain the differences in substrate specificities observed between subfamilies (Fig. 6).

The most significant structural changes were identified in the Uhgb_MP Gly121–Pro125 loop, which is 11 residues longer in GH130_1 enzymes compared with those belonging to the GH130_2 and GH130_NC clusters (Fig. 6). This longer loop appeared at the extremity of the catalytic tunnel and, in the case of *Bf*MGP, actually filled it. Therefore, the accommodation of longer substrates than disaccharide would be impossible for GH130_1 enzymes, which is in accordance with the biochemical data published to date (Senoura *et al.*, 2011; Kawahara *et al.*, 2012; Jaito *et al.*, 2014). In GH130_2 enzymes the shorter loop would enable the entry of longer substrates, such as long manno-oligosaccharides or even mannans for Uhgb_MP, into the large cavity formed inside the quaternary structure between the three lateral planes of the homohexamer. Moreover, the +1 subsite flexibility of the GH130_2 members, which are able to accommodate *N*-acetylglucosamine and the C_2 epimer of glucose, would be explained by the location of the Uhgb_MP Arg65 residue. Indeed, the arginine side chain is at a distance of 6.43 Å from the O_2 atom compared with 2.96 Å for the side chain of the corresponding *Bf*MGP residue, Arg94, which would be responsible for the specificity of the GH130_1 members for β-ᴅ-Manp-1,4-ᴅ-Glc through hydrogen bonding to O_2 of glucose.

In addition, the stronger specificity of GH130_2 towards *N*-acetylglucosamine at the +1 subsite compared with glucose or mannose is explained by hydrophobic interactions of the *N*-acetyl-methyl moiety with Met67, which is not conserved in the GH130_1 subfamily,
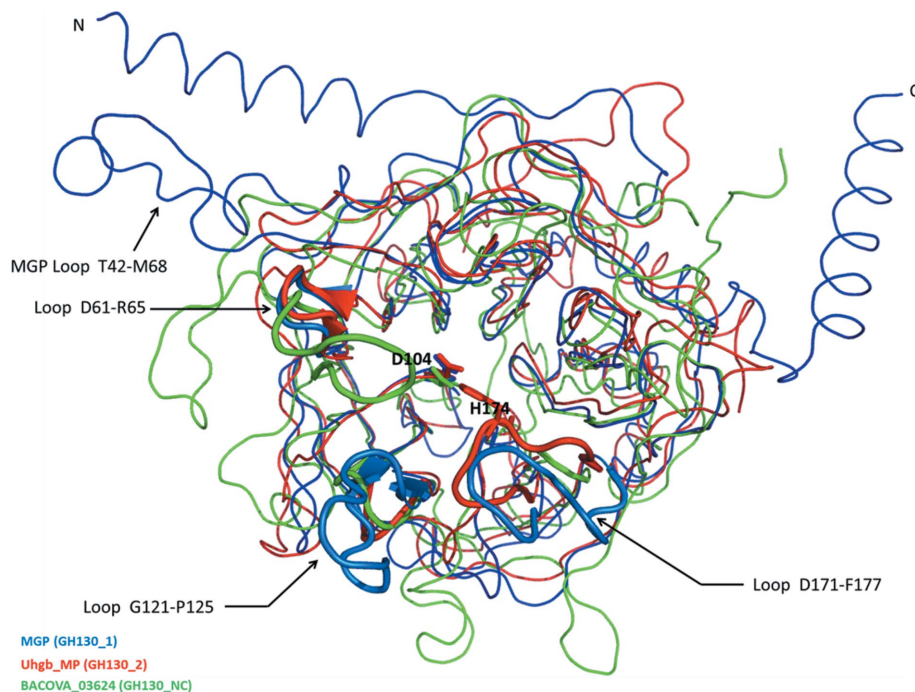
Figure 6
Superposition of GH130 structures. Superposition of Uhgb_MP (GH130_2; red) with mannosyl-glucose phosphorylase (*Bf*MGP) from *B. fragilis* NCTC 9343 (GH130_1; PDB entry 4kmi; blue), and BACOVA_03624 from *Bacteroides ovatus* ATCC 8483 (GH130_NC; PDB entry 3qc2; green), illustrating the structural differences between GH130 subfamilies. The catalytic Asp104 is shown in the *B* conformation. With the exception of the *Bf*MGP loop Thr42–Met68, which does not exist in the Uhgb_MP structure, the loops are numbered with respect to the Uhgb_MP sequence. The 11-residue longer Gly121–Pro125 loop, which is specific to GH130_1, fills the entrance to the Uhgb_MP tunnel. In place of the Asp61–Arg65 loop, an extension is observed for GH130_NC, capping the catalytic site. These two loops, which are specific to GH130_1 and GH130_NC, respectively, may explain the inability of enzymes belonging to these subfamilies to phosphorolyze long substrates. Loop Asp171–Phe177 (the so-called 'lid loop' in *Bf*MGP) is shorter in GH130_NC enzymes than in GH130_1 and GH130_2, thus allowing access to the active site. In addition, in GH130_1 enzymes loop Asp171–Phe177 is very mobile because of the GSGGG motif located at its base, which is locked close to the catalytic site only when a substrate is bound, as shown for *Bf*MGP structures. In contrast, in Uhgb_MP the loop is not so mobile and holds His174, which is conserved in GH130_2 and which has already been shown to be involved in the +1 subsite. The *Bf*MGP loop Thr42–Met68 in contact with the N-terminal helix involved in oligomerization is completely absent in Uhgb_MP even when both proteins are assembled as homohexamers.

while being present in half of the GH130_2 members, especially those acting on β-D-Man$p$-1,4-D-GlcNAc (Bt1033 and RaMP2).

Moreover, in the structures containing N-acetylglucosamine, the Phe203 side chain of a cognate monomer was found close to Met67; these two residues form a hydrophobic pocket that interacts with the methyl group of the N-acetyl moiety. On the contrary, in the apo form and in the structure containing mannose alone in the −1 subsite, the Phe203 side chain was rather found rotated towards the exit of the catalytic tunnel. Therefore, Phe203, which is not conserved in the GH130_1 subfamily, while being present in 25% of GH130_2 members, would thus be a specific feature of GH130 enzymes that are able to phosphorolyze β-D-Man$p$-1,4-β-D-Glc$p$NAc.

The C_3 stereochemistry at the +1 subsite also appears to be critical since all pyranoside inhibitors of Uhgb_MP (allose, L-rhamnose and altrose) share an inversion of configuration at this position compared with that of mannose. This effect is probably owing to the proximity of Tyr103 (or the equivalent Tyr130 in BfMGP), thus implying a steric constraint that would select an equatorial hydroxyl at this position in the case of a $^4C_1$ chair, which is the case for N-acetylglucosamine in the structure of the corresponding complex. We previously observed that a Y103E mutation strongly destabilizes Uhgb_MP, as is the case for the wild-type enzyme without phosphate. The Y103E mutation also significantly increases the ratio between hydrolysis and phosphorolysis, with the glutamic acid playing the role of the second catalytic acid required for hydrolysis (Ladevèze *et al.*, 2013). The role of Tyr103 in stabilizing the active-site conformation in the presence of phosphate is owing to hydrogen-bonding interactions between its lateral chain and that of Arg150, which interacts with phosphate (Fig. 1). The Y103E mutation would decrease hydrogen-bonding interactions, while positioning the glutamic acid at a hydrogen-bonding distance (less than 4 Å) from the interosidic O atom to allow hydrolysis to occur.

3.5. Significance

In this paper, we present the first structure of a phosphorolytic enzyme involved in N-glycan degradation in its apo form and in complex with mannose and N-acetylglucosamine. As previously highlighted by the integration of biochemical, genomic and metagenomic data, Uhgb_MP and GH130 enzymes more generally can be considered as interesting targets to study interactions between host and gut microbes, especially since GH130_2 sequences are overrepresented in the metagenomes of IBD patients. Further studies will be needed to confirm the physiological role of these enzymes and their potential involvement in damage to the intestinal barrier, such as metabolomic and transcriptomic analyses of the gut bacteria that produce them in the presence of N-glycans as a carbon source or inoculated in model animals with and without inhibitors. This three-dimensional structure paves the way for such studies through the design of specific GH130_2 inhibitors that could mimic substrate binding.

In addition, analysis of tertiary and quaternary structures led to the identification of structural features involved in the accommodation of long oligosaccharides and polysaccharides. This is a key feature that is unique to Uhgb_MP and is of great biotechnological interest for the conversion of hemicellulose into compounds with high added value. Identification of the structural determinants of the strong specificity of Uhgb_MP and other GH130_2 enzymes towards Man-GlcNAc also paves the way for the rational engineering of GH130 enzymes optimized for manno-oligosaccharide synthesis and diversification. Functional investigations of structurally characterized enzymes classified in the different GH130 sequence clusters would significantly advance our knowledge of the molecular bases of substrate specificities and improve our understanding of their role in key catabolic pathways, especially in the mammalian gut.

4. Related literature

The following references are cited in the Supporting Information for this article: Studier (2005).

References

Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. (2010). *Trends Biochem. Sci.* **35**, 74–82.

Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H. & Henrissat, B. (2012). *BMC Evol. Biol.* **12**, 186.

Burnaugh, A. M., Frantz, L. J. & King, S. J. (2008). *J. Bacteriol.* **190**, 221–230.

Chiku, K., Nihira, T., Suzuki, E., Nishimoto, M., Kitaoka, M., Ohtsubo, K. & Nakai, H. (2014). *PLoS One*, **9**, e114882.

David, G. & Pérez, J. (2009). *J. Appl. Cryst.* **42**, 892–900.

research papers

Egloff, M. P., Uppenberg, J., Haalck, L. & van Tilbeurgh, H. (2001). *Structure*, **9**, 689–697.

Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.

Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. (2006). *Anal. Biochem.* **357**, 289–298.

Jaito, N., Saburi, W., Odaka, R., Kido, Y., Hamura, K., Nishimoto, M., Kitaoka, M., Matsui, H. & Mori, H. (2014). *Biosci. Biotechnol. Biochem.* **78**, 263–270.

Juanhuix, J., Gil-Ortiz, F., Cuní, G., Colldelram, C., Nicolás, J., Lidón, J., Boter, E., Ruget, C., Ferrer, S. & Benach, J. (2014). *J. Synchrotron Radiat.* **21**, 679–689.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kawahara, R., Saburi, W., Odaka, R., Taguchi, H., Ito, S., Mori, H. & Matsui, H. (2012). *J. Biol. Chem.* **287**, 42389–42399.

Ladevèze, S., Tarquis, L., Cecchini, D. A., Bercovici, J., André, I., Topham, C. M., Morel, S., Laville, E., Monsan, P., Lombard, V., Henrissat, B. & Potocki-Véronèse, G. (2013). *J. Biol. Chem.* **288**, 32370–32383.

Ladevèze, S., Tarquis, L., Henrissat, B., Monsan, P., Laville, E. & Potocki-Veronese, G. (2014). International Patent WO/2015/014973.

Larkin, A. & Imperiali, B. (2011). *Biochemistry*, **50**, 4411–4426.

Lehle, L., Strahl, S. & Tanner, W. (2006). *Angew. Chem. Int. Ed.* **45**, 6802–6818.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.

Martens, E. C., Chiang, H. C. & Gordon, J. I. (2008). *Cell Host Microbe*, **4**, 447–457.

Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon, J. I. (2009). *J. Biol. Chem.* **284**, 24673–24677.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Mirza, O., Skov, L. K., Sprogøe, D., van den Broek, L. A. M., Beldman, G., Kastrup, J. S. & Gajhede, M. (2006). *J. Biol. Chem.* **281**, 35576–35584.

Motomura, K., Hirota, R., Ohnaka, N., Okada, M., Ikeda, T., Morohoshi, T., Ohtake, H. & Kuroda, A. (2011). *FEMS Microbiol. Lett.* **320**, 25–32.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.

Nagae, M. & Yamaguchi, Y. (2012). *Int. J. Mol. Sci.* **13**, 8398–8429.

Nakae, S., Ito, S., Higa, M., Senoura, T., Wasaki, J., Hijikata, A., Shionyu, M., Ito, S. & Shirai, T. (2013). *J. Mol. Biol.* **425**, 4468–4478.

Nihira, T., Suzuki, E., Kitaoka, M., Nishimoto, M., Ohtsubo, K. & Nakai, H. (2013). *J. Biol. Chem.* **288**, 27366–27374.

Petoukhov, M. V., Konarev, P. V., Kikhney, A. G. & Svergun, D. I. (2007). *J. Appl. Cryst.* **40**, s223–s228.

Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. (2003). *Acta Cryst.* D**59**, 1131–1137.

Renzi, F., Manfredi, P., Mally, M., Moes, S., Jenö, P. & Cornelis, G. R. (2011). *PLoS Pathog.* **7**, e1002118.

Roberts, G., Tarelli, E., Homer, K. A., Philpott-Howard, J. & Beighton, D. (2000). *J. Bacteriol.* **182**, 882–890.

Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H., Ozawa, T., Jin, S., Watanabe, J., Wasaki, J. & Ito, S. (2011). *Biochem. Biophys. Res. Commun.* **408**, 701–706.

Sheng, Y. H., Hasnain, S. Z., Florin, T. H. J. & McGuckin, M. A. (2012). *J. Gastroenterol. Hepatol.* **27**, 28–38.

Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

Suzuki, T. & Harada, Y. (2014). *Biochem. Biophys. Res. Commun.* **453**, 213–219.

Varki, A., Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W. & Etzler, M. E. (2009). Editors. *Essentials of Glycobiology*, 2nd ed. New York: Cold Spring Harbor Laboratory Press.

Zhu, Y., Suits, M. D. L., Thompson, A. J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K. W., Xiang, Y., Siriwardena, A., Williams, S. J., Gilbert, H. J. & Davies, G. J. (2010). *Nature Chem. Biol.* **6**, 125–132.