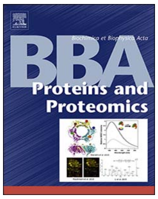
BBA - Proteins and Proteomics 1866 (2018) 292–306

Contents lists available at ScienceDirect

BBA - Proteins and Proteomics

journal homepage: www.elsevier.com/locate/bbapap



Structural insights of *Rm*Xyn10A – A prebiotic-producing GH10 xylanase with a non-conserved aglycone binding region

Anna Aronsson[a,*], Fatma Güler[a,b,c], Maxim V. Petoukhov[d,e,f], Susan J. Crennell[g], Dmitri I. Svergun[d], Javier A. Linares-Pastén[a], Eva Nordberg Karlsson[a]

[a] Biotechnology, Department of Chemistry, Lund University, PO Box 124, SE-221 00 Lund, Sweden
[b] Department of Food Engineering, Faculty of Engineering, Ankara University, 06830 Gölbaşı, Ankara, Turkey
[c] Department of Food Engineering, Faculty of Engineering, Bülent Ecevit University, 67900 Çaycuma, Zonguldak, Turkey
[d] European Molecular Biology Laboratory Hamburg Outstation, EMBL c/o DESY Notkestrasse 85, 22607 Hamburg, Germany
[e] A.V. Shubnikov Institute of Crystallography of Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Leninsky prospect 59, 119333 Moscow, Russia
[f] A.N. Frumkin Institute of Physical chemistry and Electrochemistry of Russian Academy of Sciences, Leninsky prospect 31, 119071 Moscow, Russia
[g] Department of Biology & Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom

ARTICLE INFO

Keywords:
Homology modelling
Manual docking
Molecular dynamics
SAXS
Rhodothermus marinus
(A)XOS

ABSTRACT

Hydrolysis of arabinoxylan (AX) by glycoside hydrolase family 10 (GH10) xylanases produces xylo- and arabinoxylo-oligosaccharides ((A)XOS) which have shown prebiotic effects. The thermostable GH10 xylanase *Rm*Xyn10A has shown great potential to produce (A)XOS. In this study, the structure of *Rm*Xyn10A was investigated, the catalytic module by homology modelling and site-directed mutagenesis and the arrangement of its five domains by small-angle X-ray scattering (SAXS). Substrate specificity was explored *in silico* by manual docking and molecular dynamic simulations. It has been shown in the literature that the glycone subsites of GH10 xylanases are well conserved and our results suggest that *Rm*Xyn10A is no exception. The aglycone subsites are less investigated, and the modelled structure of *Rm*Xyn10A suggests that loop $\beta_6\alpha_6$ in the aglycone part of the active site contains a non-conserved α-helix, which blocks the otherwise conserved space of subsite +2. This structural feature has only been observed for one other GH10 xylanase. In *Rm*Xyn10A, docking revealed two alternative binding regions, one on either side of the α-helix. However, only one was able to accommodate arabinose-substitutions and the mutation study suggests that the same region is responsible for binding XOS. Several non-conserved structural features are most likely to be responsible for providing affinity for arabinose-substitutions in subsites +1 and +2. The SAXS rigid model of the modular arrangement of *Rm*Xyn10A displays the catalytic module close to the cell-anchoring domain while the carbohydrate binding modules are further away, likely explaining the observed lack of contribution of the CBMs to activity.

1. Introduction

Arabinoxylan (AX) is the second largest carbohydrate component after cellulose in cereal bran, a by-product from the production of flour. With an estimated annual world production of 2.0 million tonnes of cereal grains [1], cereal bran makes up a large source of AX with the potential to be utilised in value-adding applications. AX in cereal crops such as wheat and rye is present in the cell walls and there tightly associated with cellulose and lignin. The bran of these cereals contains around 50 to 250 mg AX per g dry matter, and the more easily-accessible water extractable AX makes up 2 to 10, and in the case of rye up to 14, mg per g dry matter [2–4]. AX consists of a backbone of 1,4 linked ß-D-xylopyranosyl units decorated with α-L-arabinofuranose units linked by an α-1,3 and/or α-1,2 bond with a degree of substitution (DS) between 0.4 and 0.8. Additional substitutions found in AX are ferulic and *p*-coumaric acid linked by an ester linkage to the O5-position of the arabinofuranose-unit as well as α-1,2 linked units of glucuronopyranosyl and its 4-*O*-methyl ether [3,5]. The two latter substitutions are

---

Abbreviations: X_1, D-xylose; X_2, 1,4-β-D-xylobiose; X_3, 1,4-β-D-xylotriose; X_4, 1,4-β-D-xylotetraose; X_5, 1,4-β-D-xylopentaose; X_6, 1,4-β-D-xylohexaose; A^3X, $1,3^2$-α-L-arabinofuranosyl-1,4-β-D-xylobiose; A^2XX, $1,2^3$-α-L-arabinofuranosyl-1,4-β-D-xylotriose; A^3XX, $1,3^3$-α-L-arabinofuranosyl-1,4-β-D-xylotriose; XA^3X, $1,3^2$-α-L-arabinofuranosyl-1,4-β-D-xylotriose; XA^2XX, $1,2^3$-α-L-arabinofuranosyl-1,4-β-D-xylotetraose; XA^3XX, $1,3^3$-α-L-arabinofuranosyl-1,4-β-D-xylotetraose; $A^{2+3}XX$, $1,2^3,1,3^3$-di-α-L-arabinofuranosyl-1,4-β-D-xylotetraose; $XA^{2+3}XX$, $1,2^3,1,3^3$-di-α-L-arabinofuranosyl-1,4-β-D-xylotetriose; $U^{4m2}XX$, $1,2^3$-α-D-(4-*O*-methyl-glucuronyl)-1,4-β-D-xylotriose

[*] Corresponding author.
E-mail address: anna.aronsson@biotek.lu.se (A. Aronsson).

https://doi.org/10.1016/j.bbapap.2017.11.006
Received 20 March 2017; Received in revised form 5 October 2017; Accepted 12 November 2017

common in the vegetable part of the plant but not in the AX present in the grain [6]. The hydrolysis products of AX from cereal grain are xylose, xylooligosaccharides (XOS) and arabinoxylooligosaccharides (AXOS), the latter two collectively termed AX-oligosaccharides ((A)XOS). XOS are abbreviated X_n where n is the degree of polymerisation (DP). For the nomenclature of the more complex AXOS, in 2009 Fauré and colleagues proposed a system for oligosaccharides derived from heteroxylans. The system is based on a one-letter code starting in the non-reducing end and where substitutions of the backbone chain xylose units are given unique letters associated with uppercase letters and numbers describing non-glycoside substitutions and position of substituents [7]. For example, an unsubstituted, a 1,3-arabinofuranosyl-substituted and a 1,2-(4-*O*-methyl-glucuronopyransyl)-substituted xylose unit are abbreviated X, A^3 and U^{4m2}, respectively.

(A)XOS have been shown to have prebiotic effects [8]. Reported health improvements resulting from ingestion of such prebiotics are reduced gut infections, better adsorption of minerals, and suppression of colon cancer [8]. Prebiotics are compounds that are neither degraded by gastric acid or host enzymes, nor adsorbed in the gastrointestinal tract, but fermented by intestinal bacteria that contribute to the health and well-being of the host, known as probiotic intestinal bacteria [9]. Stimulation of probiotic bacteria by (A)XOS has mainly been ascribed to strains of *Bifidobacterium* and a few strains of *Lactobacillus*, *e.g. brevis* [10–12]. Utilisation of (A)XOS varies between strains, even within the same genus, and depends on the length of the xylose chain and presence of arabinose-substitutions. AXOS can only be utilised by a limited number of species in the gut, mainly from *Bifidobacterium* but also a few from *Bacteroides* and *Lactobacillus*.

The hydrolysis of cereal bran AX into prebiotic (A)XOS can be achieved by endo-β-1,4-xylanases (often only called xylanases), which hydrolyse the internal glycosidic linkages between xylose units in the AX backbone. Endo-β-1,4-xylanases belong mainly to glycoside hydrolase (GH) family 10 and GH11, but can also be found in GH families 5, 8, 30, 43, 51 and 98 in the Carbohydrate Active Enzyme database (CAZy, www.cazy.org) [13]. Enzymes from the same family share a catalytic mechanism and overall fold. However, what compounds can be hydrolysed by an enzyme and products that can be formed from a given substrate, the substrate specificity, is not necessarily conserved among enzymes within one family [14]. Production of prebiotic (A)XOS has been performed by xylanases belonging to GH10, GH11 and GH30 [15]. In general, xylanases from GH11, which have a relative narrow active site, are less tolerant to arabinose-substitutions than xylanases from GH10 while most xylanases from GH30 have been shown to require a methyl glucuronic acid-substitution for hydrolysis to occur [16–18]. GH10 is thus a good choice for hydrolysis of cereal bran AX.

For the production of prebiotic (A)XOS from AX, understanding of how the active site residues of GH10 xylanases bind AX and (A)XOS is crucial. Improvement of purity and yield of the production by rational engineering is possible when the structure of the enzyme and its substrate binding interactions are known. In the active site of glycoside hydrolases, each sugar unit is bound into a subsite. According to the nomenclature of the field [19], the two subsites on either side of the cleavage point are named −1 and +1 respectively, and the number increases by one integer for each subsite away from the cleavage point. The subsites towards the reducing end of the substrate are the plus- or aglycone subsites and the subsites towards the non-reducing are consequently the minus- or glycone subsites. GH10 xylanases produce small oligosaccharides which can be explained by the strong binding of the well-conserved subsites −1, −2 and +1 [20]. More than two glycone subsites exist in the family but are not conserved [21]. The more distant aglycone subsites are not conserved, neither the interacting amino acids, nor the substrate specificity. Up to four aglycone subsites have been reported [22,23] and mainly hydrophobic interactions are observed [24]. The ability to bind arabinose-substituted substrates also varies within the family. Arabinose-substitutions in subsite −2 are permitted while substitutions in the aglycone subsites are reported but not conserved [20]. Considering the low conservation of the aglycone subsites and substrate specificity in the family, the xylanase for a given application must be carefully chosen based on structural and biochemical information.

To date, only three studies have presented crystallographic structures of GH10 xylanases in complex with AXOS [25–27]. The arabinose substitution is located at subsite −2 in all structures. This position is highly interesting since arabinose-substitutions at this position can have a role in substrate recognition as discovered by Xie and co-workers. They found four GH10 xylanases that had higher activity on A^3XX than X_3 [28]. However, structures of complexes with arabinose-substitutions in other positions are necessary for the prediction of hydrolysis products and as a base for tailor-made prebiotic (A)XOS.

A modular GH10 xylanase with valuable properties, *Rm*Xyn10A, has been cloned from the thermophilic bacterium *Rhodothermus marinus*, which was isolated from a hot spring in Iceland [29,30]. *Rm*Xyn10A comprises one signal peptide, two tandem carbohydrate binding modules (CBMs) from family 4, one domain with unknown function, one catalytic module and one putative cell anchoring domain [31]. Characterisation of the catalytic module revealed thermostability, activity remained for 24 h in 70 °C [32], which makes it suitable for industrial applications. The non-catalytic modules have no impact on the specific activity. However, the full-length enzyme has a higher stability than the catalytic module alone, indicating thermostabilising interactions between the domains [33]. *Rm*Xyn10A has been used for production of (A)XOS from various xylan-containing sources, *e.g.* rye bran, wheat bran, birchwood and quinoa stalks [11,12,34,35]. The produced oligosaccharides are mainly X_2, X_3 and A^3X, but also unidentified AXOS have been detected. Efforts have been made to determine the three-dimensional structure of the catalytic module in order to study substrate binding but several attempts have failed, showing that the module is difficult to crystallise.

In this paper we use a different approach by applying computational methods to provide deeper insight into substrate specificity and binding. The structure of the catalytic module of *Rm*Xyn10A is described through a homology model. Subsite affinities and substrate preferences are explored by molecular dynamics and docking techniques. Interesting structural features in the active site are further explored by site-directed mutagenesis. Valuable insights in the modularity of the full-length enzyme are gathered by combining structural data from small-angle X-ray scattering (SAXS) of the full-length enzyme with NMR data of the structure of CBM4-2 [36] and the homology models of the catalytic module and CBM4-1. The results are evaluated and compared to existing biochemical data of *Rm*Xyn10A as well as to the entire GH10.

2. Material and methods

2.1. Homology modelling

Homology modelling of the catalytic module of *Rm*Xyn10A (*Rm*Xyn10A_CM), residues 549–870 according to the full-length numbering, was carried out using the YASARA program [37,38]. Parameters used for the homology modelling are presented in Table S1.

Three crystal structures of XynB from *Xanthomonas axonopodis* pv. citri (Xac), PDB ID: 4PMZ-B (no ligand), 4PN2-A (X_2) and 4PMX-A (X_3), were identified as templates after running the PSI-BLAST algorithm [39] to extract a position specific scoring matrix, PSSM, from UniRef90 [40] and then searching for a match in the Protein Data Bank (PDB) [41]. The matches were ranked based on a total score defined as the product of the BLAST alignment score, the WHAT_CHECK [42] quality score obtained from the PDBFinder2 database [43] and the target coverage, see Table 1. The BLAST E-values are identical as the sequences are the same.

The following procedure was applied to generate alignments between the target and the identified templates. A secondary structure

Table 1
Templates for homology modelling of the catalytic module of *Rm*Xyn10A scored based on BLAST alignment score, the WHAT_CHECK quality score and target coverage.

PDB ID	Resolution (Å)	BLAST E-value	Align score	Coverage (%)	Total score
4PMZ-B	1.40	$6e^{-91}$	857	93	424
4PN2-A	1.42	$6e^{-91}$	855	93	406
4PMX-A	1.30	$6e^{-91}$	856	93	392

prediction of the target sequence was generated by the PSI-Pred secondary structure prediction algorithm [44,45] and a target sequence profile was generated from multiple sequence alignments based on related UniRef90 sequences. Sequence profiles of the templates were generated by a combination of sequence alignments from UniRef90 sequences and several structural alignments [46–48] of 21 structurally similar proteins from PDB. Alignments of the target sequence and the templates were generated by aligning the target and the template sequence profiles based on SSALN scoring matrices [49] including structural information of the template and the predicted target secondary structure.

For each alignment, five models were created as follows. First, the backbone was built based on a stochastic alignment [50], gaps and insertions were modelled based on anchor points and loop conformations from PDB [51]. A side-chain rotamer network was constructed [52] and ligands present in the templates were parameterised and considered in the procedure. Next, to optimise the loops, 50 iterations of applying different loop conformations and subsequent side-chain optimisation in implicit solvent [53] were made. Side-chains were fine-tuned by steepest descent energy minimisation (EM) in dihedral angle space with implicit water and YASARA2 force field [38] and the hydrogen bonding network of the model was optimised [54]. Last, EM with explicit solvent molecules was run in two steps, first a combined steepest descent and simulated annealing minimisation on the modelled loops and second, a full unrestrained simulated annealing minimisation on the entire model.

The 15 models were ranked based on an overall Z-score, see Table S2, which defines how many standard deviations the model quality is from an average high-resolution X-ray structure, where a negative value indicates a lower quality. The overall Z-score is a weighted average of individual Z-scores of backbone- and side-chain dihedrals and packing interactions: $\text{Overall Z-score} = 0.145 \cdot \text{Dihedrals} + 0.390 \cdot \text{Packing1D} + 0.465 \cdot \text{Packing3D}$.

A hybrid model containing three xylose-like ligands and a calcium ion (Ca^{2+}) was generated by combining the best parts of the 15 models, see Table S3, and evaluated by Z-scores.

2.2. Refinement and validation

Two structures were created from the hybrid model, one where the ligands present were removed and one where both the ligands and Ca^{2+} were removed. Refinement was performed by the software YASARA. A 500 ps simulation [55] of molecular dynamics (MD) was run at a temperature of 70 °C, a water density of 0.978 g/mL, a pH of 7.5 and a sodium chloride concentration of 0.9% (w/v). The YASARA2 force field, periodic boundary with a cell extension of 7.5 Å on each side of the protein, long-range coulomb forces calculated by the particle-mesh Ewald (PME) method [56], temperature control by a Berendsen thermostat based on time-averaged temperature and intramolecular and intermolecular time steps of 2 and 1 fs respectively, were used. Before each simulation, neutralisation, EM and setting of temperature were carried out. Neutralisation was done by calculating pK_a [57] and protonation states, adding TIP3P water molecules, running steepest descent and simulated annealing of the solvent and adding sodium and chloride ions to the final concentration. EM was done by the steepest descent method followed by simulated annealing. The temperature was reached by setting initial atom velocities according to a Boltzmann distribution. During the simulation, 20 trajectories were saved, energy minimised and analysed by checking the energy of the system as well as dihedral angles, packaging1D and packaging3D. The best trajectory was further evaluated as described below. Both structures were refined until no overall improvement was seen during evaluation of the refined structures.

Evaluation of the refined structures was done by average 3D-1D scores generated by Verify 3D [58,59], overall C^{α} root mean square deviation (RMSD) between the models and the native structure of the template protein (PDB ID: 4PMX) [60] generated by Chimera [61], Ramachandran plots generated by PROCHECK [62], Z-scores generated by ProSA-web [63,64], WHAT_CHECK [42] analysis, overall quality factors generated by ERRAT [65] and Z-scores and non-local high-energy profiles generated by ANOLEA [66–68].

The refined hybrid model without Ca^{2+} was determined to be the final model. A second variant of the final model was created by adding a His-tag, consisting of six Histidine residues at the C-terminus, and a Methionine at the N-terminus. The residues were added one by one in YASARA and EM was applied to the entire protein after the addition of each residue.

2.3. Molecular dynamics

The final model, with and without added residues, was subjected to MD simulation in GROMACS 2016 [69,70] at 70 °C, pH 7.5 and 1 bar for 500 ns. The AMBER03 force field [71], 2.5 fs time steps, periodic boundaries, 8 Å cutoff of short-range electrostatic and van der Waals forces and long-range forces calculated by PME was applied. Before each simulation, solvation, EM and equilibration were carried out. A cubic cell with a minimum of 10 Å extension from the protein filled with TIP3P water molecules and with a sodium chloride concentration of 0.9% (w/v) was created around the protein. EM was achieved by a steepest descendent algorithm including a step size of 0.1 Å, a tolerance of 1000 kJ/mol and with maximum 50,000 steps. Equilibration was done in two steps. First, the temperature was stabilised under a 100 ps long NVT ensamble with time steps of 2 fs and with temperature coupling by a modified Berendsen thermostat using velocity rescaling with a stochastic term [72]. Second, the pressure was stabilised under a 100 ps long NPT ensamble with times steps of 2 fs and with a pressure coupling by the Parrinello-Rahman method [73]. Potential energy during EM, temperature during NVT equilibration as well as pressure during NPT equilibration were monitored for the two models and can be seen in Fig. S1. During the MD simulations, trajectories were saved every 1.25 ns. RMSD C^{α} for the entire protein and root mean square fluctuation (RMSF) for each residue were calculated for each simulation.

2.4. Docking arabinoxylan-oligosaccharides

A general working scheme for the docking procedure is illustrated in Fig. 1. The template protein in complex with X_3 docked in subsites −3 to −1 (PDB ID: 4PN2) was superimposed onto the final model. Manual docking, using Chimera, was applied to the final model with the ligand by changing rotamers of the interacting amino acid residues manually with guidance from the template and other glycoside hydrolase family 10 (GH10) xylanases. Finally, only the ligand and the final model were kept and subjected to EM followed by a 10 ns MD in YASARA at 70 °C, pH 7.5, with a water density of 0.978 g/mL and in a sodium chloride concentration of 0.9% (w/v). The AMBER03 force field, periodic boundaries with a cubic cell with a minimum of 10 Å extension from the protein, long-range coulomb forces calculated by PME, temperature control by a Berendsen thermostat based on time-averaged temperature and intramolecular and intermolecular time steps of 2 and 2.5 fs, respectively, were used. Neutralisation, EM and setting of temperature were done before each simulation as described for the refinement. The

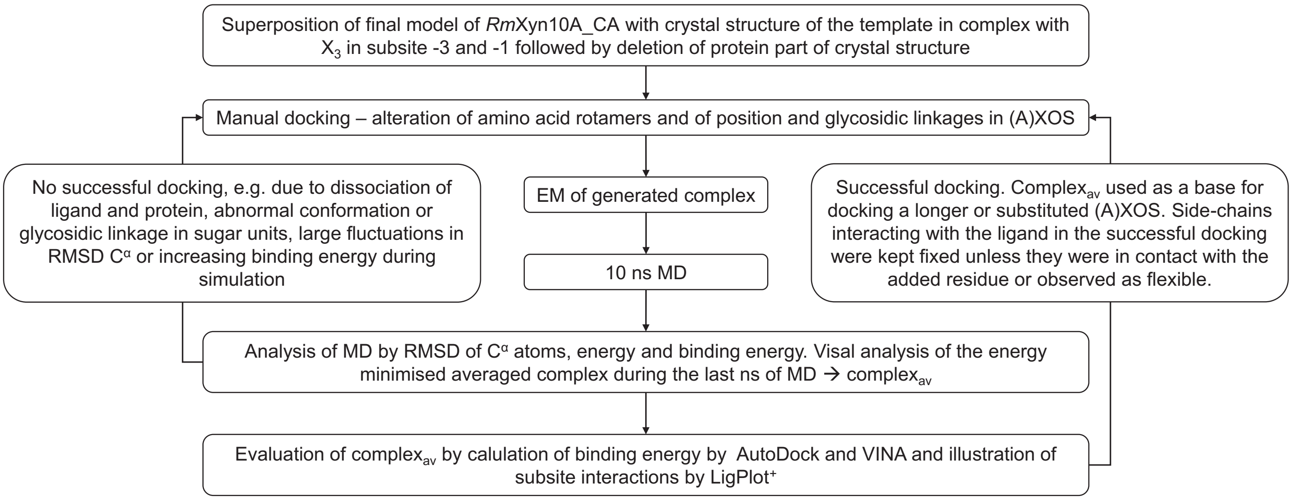


Fig. 1. Working procedure for docking (A)XOS into active site of the model of *Rm*Xyn10A_CM.

catalytic residues (Glu667 and Glu780) were kept fixed during EM and MD. During the simulation, trajectories were saved every 25 ps. The total energy of the system, RMSD C^{α} for the entire protein and binding energy between the protein and the ligand were calculated for each simulation. Equilibrium was assumed when the fluctuation of RMSD C^{α} for the entire protein was lower than 1 Å for the remaining simulation. The averaged structure of the last ns was energy minimised and further analysed, given that the equilibrium was reached after 5 ns, and used as the final complex.

Binding energies of the final complex were calculated after local docking by use of two methods: VINA [74] using default parameters and AutoDock [75] using the default parameters, and point charges initially assigned according to the AMBER03 force field, and damped to mimic the less polar Gasteiger charges used to optimise the AutoDock scoring function. Both docking methods were performed from the YASARA interface [76] and with the ligand kept rigid. During local docking with AutoDock, 100 cycles were run, and the binding energy for the best hit was saved. For analysis of subsite interactions, 2-D illustrations of ligand-protein interactions were prepared by Ligplot+ [77]. For visualisation and production of pictures of the 3-D structure, Chimera was used.

Xylose units were added to the reducing end of the ligand in the resulting complex with guidance from structures containing complexes of GH10 xylanases with ligands docked in the aglycone region and the same procedure as described above was applied. During dockings in the aglycone region, residues in the glycone were kept fixed to minimise unrealistic simulation results. Arabinofuranosyl-substitutions were later added manually to the existing ligands in their desired position using torsion angles of the glycosidic linkage as reported by Yui with coworkers for arabinoxylan isolated from rice [78]. If steric hindrance occurred, the torsion angles were changed to the closest position without any clashes with the protein before EM was applied.

2.5. Site-directed mutagenesis of RmXyn10A_CM

The following seven mutants of *Rm*Xyn10A_CM were designed in order to investigate features in the aglycone part of the active site: G720N, S724G, D782G, D784G, F754N, F754Del and R833G. The mutations were generated through synthetizing whole plasmids by PCR and using mutagenic primers, see Table 2. The template was the plasmid pET21b::*Rm*Xyn10A harboring the wild type (*wt*) gene. High-fidelity DNA polymerase iProof™ kit (Bio-Rad, California, USA) was used for the 29 cycles of amplification. Thereafter, the PCR reaction mixes were treated with *Dpn*I endonuclease, at 37 °C per 2 h, in order to destroy the *wt* template.

Chemically competent *Escherichia coli* NovaBlue strain (Novagen brand, Merck KGaA, Darmstadt, Germany) was transformed directly with the PCR products (without any purification) for the proliferation of the mutant plasmids. The cells were grown in Luria Bertani (LB) medium supplemented with 100 μg/mL ampicillin, at 37 °C for 12 h. Then, the plasmids were extracted using QIAprep Spin Miniprep Kit (QIAGEN, Germany) and sequenced in order to verify the mutations.

2.6. Production and purification of recombinant proteins: wt and mutant proteins

Wild type and mutant proteins were produced in *E. coli* BL21 (DE3) strain (Novagen brand, Merck KGaA, Darmstadt, Germany). The corresponding plasmids were introduced into the competent cells by thermal shock. The cells harboring plasmids were cultivated in 1000 mL LB supplemented with 100 μg/mL ampicillin at 37 °C. The expression of the recombinant proteins was induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) when the optical density (λ = 600 nm) reached 0.5–1, and the production phase was 15 h. Thereafter, the cell pellets were harvested by centrifugation and resuspended in binding buffer (50 mM TRIS, 0.5 M NaCl, pH 7.4). The resuspended cells were lysed by ultrasonication and the cell extracts were centrifuged at 25931 × *g*, 20 min. The supernatants were used for the protein purification by immobilized metal ion affinity chromatography (IMAC) using an ÄKTA Prime system with a HisTrap™ FF crude column (GE Healthcare) as described elsewhere [79]. Purity of the proteins was determined by SDS-PAGE and their respective concentration was estimated by NanoDrop® (Spectrophotometer 1000, Saveen Werner, Sweden) at 280 nm, using the approximation: 1 A_{280nm} = 1 mg/mL.

2.7. Enzyme activity measurements

The activity of the produced *wt* and mutant proteins of *Rm*Xyn10A_CM was determined by the dinitrosalicyclic acid (DNS) stopping method [80]. 20 μL of a defined enzyme concentration, prepared by dilution in 20 mM sodium phosphate buffer (SPB), pH 7.4, was added to 40 μL of 1% birchwood xylan (Sigma, St Louis, MO, USA) in 20 mM SPB, pH 7.4. The reaction mixture was incubated in 70 °C for 10 min before the reaction was stopped by adding 90 μL of DNS reagent to the reaction mixture and incubated at 98 °C for 10 min. Measurements of absorbance was done at 540 nm. The concentration of reducing ends was calculated from a standard curve using xylose of defined concentration, dissolved in the same substrate.

Table 2
Primers used for constructing site-directed mutagenesis, from 5′ to 3′ end. The horizontal and vertical lines indicate the mutation.

Mutant	Forward primer	Reverse primer
G720N	caatgactacaacatcctcagcagtctgg	ctgctgaggatgttgtagtcattgatcatc
S724G	catcctcagcggcctggaaacggcccagc	ccgtttccaggccgctgaggatgccgtag
D782G	gttaccgagatgggcatcgacggcaatcccaatc	gattgccgtcgatgcccatctcggtaacctgaatc
D784G	gatggatatcggcggcaatcccaatcagagc	gattgggattgccgccgatatccatctcggtaac
F754N	gggcatgctaactcgacgcgttccggggc	gaacgcgtcgagttagcatgcccctgcacc
F754Del[a]	caggggcatgct │ tcgacgcgttccggggcg	ggaacgcgtcga │ agcatgcccctgcacccc
R833G	cggcctgtggggcaatgattacgaagcctac	cgtaatcattgccccacaggccgggccgc

[a] Deleted residue.

2.8. Hydrolysis of arabinoxylan extracted from rye bran

The product profiles of the produced *wt* and mutants of *Rm*Xyn10A_CM from hydrolysis of arabinoxylan (AX) extracted from rye bran were investigated. AX from an upscaled extraction with rye bran as starting material [34] was used as substrate with a concentration of 1% in 20 mM SPB, pH 7.4. 60 nmol of respective protein/g AX (62% (w/w) in the substrate) were added to the substrate and the reaction mixtures were incubated at 70 °C. Samples were taken after 1 and 20 h, the reaction was stopped by diluting the samples in NaOH to pH 10.

The oligosaccharides produced during the reactions was analysed by High Performance Anionic Exchange Chromatography coupled with Pulsed Amperometric Detection (HPAEC-PAD) as described elsewhere [34]. Xylose and the oligosaccharides X_2, X_3, X_4, X_5, X_6, A^2XX, A^3X and XA^3XX purchased from Megazyme (Wicklow, Ireland) were used as standards for quantification. A mixture of XA^2XX and XA^3XX (Megazyme) was used as the standard to identify XA^2XX.

2.9. Small angle X-ray scattering (SAXS)

The structure of the full-length version of *Rm*Xyn10, containing two tandem carbohydrate binding modules from family 4 (CBM4-1 and CBM4-2), a third domain with unknown function (D3), the catalytic module (CM) and a fifth domain with putative cell anchoring function (D5), was analysed in solution by SAXS. *Rm*Xyn10A was produced as described elsewhere [33].

The synchrotron radiation X-ray scattering data were collected on the X33 camera [81,82] at the European Molecular Biology Laboratory (EMBL) on the storage ring DORIS III of the Deutsches Elektronen Synchrotron (DESY) using multi-wire proportional chambers with delay line readout [83]. The scattering patterns were recorded at a sample-detector distance of 2.3 m covering the range of momentum transfer $0.15 < s < 3.5\ \mathrm{nm}^{-1}$, defined as $s = \frac{4\pi \sin\theta}{\lambda}$, where 2θ is the scattering angle and λ (= 0.15 nm) is the X-ray wavelength. The solutions were measured at protein concentrations of 2.5, 6.8 and 13.7 mg/mL.

To check for radiation damage during the scattering experiments, the data were collected in 20 successive one-minute frames. The individual frames were averaged after normalisation to the intensity of the incident beam, corrected for the detector response and the scattering of the buffer was subtracted using the program PRIMUS [84]. The difference curves were scaled for the solute concentrations and extrapolated to infinite dilution following standard procedures [85].

The maximum dimensions of the particles (D_{max}) were estimated using the indirect Fourier transform package GNOM [86,87], which also provides the distance distribution function (p(r)) of the particle. The forward scattering (I(0)) and the radii of gyration (R_g) were evaluated using the Guinier approximation [88] assuming that at very small angles, $s < 1.3/R_g$, the intensity can be represented as $I(s) = I(0)\exp\left(-\frac{(sR_g)^2}{3}\right)$. These parameters were also computed from the entire scattering pattern using GNOM. The molecular mass (MM) of the solute was evaluated by comparison of the forward scattering with that from a reference solution of bovine serum albumin with MM = 66 kDa.

Low-resolution models of the protein were generated *ab initio* by the program GASBOR [89]. The program represents the protein by an assembly of dummy residues (DRs) and uses simulated annealing to build inside a sphere with the diameter D_{max} a locally "chain-compatible" DR-model that fits the experimental data $I_{exp}(s)$ to minimise the discrepancy: $\chi^2 = \frac{1}{N-1}\sum_j \left[\frac{I_{exp}(s_j) - cI_{calc}(s_j)}{\sigma(s_j)}\right]^2$, where N is the number of experimental points, c is a scaling factor and $I_{calc}(s)$ and $\sigma(s_j)$ are the calculated intensity and the experimental error at the momentum transfer s_j, respectively.

Results from at least ten separate GASBOR runs were averaged to determine common structural features using the programs DAMAVER [90] and SUPCOMB [91]. The latter program aligns two arbitrary low or high resolution models represented by ensembles of points by minimising a dissimilarity measure called normalised spatial discrepancy (NSD). For every point (bead or atom) in the first model, the minimum value among the distances between this point and all points in the second model are found, and the same is done for the points in the second model. These distances are added and normalised against the average distances between the neighbouring points for the two models. Generally, NSD values close to unity indicate that the two models are similar. The program DAMAVER generates the average model of the set of superimposed structures and also specifies the most typical model *i.e.* that having the lowest average NSD with all the other models of the set.

Rigid body modelling of the tertiary structure of full-length *Rm*Xyn10A in solution was performed based on the available high-resolution atomic coordinates of the two carbohydrate binding modules CBM4-1 (homology model with CMB4-2 as template (PDB ID: 1K45)) and CBM4-2 (PDB ID: 1K45), and CM (homology model presented in this study). The fragments with the unknown structures, D3, D5 and linkers, were represented as interconnected chains composed of DRs [92]. The positions of CBM4-1 and CBM4-2 were fixed based on previous modelling.

A simulated annealing protocol was employed to find the optimal positions and orientations of available high-resolution models of domains and the probable conformations of the DR chains attached to the appropriate residues. The theoretical scattering pattern I(s) was calculated from the available high-resolution coordinates of the domains with known structure and from the portion with unknown structure represented as DRs using spherical harmonics: $I(s) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-l}^{l} \left|\sum_n A_{lm}^n(s) + \sum_k B_{lm}^k(s)\right|^2$. The complex functions $A_{lm}^n(s)$ are the partial scattering amplitudes of the domains in the given positions and orientations which depend on the scattering amplitudes in the reference positions calculated using the program CRYSOL [93] and on three rotational and three translational parameters. $B_{lm}^k(s)$ are the partial amplitudes of DRs comprising the unknown parts, calculated using the form factor of a DR [89,92]. The algorithm is a particular case of the program BUNCH, which allows determination of three-dimensional domain structure of proteins based on multiple scattering data sets from deletion mutants when the structure(s) of individual domains are available [94].

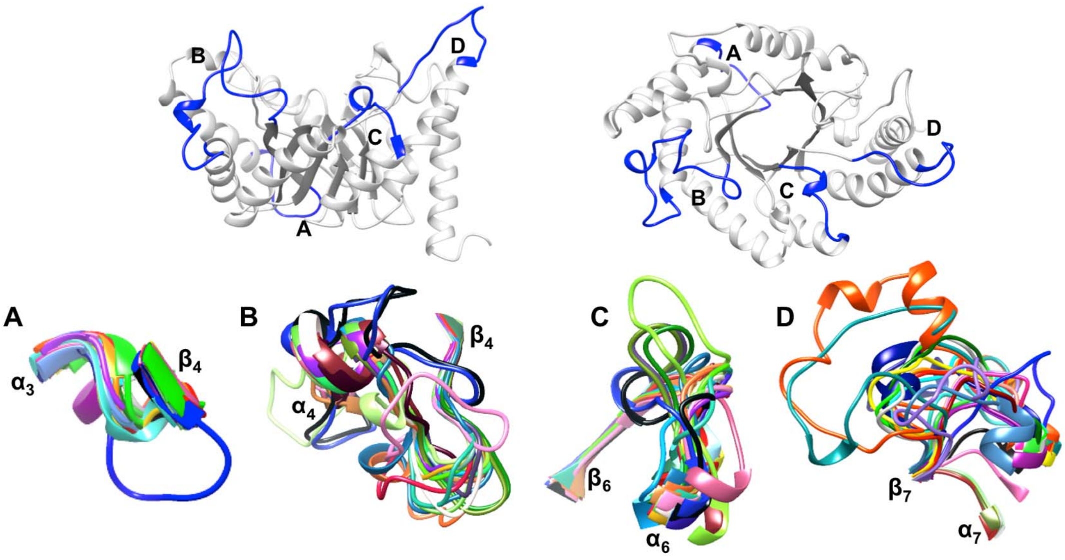


Fig. 2. Top: Model of *Rm*Xyn10A_CM in two perpendicular orientations, highlighting non-conserved loops in blue. Below: superposition of A: loop $\alpha_3\beta_4$, B: loop $\beta_4\alpha_4$, C: loop $\beta_6\alpha_6$ and D: loop $\beta_7\alpha_7$ with template protein, 4PMX, (black) and 27 crystal structures of GH10 proteins named in Table S4. Loop A, C and D are in the same orientation as the model to the left and loop B as the model to the right.

3. Results and discussion

3.1. The catalytic module of RmXyn10A is folded as a typical TIM-barrel with several non-conserved loop-regions

A homology model of the catalytic module of xylanase *Rm*Xyn10A from *Rhodothermus marinus* (*Rm*Xyn10A_CM) was successfully generated from a template with 51% identity. The model contains a $(\beta/\alpha)_8$ TIM-barrel fold, the overall fold conserved among glycoside hydrolase (GH) family 10. By comparing the model to GH10 xylanases of known structure, major differences in four loops were observed, see Fig. 2. Two of these loops formations, $\beta_4\alpha_4$ and $\beta_6\alpha_6$ illustrated in Fig. 2-B, C, show similarity only to the template protein in the homology modelling, XynB from *Xanthomonas axonopodis* pv. citri (Xac) (PDB ID: 4PN2) [95]. These loops are both located in the presumed aglycone region of the active site. Only one residue in the α-helix of loop $\beta_6\alpha_6$ differs between *Rm*Xyn10A and XynB from *X. axonopodis* whereas the other GH10 xylanases show less similarity to *Rm*Xyn10A in this region. The residues in loop $\beta_4\alpha_4$ show low similarity, both to XynB and other GH10 xylanases. Thus, it is likely that the prediction of loop $\beta_6\alpha_6$ is more correct than the prediction of $\beta_4\alpha_4$. The other two loops contain large insertions compared to XynB, Thr656-Phe660 in loop $\alpha_3\beta_4$ on the bottom of the barrel and Pro787-Gln797 in loop $\beta_7\alpha_7$ on the opposite side. As can be seen in Fig. 2-A, D, these loop conformations show no similarity to any other GH10 xylanase structure. Loop $\alpha_3\beta_4$ seems to be well conserved among GH10 proteins (apart from *Rm*Xyn10A), whereas loop $\beta_7\alpha_7$ varies considerably in both length and conformation.

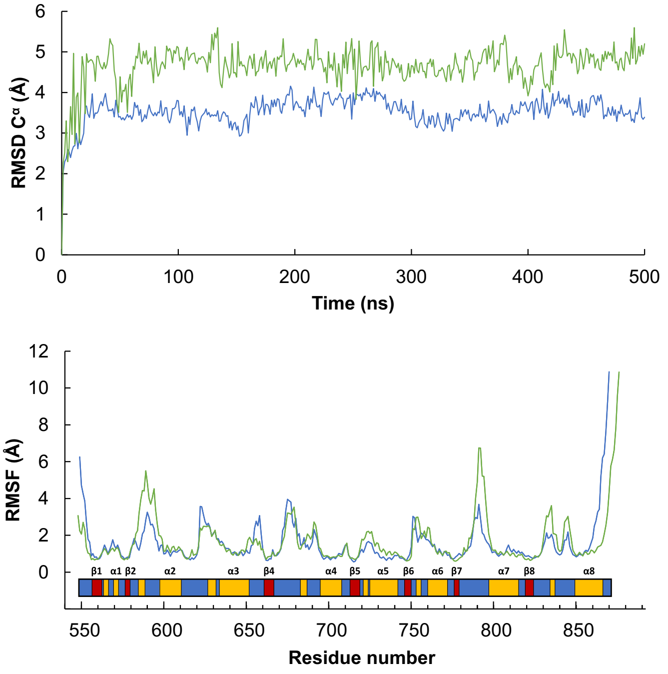


Fig. 3. MD of the model of *Rm*Xyn10A_CM without (blue) and with (green) added residues for expression and purification. On top, RMSD of C$^\alpha$ atoms during the simulations. On bottom, RMSF for each residue after an equilibrium time of 100 ns. Secondary structure assignments of the final model are shown in the bottom of the graph, red and yellow bars correspond to β-strands and α-helices, respectively.

The four non-conserved loops were all shown to be flexible regions of the protein. Stability of the model was investigated by molecular dynamics (MD) in conditions where the protein has been shown to function well [32]. Root mean square deviation (RMSD) of C$^\alpha$ atoms during the simulation are presented in Fig. 3 in blue, showing a stabilisation in RMSD at around 3.6 Å after 50 ns with fluctuations within 1.3 Å. Root mean square fluctuation (RMSF) for each residue are also presented in blue in Fig. 3 clearly showing that the largest fluctuations are within the loop regions of the protein rather than in the secondary structures. The largest fluctuations can be seen in the C- and N-termini and the four non-conserved loops $\alpha_3\beta_4$, $\beta_4\alpha_4$, $\beta_6\alpha_6$ and $\beta_7\alpha_7$ (Fig. 2) are among the regions with the highest fluctuation possibly indicating a weakness in the prediction of these regions. The non-conserved α-helix on loop $\beta_6\alpha_6$ remained throughout the simulation and was even elongated towards α-helix 6.

During previous biochemical characterisation and when used for production of prebiotic arabinoxylan-oligosaccharides ((A)XOS), a His-tag at the C-terminus and a Methionine at the N-terminus have been appended to *Rm*Xyn10A_CM to simplify expression and purification by metal-ion-affinity chromatography. Previous studies have shown that the enzyme is active and the His-tag is expected to have no significant impact. According to the model, the His-tag is situated on the bottom of the barrel and, thus, does not interfere with the active site situated on top of the barrel. This conclusion was supported by running MD of the model with a His-tag and Met549 added. However, some instability is

Table 3
Validation of the hybrid model and two refined structures of the hybrid model.

Structure	Verify 3D	Chimera	PROCHECK	ProSA	ERRAT	ANOLEA
	Average 3D-1D score ≥ 0.2[a] (%)	Overall RMSD C^{α} to 4PMX[b] (Å)	Ramachandran plot[c]	Z-score	Overall quality factor[d] (%)	Z-score
MT	95.75		88.4;11.6;0.0;0.0	−8.86	98.649	−0.31
HM	88.82	0.908	88.6;10.7;0.4;0.4	−9.44	94.904	0.26
rHM	90.37	1.044	87.5;12.1;0.4;0.0	−9.83	93.949	0.11
rHMnoCa	94.41	1.023	88.6;11.1;0.4;0.0	−9.59	97.452	−0.11

[a] 80% of the residues in a structure is required at an averaged 3D-1D score of 0.2 to be considered a good model.

[b] Calculated from 296 residues aligned after superimposing the structures with a match alignment cut off of 5.0 Å. A rule of thumb from the Critical Assessment of protein Structure Prediction (CASP) meetings is that an overall RMSD of C^{α} should be < 1 Å if the sequence identity between target and template is over 60% [60]. In this case, the sequence identity is 51% and a slightly higher overall C^{α} RMSD can be expected.

[c] Numbers corresponds to percentage of residues in: Most favourable regions; additionally allowed regions; generously allowed regions; disallowed regions.

[d] For high-resolution X-ray structures an overall quality factor of 95% is considered a good protein structure.

introduced to the protein as seen in Fig. 3 where the green line shows the His-tagged protein. Longer time until equilibration, larger fluctuations in RMSD of C^{α} atoms (around 1.8 Å) and higher fluctuation in some loop regions, especially loops $\beta_2\alpha_2$ and $\beta_7\alpha_7$, are observed for the His-tagged protein compared to the native. The overall higher RMSD can to some extent be explained by the added His-tag, forming a flexible tail in the C-terminus.

3.2. Refinement of the homology model exposes an incorrect calcium ion

A calcium ion (Ca^{2+}) interacting with residues Glu589, Asp593, Glu640 and Glu643 was added during the homology modelling, but several validation methods and biochemical data suggest that no metal ion-binding site is present in the structure. Refinement of the hybrid model and a second structure without Ca^{2+} were performed. Table 3 summarises the results of the validation of the main template of the homology modelling (MT), the hybrid model (HM), the hybrid model without ligands after refinement (rHM) and the hybrid model without ligands and Ca^{2+} after refinement (rHMnoCa). Detailed results from Verify-3D, ProSa-web, and ANOLEA are presented in Figs. S2, S3 and S4, respectively. All validation methods suggested that a higher quality model was achieved by removing Ca^{2+} following refinement compared to refining the homology model including Ca^{2+}.

Several quality assessments highlight the Ca^{2+} binding site as a problematic region. A large difference in the presence of abnormally short interatomic distances was detected by WHAT_CHECK. HM and rHM had nine and eleven respectively, whereas only two were found in rHMnoCa and none in MT. The majority of the abnormally short distances in the models with Ca^{2+} were between oxygen and carbon atoms of the carboxyl groups of the residues binding Ca^{2+}. In rHMnoCa, the side-chains previously involved in Ca^{2+} binding had moved, separated in space and thus lessened the error of abnormally short distances. Furthermore, the short distances in the Ca^{2+} binding site of the model fails these models by ERRAT, as seen in Fig. 4. Additionally, non-local energy profiles of HM and the refined models generated by ANOLEA reveal high-value zones in the Ca^{2+} binding site of rHM.

Biochemical data provides more evidence against the existence of Ca^{2+} in *Rm*Xyn10A_CM as seen in HM. Previous studies by Abou-Hachem and co-workers on several constructs of *Rm*Xyn10A where modular stability were investigated by differential scanning calorimetry did not find any influence of Ca^{2+} on the catalytic module [33]. Evidence for the presence of Ca^{2+} in *Rm*Xyn10A is however available for its two carbohydrate binding modules (CBMs), which have been shown to bind Ca^{2+} and thus improve the stability of the full length enzyme [96]. A likely explanation for the presence of Ca^{2+} in HM is that Ca^{2+} is present in the template protein XynB in the same position as HM. However, in contrast to the studies on *Rm*Xyn10A, a comparison of XynB including Ca^{2+} and where the Ca^{2+} were removed by treatment with EDTA revealed that Ca^{2+} plays a major role in both the stability of the protein and in catalytic activity [95]. Ca^{2+} is not common in the catalytic modules of enzymes from GH10. Out of the 34 GH10 proteins of known structure present in the Carbohydrate Active Enzymes (CAZy) database, only four have Ca^{2+} present in the catalytic module, the template protein of the homology model being one of them. The Ca^{2+} binding sites are all on the outside of the TIM-barrel structure but their positions are widely spread and thus not conserved.

The structure rHMnoCA was used as the final model and all simulation and discussion is based on this structure – this choice was based on three observations. First, the quality of the refined model excluding Ca^{2+} was more highly rated by all validation methods completed compared to the refined model including Ca^{2+}. Secondly, the hybrid model including Ca^{2+} failed several quality assessments, which was directly caused by the binding of Ca^{2+}. Thirdly and most importantly, the biochemical studies do not support the presence of Ca^{2+}.

3.3. The glycone subsites are conserved and allows O3-bound arabinose in subsite −2

The active site is situated on top of the wider part of the barrel. Residues Glu667 and Glu780 were identified as the catalytic acid/base and nucleophile of a retaining glycoside hydrolase, as these residues are situated at the end of β-strand 4 and 7, respectively, with a distance of 6 Å between the carboxyl groups. The glycone part of the active site of GH10 xylanases is highly conserved with well-defined subsites −1 and −2 [20]. The residues building up the glycone subsites were identified by comparing the crystal structures of GH10 xylanases presented in Table S5, which contain various xylooligosaccharides (XOS) bound in active site. The conserved residues were confirmed to be involved in ligand interaction by manual docking and MD of X_3 in the presumed −3 to −1 subsites. Fig. 5 clearly shows defined subsites −2 and −1, whereas no interactions for the −3 subsite were observed. However, the more distal residues Lys584, Asn581 and Gln624 were occasionally observed in a rotamer not interacting with the xylan chain. It is difficult to know whether this is due to flexibility in these residues or a less realistic simulation result. Interesting to note is that the aromatic group of Trp832, on top of the glycone subsites in Fig. 5, is turned 180° compared to all other structure-determined GH10 with the exception of the template XynB from *X. axonopodis*. As a consequence, the amide group, which is capable of forming a hydrogen bond, is exposed to the surrounding solvent rather than the interior of the protein structure. If this observation is due to an inaccuracy in the determined structures of the template or is a correct prediction is impossible to say without a crystal structure. However, simulations were run with both rotamer alternatives and showed no significant impact on the binding of XOS or arabinoxylooligosaccharides (AXOS).

As no interaction was observed for more distal glycone subsites than −2, investigation of possibilities of arabinose-substitutions was limited to subsite −2. Steric hindrance prevents any arabinose bound to the

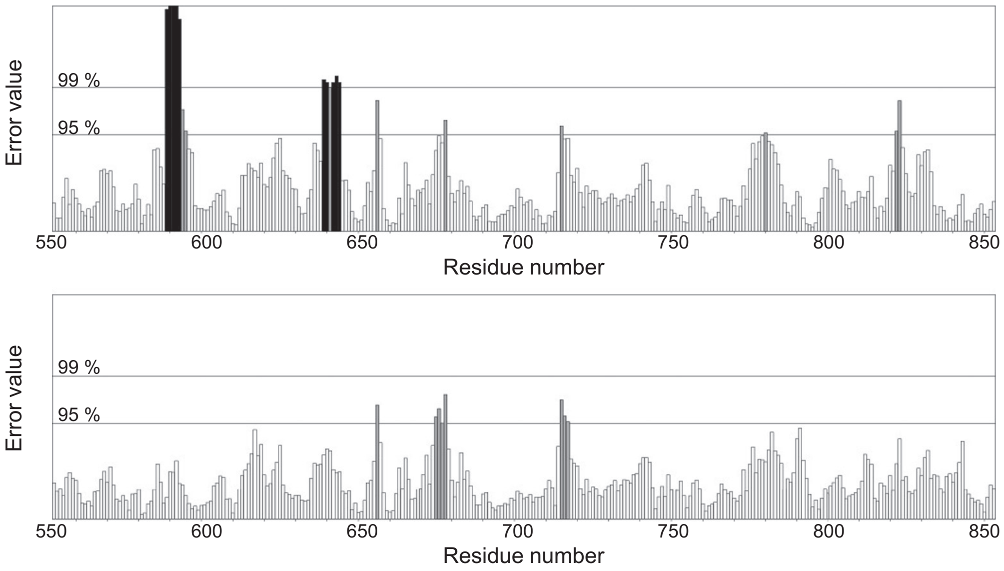


Fig. 4. Error-value generated by ERRAT for each residue in the hybrid model without ligands after refinement (upper) and the hybrid model without ligands or Ca^{2+} after refinement (lower). ERRAT is based on statistics from high-resolution crystal structures and calculates the error of the protein regarding non-bonded interactions between different kinds of atoms [65]. The two lines in each graph indicate the confidence to which it is possible to reject regions that exceed the error value of the line.

O2-position of subsite -2 as well as any substitutions in subsite -1. Hydrogen bonds from atom HO2 of the arabinose bound in O3-position in subsite -2 to each of the carboxyl oxygen of Glu580 were observed. These hydrogen bonds have been observed in all GH10 structures in complex with arabinose in this position (PDB ID: 2CNC, 2FGL, 1UR1, 1VGU and 2BNJ). The possibility of a hydrogen bond to the backbone of Leu831 on loop $\alpha_8\beta_8$ above active site was observed in some dockings. The same bond was also observed in Xyn10A from *Streptomyces olivaceoviridis* (PDB ID: 1VGU). Other hydrogen bonding residues observed in the above mentioned complexes are not present in *Rm*Xyn10A.

3.4. The aglycone subsites are non-conserved with two possible clefts

Interactions in the aglycone region of *Rm*Xyn10A were more difficult to predict as several alternative positions were observed rather than a single fit between protein and ligand. The aglycone subsites are generally not conserved among xylanases in the family with the exception of an aromatic residue stacking against the sugar in $+1$ [25] and an Asparagine making hydrogen bonds to the sugar in subsite $+3$. The aromatic residue is conserved in *Rm*Xyn10A as Tyr719, but the Asparagine is not, instead Gly720 is present in the position with no hydrogen bonding possibilities. Interestingly, in the model, a short α-helix containing Phe754, situated on loop $\beta_6\alpha_6$, see Fig. 2-C, is blocking subsite $+2$ of the active site groove normally found in GH10 xylanases. This structural feature has only been observed in the template protein, XynB from *X. axonopodis*. Consequently, the reducing end of the substrate binding into the active site needs to take another route than has been found in GH10 xylanases so far. The starting point of the docking in aglycone cleft was based on the structures of complexes of GH10 and XOS in the aglycone region. Finding a stable complex by MD was challenging and several attempts where the ligand position and side chains were initially altered were necessary to dock the ligand in possible aglycone subsites. Two main conformations of the xylan chain with corresponding aglycone clefts, exemplified in Fig. 6, were identified and further evaluated.

In both the alternatives, the xylose unit in subsite $+1$ is similarly positioned with no hydrogen bond interactions. Substitution with arabinose in O2, O3 and O2/O3 position in this subsite were successfully docked and several hydrogen bond interactions were observed. The non-conserved loop $\beta_4\alpha_4$ is highly involved in interacting with substituents in subsite -1, especially Glu671 but also the more distal part of the loop was observed to move during simulation to facilitate interactions with arabinose bound in O2-position, see Fig. 7.

The first alternative for the aglycone subsites is illustrated in Fig. 6-A, and is the closest to other GH10 complexes, but the cleft is slightly pushed down because of the α-helix in loop $\beta_6\alpha_6$. The atypical

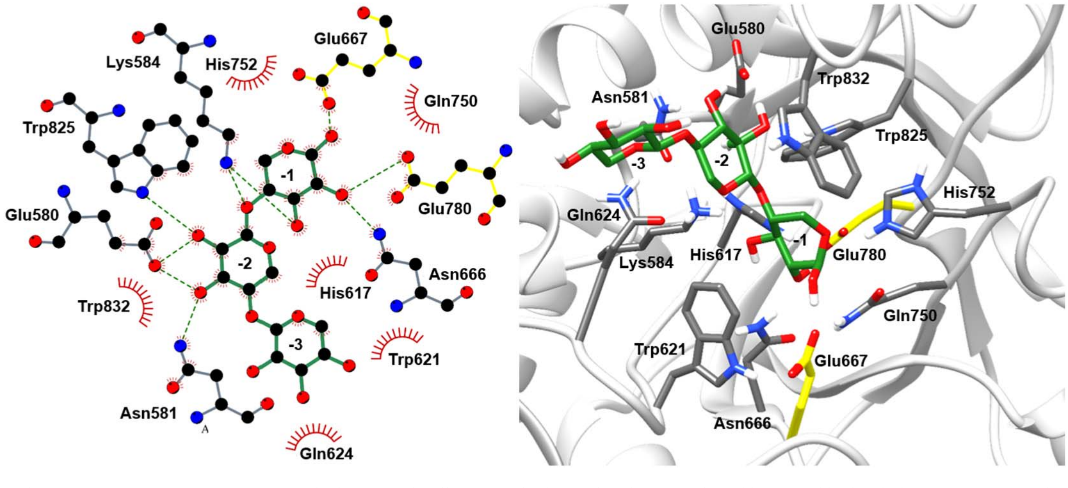


Fig. 5. X_3 docked into subsites -3 to -1 of *Rm*Xyn10_CM. To the left, 2-D plot illustrating hydrogen bonds (green dotted lines) and hydrophobic interactions (red stripes). To the right, 3-D representation. The catalytic residues are drawn in yellow.

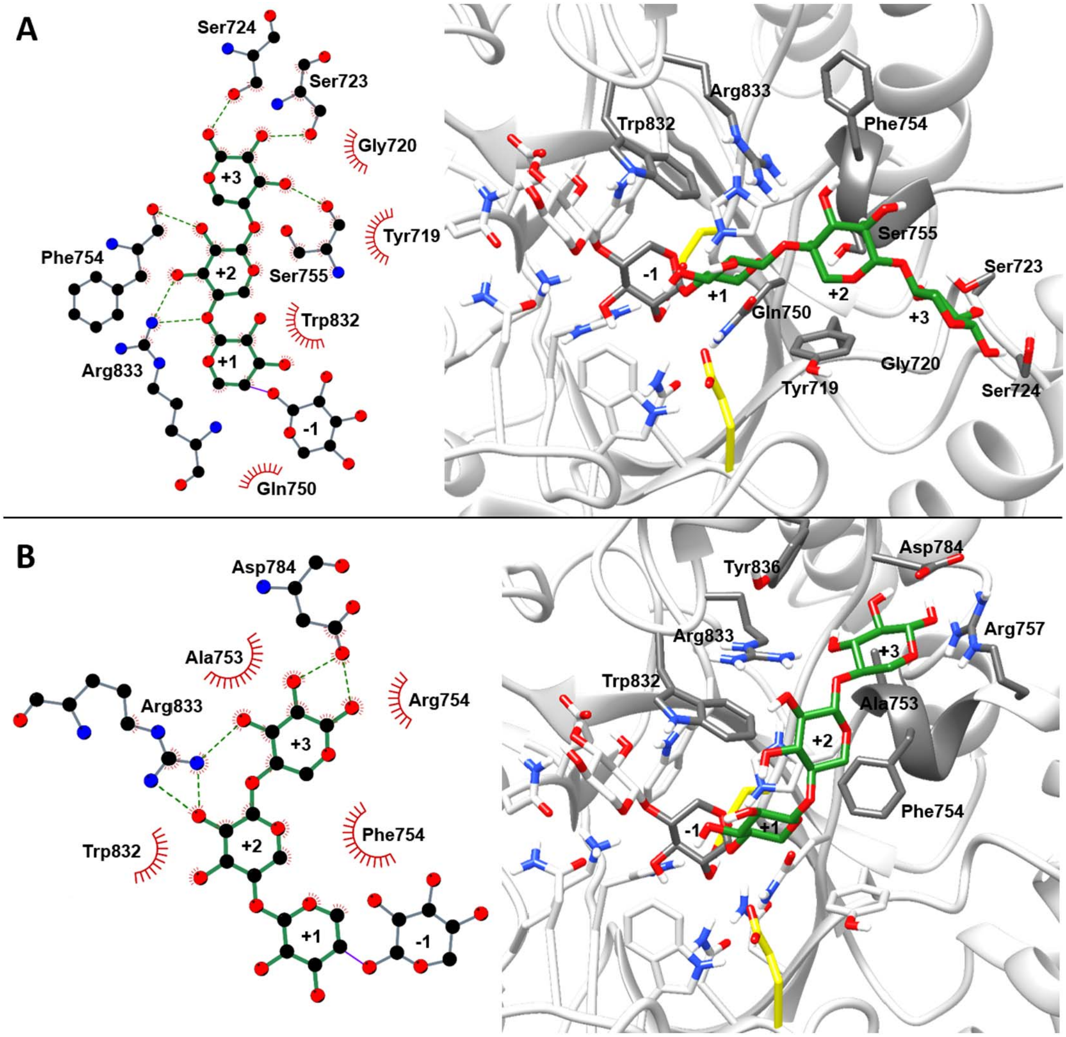


Fig. 6. Illustrations of the *Rm*Xyn10A aglycone-binding region in 2-D plots (left) and 3-D plots (right). A: alternative 1, the aglycone lies along the GH10 conserved binding cleft and B: alternative 2. Hydrogen bonds are illustrated in the 2-D plots by green dotted lines and hydrophobic interactions as red stripes. The catalytic residues are coloured in yellow and the non-conserved α-helix on loop $\beta_6\alpha_6$ is marked in dark grey.

conformation of loop $\beta_4\alpha_4$ combined with Gly720 instead of an Asparagine opens up the surface of the structure below subsite +3 as compared to other structures of GH10, and the distal part of the aglycone site is also lowered in comparison, with interactions between the xylose units and residues on loop $\beta_5\alpha_5$ rather than on $\beta_6\alpha_6$ or $\beta_7\alpha_7$. The position of Arg833 is conserved in several other GH10 xylanases forming hydrogen bonds to the cyclic oxygen between subsite +1 and +2 as well as to hydroxyl group of the xylose unit in subsite +2 (*e.g.* in *So*Xyn10A from *Streptomyces olivaceoviridis*, the GH10 xylanase with highest sequence identity to *Rm*Xyn10A having a complex with XOS in the aglycone region), but because of the somewhat lower position of the xylose chain in the aglycone part of *Rm*Xyn10A, the same hydrogen bonds were only obtained in a minority of the simulations. The aromatic Phe754 was highly flexible during the simulation and no single rotamer was favoured. Docking of AXOS with arabinose in the O3-position in subsite +2 requires that Arg833 is bent towards the backbone rather than reaching out to hydrogen bond with the xylose units. The arabinose bound in the O3-position in subsite +2 had many possible hydrogen bond- and hydrophobic interactions, so although a single fit was not possible to dock, it is evident that the affinity for these substitutions is high. Double-substitution in subsite +2 is possible although no strong interactions for the O2-bound arabinose were observed during the simulations. The hydroxyl groups of the xylose in subsite +3 are directed towards the protein surface and, thus, major steric hindrance prevents any arabinose-substitution in subsite +3. Fig. 7 presents the residues involved in interacting with possible arabinose substitutions in subsite −2 to +3 for the first alternative of the aglycone binding region.

In the second alternative aglycone-binding cleft, exemplified in Fig. 6-B, the xylan chain is more bent than in other GH10 complexes and runs along with β-strand 6 towards loop $\beta_7\alpha_7$. This conformation has not been observed in any other GH10 xylanase so far. Arabinose-substitutions in subsite +2 were not possible to dock in this second alternative. Hydroxyl-groups 2 and 3 of the xylose are directed towards the solvent and subsequently there is no affinity for the arabinoses unless these are positioned with abnormally bent glycosidic linkages. Also in subsite +3, arabinose substitutions are not possible due to steric hindrance. The simulations in the second cleft revealed a higher degree of uncertainty regarding the subsites and their interactions as the variation of the complexes obtained was higher than in the first alternative. However, the affinity for this region of the structure is relatively high as the xylan chain moved towards this cleft more often than towards the first alternative cleft during the simulations. These observations are not statistically proven and it must be emphasised that the results of MD are highly influenced by the starting complex which in turn is strongly influenced by the authors' idea about the how the ligand is supposed to dock given the structure, available structure-determined complexes, experience and continuous information obtained from simulations

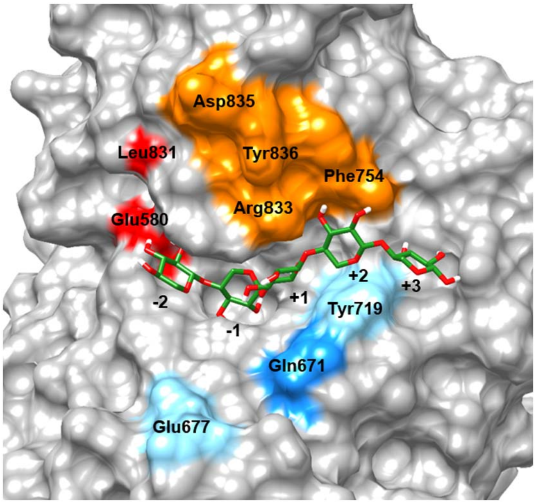


Fig. 7. Surface of *Rm*Xyn10A_CM docked with X_5 in the first potential cleft illustrating residues capable of interacting with arabinose substituents in subsite −2 (red), +1 (blue) and +2 (orange). The following positions have successfully been docked: O3 in subsite −2, O2, O3 and O2/O3 in subsite +1 and O3 and O2/O3 position and in subsite +2. Light blue indicates interacting residues unique to arabinose bound in the O2-position of subsite +1. Interaction with the highlighted residues has been observed in at least two simulations.

throughout the study.

In order to compare the two alternative aglycone clefts, Table 4 presents several evaluating parameters of the two alternative complexes formed. The binding energy of the ligand is slightly more negative for the first alternative and the C^{α} RMSD is lower, suggesting that this alternative is more probable. The bend in the xylan chain necessary for binding in the second active site cleft gives the glycosidic bond between the xylose units in subsite +1 and +2 abnormal torsion angles. Xylan has been reported to have a three-fold helix conformation with torsion angles φ, ψ_{O3} and ψ_{O5} in the range of −50 to −105, +135 and −59 to −116, respectively [22,78,97]. The glycosidic bond between subsite +1 and +2 of the second alternative falls outside these reported observations, suggesting that this conformation is atypical, has a higher energy and thus, disfavours the binding. Another aspect is the loop $\beta_7\alpha_7$ which constitutes a highly variable region within GH10, and longer loops are in several cases involved in facilitating distal aglycone subsites (*e.g.* BSX from *Bacillus subtilis* (PDB ID: 2FGL). Xyn10A from *Cellvibrio japonicas* (PDB ID: 1E5N), Xyn10B from *Cellvibrio mixtus* (PDB ID: 1UQZ) and extracellular xylanase T-6 from *Geobacillus stearothermophilus*, (PDB ID: 4PUD)). This loop in *Rm*Xyn10A has a length in between the longer ones involved in substrate binding and the shorter ones which only connect β-strand 7 and α-helix 7. If the binding cleft runs towards the loop, the length is enough to facilitate interaction between the amino acids of the loop and xylose units of subsite +3 and further out. The loop conformation is difficult to predict as it contains a large insertion compared to the template, and molecular dynamics in not a powerful enough tool to simulate large conformational change and loop flexibilities. The conformation of loop $\beta_7\alpha_7$ remains to be explored, and its putative role in substrate binding remains unclear.

For arabinoxylan (AX) to be hydrolysed, the first alternative of the two clefts has to be used in order to facilitate arabinose substitutions. However, the region containing Arg833, Tyr836 and Phe754 are involved in the binding of the arabinose in O3-position in subsite +2 of the first alternative but also the xylose in subsite +2 of the second alternative. Thus, this region has high affinity and it is possible that an unsubstituted xylan chain would bind differently than a substituted. The number of GH10 xylanases crystallised with a ligand in the aglycone subsites is limited, only five are reported in the CAZy database. The α-helix on loop $\beta_6\alpha_6$ has only been observed in one other GH10 xylanase, the template protein, XynB, and attempts to model a ligand into the aglycone region of XynB based on GH10 complexes were not successful [95]. It is, therefore, possible that XynB and *Rm*Xyn10A is part of a group within GH10 with a different aglycone conformation from the other previously-determined complex structures.

3.5. Site-directed mutagenesis in the aglycone region supports the more conserved GH10 binding cleft

A biochemical study of several mutants of *Rm*Xyn10A provides support for substrate binding in the first alternative of binding cleft (Fig. 6-A) rather than in the second one (Fig. 6-B). Seven mutants, namely G720N, S724G, D782G, D784G, F754N, F754Del and R833G were designed in order to investigate features of the aglycone region of the active site and are illustrated in Fig. 8. The wild-type (*wt*) and all mutants were constructed, expressed and purified successfully. The activity for each protein was determined by the DNS stopping method with birchwood xylan as substrate and can be found in Table 5. A longer reaction was carried out to study the product profiles. The same molar concentration of the proteins was used to degrade AX extracted from rye bran, samples were taken at 1 and 20 h and the product profiles were determined with High Performance Anionic Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). The results including all identified products can be found in Table S6 and selected data is presented in Fig. 9. Several non-identified peaks were observed, the major one eluting right after A^3X was assigned the name AXOS 1 and will be discussed further in the next section. The ratio of peak area between A^3X and AXOS 1 as well as the amount of end products are presented in Table 6. The products X_1, X_2, A^2XX and A^3X are considered end products as they previously have been seen to accumulate compared to the other identified products which are produced and later degraded [34]. The amount of end products was used in order to estimate the progression of the degradation.

Four of the mutants were designed to investigate the likelihood of binding in either of the two possible aglycone binding cleft identified in

Table 4
Evaluating parameters of X_5 docked into subsite −2 and +3 in two alternative active site clefts of *Rm*Xyn10A. Average and standard deviation for three representative simulations of each alternative are presented. The binding energies were calculated by two methods, here referred to VINA and AD_{100}, as explained in the Material and Method section.

	Potential active site cleft 1			Potential active site cleft 2		
Binding energy (kcal/mol)	VINA	AD_{100}		VINA	AD_{100}	
	−16.6 ± 0.6	−15.7 ± 0.6		−15.9 ± 0.3	−15.1 ± 0.5	
Overall RMSD C^{α} to final model (Å)	1.36 ± 0.03			1.40 ± 0.05		
Glycosidic bond angle (°)	φ	ψ_{C3}[a]	ψ_{C5}[b]	φ	ψ_{C3}	ψ_{C5}
−1/+1	−53 ± 1	172 ± 3	−65 ± 2	−51 ± 2	163 ± 4	−75 ± 4
+1/+2	−58 ± 7	−179 ± 2	−58 ± 2	54 ± 0	118 ± 4	−123 ± 5
+2/+3	−55 ± 9	176 ± 15	63 ± 15	−65 ± 37	174 ± 8	−67 ± 8

[a] ψ_{C3} is defined as the torsion angle for atoms C1′-O4-C4-C3 where ′ refers to the xylose unit towards the non-reducing end.
[b] ψ_{C5} is defined as the torsion angle for atoms C1′-O4-C4-C5 where ′ refers to the xylose unit towards the non-reducing end.

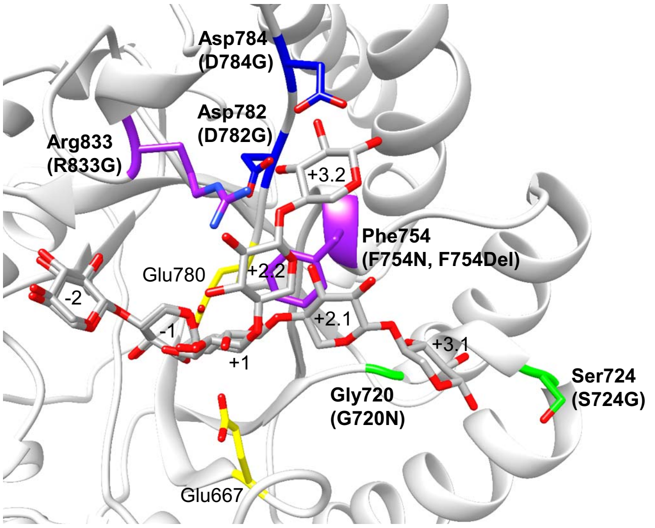


Fig. 8. Model of *Rm*Xyn10A with superimposition of two X_5 docked in subsites −2 to +3 illustrating the first and second alternative aglycone binding clefts. Mutated residues in the first (green) and second (blue) cleft, mutated residues involved in arabinose binding in subsite +2.1 (purple) are marked as well as the catalytic residues (yellow).

the MD simulations. G720N was designed to increase affinity for subsite +2 and +3 of the first alterative binding cleft (Fig. 6-A). With the exception of *Rm*Xyn10A, this residue is conserved in GH10 as an Asparagine. The mutation increased the activity of *Rm*Xyn10A with 15% (Table 5). S724G is situated in subsite +3 of the first alternative binding cleft and was designed to decrease affinity in this region. A minor decrease in activity was seen for this mutation. Both mutations G720N and S724G have a similar progression in the reaction as the *wt*. However, the mutants showed a slightly altered product profile comparted to *wt*, the amount of X_3 was lower while the amount of X_2 and X_1 was higher. These observations can be caused by a decreased affinity for subsite +3 and/or increased affinity for subsite +2 resulting in less X_3 being formed.

D782G and D784G were designed to reduce the affinity of the second alternative binding cleft (Fig. 6-B), where the residues are present in subsite +2 and +3 respectively. The mutations show a significant reduction in activity compared to *wt* (18 and 4% of *wt* activity, respectively). As the reaction proceeds at a slower pace for these mutants, a straight comparison of the product profiles is difficult to make. However, the 1 h time point of the *wt* and the 20 h time point of D784G shows the same amount of end products and comparison of these product profiles shows only minor differences. Neither any major differences, nor identical product profiles of *wt* and mutant D782G can be determined. The lowered activity but only slightly changes observed in product profiles caused by mutant D784G and D782G indicates that these residues are more involved in an overall stability of the protein rather than substrate binding. With the same reasoning, the remaining and increased activity in combination with altered product profiles of mutants in the first aglycone binding cleft suggest that these mutated residues are involved in substrate binding rather that structural integrity of the protein. Thus, these results points towards the first aglycone binding cleft (Fig. 6-A) being the actual one used, even for degradation of XOS.

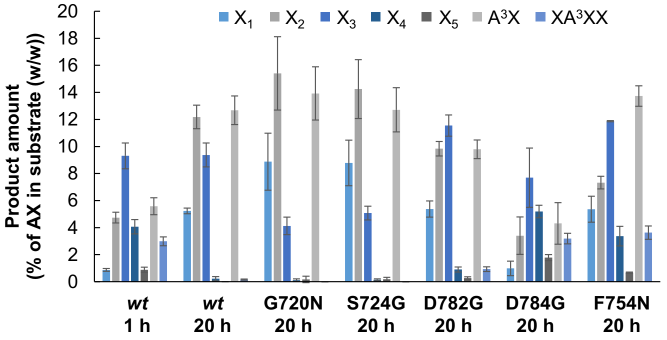


Fig. 9. Product profiles of *wt Rm*Xyn10A_CM and selected mutants during hydrolysis of AX extracted from rye bran. The reactions were carried out with 60 nmol protein per g AX in a 1% substrate solution (containing 62% AX) at 70 °C. All reactions were carried out in triplicates, the staples represent average and the error bars one standard deviation.

Table 6
Hydrolysis of AX extracted from rye bran by *wt* and mutants of *Rm*Xyn10A_CM. Amount of end products (X_1, X_2, A^2XX and A^3X) formed during the reaction as well as the peak area ratio between A^3X and the unidentified product AXOS 1 are presented in the table. The reactions were carried out with 60 nmol protein per g AX in a 1% substrate solution (containing 62% AX) at 70 °C, pH 7.4, samples were taken at 1 and 20 h. All reactions were carried out in triplicates, the numbers represent average and one standard deviation.

		Amount of end products (% of AX in substrate (w/w))[a]	Ratio of peak area A^3X/AXOS 1
wt	1 h	11.6 ± 0.7	0.64 ± 0.01
	20 h	30.3 ± 1.4	1.12 ± 0.01
G720N	1 h	11.4 ± 0.3	1.01 ± 0.07
	20 h	38.3 ± 4.0	1.55 ± 0.07
S724G	1 h	9.46 ± 0.67	0.59 ± 0.06
	20 h	35.9 ± 3.2	1.13 ± 0.05
D782G	1 h	3.92 ± 0.21	0.37 ± 0.05
	20 h	25.4 ± 1.1	0.77 ± 0.01
D784G	1 h	0.08 ± 0.02	n.d.[b]
	20 h	9.12 ± 2.14	0.53 ± 0.08
F754N	1 h	2.98 ± 0.22	2.17 ± 0.03
	20 h	26.7 ± 1.3	3.41 ± 0.32
F754Del	1 h	0.09 ± 0.02	n.d.
	20 h	7.54 ± 2.54	1.82 ± 0.44
R833G	1 h	0.02 ± 0.02	n.d.
	20 h	2.49 ± 1.17	0.97 ± 0.38
NC[c]	1 h	0.08 ± 0.00	n.d.
	20 h	0.28 ± 0.04	n.d.

[a] End products are X_1, X_2, A^2XX and A^3X. These products are accumulating during the reaction and the added amount of these products can thus be seen as a measure of how far the reaction have progressed.

[b] n.d. not detected, any or both of the two products were not detected.

[c] NC – negative control, no protein added in these reactions.

In order to study the effect of the non-conserved α-helix on loop $ß_6\alpha_6$ and the predicted region for binding an arabinose in subsite +2, F754Del, F754N and R833G were designed (Fig. 8). Substitution of the Phenylalanine with Asparagine resulted in loss of 82% of the activity. Interestingly, the ratio of peak area between A^3X and the unidentified peak AXOS 1 increased for the mutant F754N compared to *wt* even though the reaction has not proceeded as far, see Table 6. The product profile indicates that A^3X is produced in a higher amount than for the *wt*. Because XA^3XX does not seem to be affected by the mutation, the mutation seems to increase affinity for the arabinose in subsite +1

Table 5
Hydrolytic activity at 70 °C of *wt* and mutants of *Rm*Xyn10A_CM on birchwood xylan expressed as kat (mol/s reducing end) per mol protein and percentage of the *wt* activity. All measurements are done in triplicates, the numbers represent average and one standard deviation.

Activity	*wt*	G720N	S724G	D782G	D784G	F754N	F754 Del	R833G
kat/mol protein	30 ± 5	34 ± 2	28 ± 1	5.4 ± 0.6	1.2 ± 0.1	5.5 ± 0.5	0.5 ± 0.0	0.2 ± 0.0
% of *wt* activity		115 ± 20	92 ± 15	18 ± 3	4.1 ± 0.7	18 ± 3	1.8 ± 0.3	0.8 ± 0.2

rather than in subsite $+2$ (AXOS 1 is suggested to correspond to XA^{3}X as discussed in the next session). The mechanism for this increased affinity is however unclear, as the Asparagine cannot directly interact with an arabinose in subsite $+1$. Deletion of the Phenylalanine which eliminates the α-helix structure resulted in an activity loss of 98% and $>$ 99% of the activity was eliminated by the mutant R833G compared to *wt*. It is clear that these structures are important for the integrity of the protein. Thus, it is difficult to conclude their impact on the possibility of accommodating an arabinose in subsite $+2$.

3.6. RmXyn10A is a good candidate to hydrolyse highly substituted arabinoxylan

The combination of docking results and biochemical data from degradation of rye AX suggest the major hydrolysis products after hydrolysis of AX are X$_1$, X$_2$, X$_3$, A^{3}X and XA^{3}X, and minor products are A^{2}XX, XA^{3}XX, A^{2+3}XX and XA^{2+3}XX. Product profiles obtained from hydrolysis of XOS with varying degrees of polymerisation (DP) suggests that the main products after degradation of linear xylan are X$_2$ and X$_3$, and that X$_3$ slowly hydrolyses to X$_2$ and X$_1$ when longer substrates are completed [11,32,34]. This hydrolysis pattern indicates that the active site consists of five subsites, arranged two and three around the cleavage point. This arrangement is in accordance with the docking result as two glycone and at least three aglycone subsites can be identified. Since X$_3$ is the smallest substrate that *Rm*Xyn10A can hydrolyse, and because subsite -2 provides more hydrogen bonds than $+2$, occupation of subsites -2 to $+1$ is assumed to be vital for hydrolysis to occur. Subsites -2 to $+1$ are the most conserved subsites within GH10, they have been shown to have a strong binding energy, and especially subsite -2 is known to provide high affinity for xylose units [20].

During hydrolysis of AX extracted from rye bran, X$_2$, X$_3$ and A^{3}X are identified as major products. Fig. 10 presents a chromatogram from 1 h and 20 h of hydrolysis with *Rm*Xyn10A_CM which shows several other products with retention times above 16 min that are also detected but not identified. These peaks are assumed to correspond to AXOS. With support from the docking results, we suggest that the currently unidentified peak AXOS 1 corresponds to XA^{3}X.

The observed high amounts of A^{3}X and XA^{3}XX in the hydrolysate require O3-substitution possibilities in subsites -2, $+1$ and $+2$ which have been demonstrated during docking, Fig. 7. O3-substitutions in subsites $+1$ and $+2$ are not a general feature of GH10 xylanases, only substitution in subsite -2 is seen for the entire family. XA^{3}XX is expected to be an intermediate product, hydrolysed to XA^{3}X and X$_1$ upon depletion of longer substrates, in the same manner as X$_3$ is observed to be hydrolysed to X$_2$ and X$_1$. No larger O3-linked single-substituted AXOS are expected in the hydrolysate as they would be hydrolysed by *Rm*Xyn10A to smaller fragments.

O2-substituted AXOS are not expected in large amount due to the low presence of O2-substitutions in rye AX [4]. The smallest AXOS accumulated, A^{2}XX, confirms the docking results that no O2-substitutions can be accommodated in subsites -1 or -2 but can in $+1$, Fig. 7.

Double-substituted A^{2+3}XX and XA^{2+3}XX are predicted as hydrolysis products by the docking results, Fig. 7. In a previous study, the hydrolysate of the extracted rye bran AX from *Rm*Xyn10_CM was treated with α-L-arabinofuranosidases (Abf) from GH43 to determine the nature of the arabinose-substitutions [34]. Abf43 removes O3-linked arabinofuranose from double-substituted xylose units leaving a corresponding O2-substituted AXOS. The results showed an increase in both A^{2}XX and XA^{2}XX after treatment with Abf43 from *Bifidobacterium adolecentis* suggesting the presence of both A^{2+3}XX and XA^{2+3}XX in the hydrolysate after treatment with *Rm*Xyn10A_CM.

The combined results of docking and biochemical characterisation present *Rm*Xyn10A as a xylanase with a high capacity for hydrolysis of highly substituted AX to XOS with a DP up to three and AXOS with a DP up to four. *Rm*Xyn10A stands out among GH10 by allowing O3- and double-substitutions in both subsite $+1$ and $+2$ and is, thus, a good candidate when a highly substituted xylan is used as substrate and *e.g.* prebiotic (A)XOS is to be produced.

3.7. The catalytic module of RmXyn10A is positioned towards the surface of R. marinus

The full-length *Rm*Xyn10A contains five domains, including two CBM4 modules, a domain with unknown function (D3), the catalytic module and a putative cell-anchoring domain (D5) [31]. To study interactions between the domains and its position towards the cell surface of *R. marinus,* the modular arrangement was studied by small angle X-ray scattering (SAXS). The experimental scattering pattern from the full-length *Rm*Xyn10A is presented in Fig. 11, and the structural parameters computed from these data sets are given in Table 7. The estimated molecular mass (MM) of the solute agrees within the errors with the values predicted from the sequence. The SAXS data therefore indicate that the full-length xylanase is monomeric in solution at the protein concentrations used in these experiments.

A low-resolution model of the full-length *Rm*Xyn10A was reconstructed *ab initio* from the corresponding experimental scattering pattern. The reconstruction yielded superimposable results, provided good fits to the experimental data and shows that the full-length *Rm*Xyn10A is compatible with a five-domain assembly in as much as they can accommodate three domains with available high-resolution structures retaining the extra space for the remaining two domains. The averaged and the most typical *ab initio* model of the solute are presented in Fig. S5. A more detailed modelling was performed utilising the available high-resolution models of the individual domains. Several

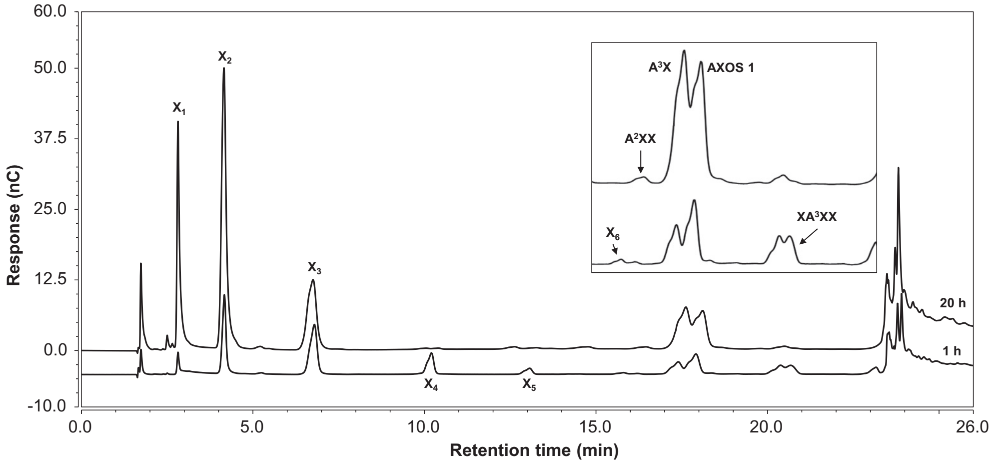


Fig. 10. HPAEC-PAD chromatogram showing products from hydrolysis of rye bran AX after 1 h (bottom) and 20 h (top) by *Rm*Xyn10A_CM. The square shows an enlargement of the response for the corresponding retention times. Peaks identified by standards are named as well as the unidentified peak AXOS 1. The suggestion for the unidentified peak AXOS 1 is XA^{3}X.

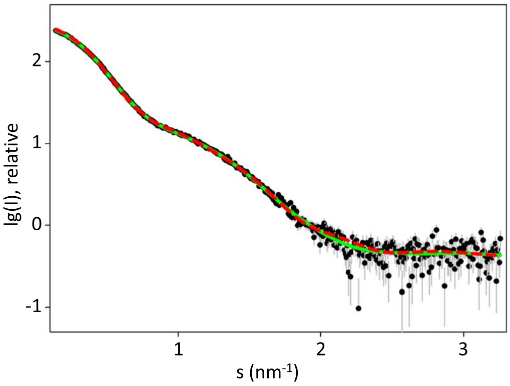


Fig. 11. X-ray scattering pattern and the scattering computed from the models of the full-length *Rm*Xyn10A. Dots with error bars: experimental data; solid green line: scattering from the typical *ab initio* model; dashed red line: scattering from the best rigid body model.

Table 7
Summary of structural parameters of the full-length *Rm*Xyn10A computed from the scattering data.

Parameter[a]	
R_g (Å)	39 ± 1
D_{max} (Å)	120
MM (kDa)	114 ± 10
MM_{seq} (kDa)	107.5
χ_S^2	1.72
χ^2	2.78

[a] R_g, D_{max} and MM are, respectively, the radius of gyration, maximum size and molecular mass, calculated from the scattering data. MM_{seq} is molecular mass of the solute predicted from the appropriate sequence. Discrepancies between the experimental data and computed scattering curves from the models are denoted as: χ_S^2 for *ab initio* shape determination; χ^2 for the models obtained by rigid body modelling.

runs were performed, which all yielded models compatible with the *ab initio* shape, although with somewhat different orientations of the three domains added. The most probable model is presented along with the *ab initio* models in Fig. S5 and in cartoon representation in Fig. 12.

In the rigid model, Fig. 12, the active site of the catalytic domain is directed away from the rest of the domains, and thus, no interference with the active site from surrounding domains occurs. This may explain why no added contribution to the specific activity is seen in trials comparing the full-length enzyme with the catalytic module alone, irrespective of whether the substrate is pure and soluble or in a complex mixture and including both soluble and insoluble components ([11], [33] and unpublished data). However, a decrease in activity has been observed when the catalytic domain is accompanied by either D3 or D5 [33]. It is therefore likely that, in the absence of CBM4-1 and CBM4-2, D3 and D5 can interfere with the catalytic module and decrease the activity. In the rigid model, the connection between the catalytic module and D5 is a linker separating the two domains in space. A possible decrease in activity suggests that this linker is flexible and could bring the two domains together. D3 is in contact with the catalytic module below the glycone part of the active site, on the outside of the barrel, over a rather large area. In the absence of the preceding domains it is likely that the interactions between the two domains are altered in a way that affects the activity.

The arrangement of the five domains of *Rm*Xyn10A opens up for speculation about potential cooperation among the domains to obtain efficiency in substrate hydrolysis. The catalytic module does not connect to the two CBMs which bind the substrate. Instead, D3 is situated in between and is thus likely to have a role in bridging the substrate from the CBMs to the catalytic module, because the space D3 occupies makes it possible for the CBMs to be further away from the cell to capture potential substrates. For the CBMs to be as far away from the cell as possible, the catalytic module would be close to the surface of the cell. In such a case, the long loop $\beta_7\alpha_7$ of the catalytic module could have a role in stabilising the position of the protein by binding to *e.g.* lipopolysaccharides covering the cell surface of *R. marinus*, explaining the atypical conformation of this loop in the homology model.

4. Conclusion

*Rm*Xyn10A is a glycoside hydrolase family 10 (GH10) xylanase well

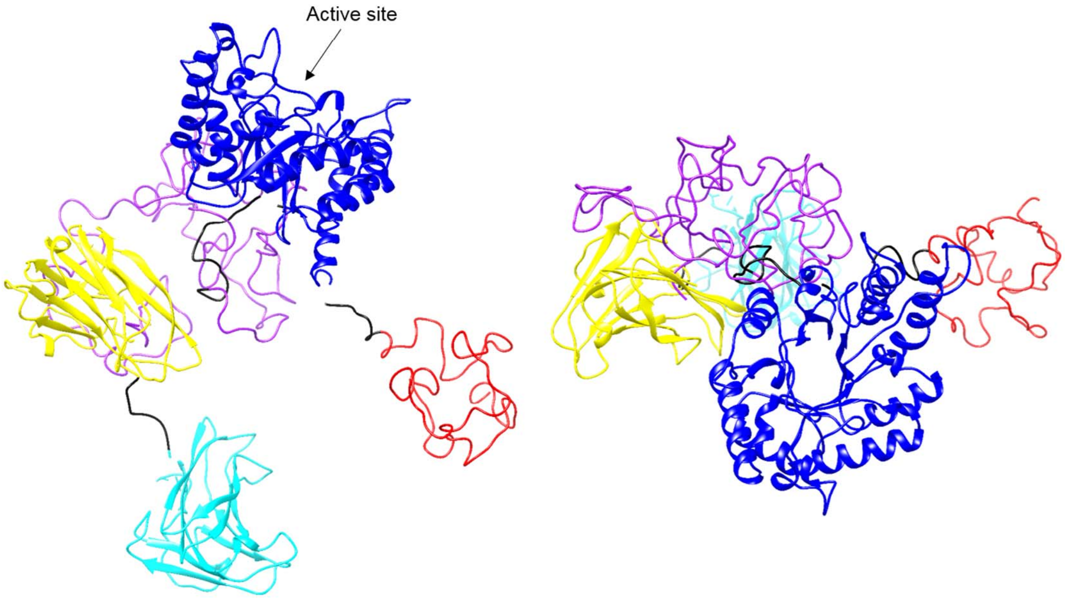


Fig. 12. Rigid model of the full-length *Rm*Xyn10A illustrating the spatial arrangement of its five domains; CMB4-1 (cyan), CBM4-2 (yellow), D3 (purple), catalytic module (blue) and D5 (red), where linkers are coloured in black. CBM2-1, CBM4-2 and the catalytic module is represented by high resolution structures while D3, D5 and linkers are represented by backbone atoms. The right panel is rotated clockwise by 90° around the horizontal axis.

adapted to hydrolyse arabinoxylan with a high degree of arabinose-substitution. The catalytic module has conserved catalytic residues and glycone subsites, but contains several non-conserved features in the aglycone region most likely responsible for the high tolerance for arabinose decorations observed. Loop $\alpha_4\beta_4$ provides interactions to a substituted arabinose in subsite $+1$. Substitution in subsite $+2$ is most likely provided by a high-affinity region above subsite $+2$ and the non-conserved α-helix on loop $\beta_6\alpha_6$ which occupies the space for the xylose in subsite $+2$ usually seen in GH10 xylanases and which forces the xylan chain to take a lower and slightly turned position. Only one other structure-determined GH10 xylanase shares the structural feature of the α-helix on loop $\beta_6\alpha_6$. The full-length *Rm*Xyn10A comprises a putative cell-anchoring domain close to the catalytic module while two carbohydrate binding modules are separated from the catalytic module by module D3 with unknown function. The position of the catalytic module, directed away from the rest of the domains, may explain the lack of contribution to the activity seen in trials comparing the full-length enzyme with the catalytic module alone.

Transparency document

The Transparency document associated with this article can be found, in online version.

Acknowledgement

This work was supported by VINNOVA *via* the Lund University Antidiabetic Food Centre (VINN Excellence Centre), by the Swedish Research Council (VR) [grant no. 2014-5038], and by the Swedish Research Council Formas [grant no. 2015-769]. Also to enable us to use the DESY-EMBL beamlines we are grateful for financial support from the European Community – Research Infrastructure Action under the FP6 "Structuring the European Research Area Programme" contract number RII3/CT/2004/5060008. The GROMACS simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at LUNARC (SNIC-2017/1-361). Björn Stenqvist, Lund University, is thanked for his computational assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbapap.2017.11.006.

References

[1] FAO Food Outlook, Biannual Report on Global Food Markets, Rome, (May 2015).

[2] J.L. Ward, K. Poutanen, K. Gebruers, V. Piironen, A.-M. Lampi, L. Nyström, A.A.M. Andersson, P. Åman, D. Boros, M. Rakszegi, Z. Bedő, P.R. Shewry, The Healthgrain cereal diversity screen: concept, results, and prospects, J. Agric. Food Chem. 56 (2008) 9699–9709.

[3] C. Maes, J.A. Delcour, Structural characterisation of water-extractable and water-unextractable arabinoxylans in wheat bran, J. Cereal Sci. 35 (2002) 315–326.

[4] C.J.A. Vinkx, J.A. Delcour, Rye (*Secale cereale* L.) arabinoxylans: a critical review, J. Cereal Sci. 24 (1996) 1–14.

[5] M.S. Izydorczyk, C.G. Biliaderis, Cereal arabinoxylans: advances in structure and physicochemical properties, Carbohydr. Polym. 28 (1995) 33–48.

[6] H.V. Scheller, P. Ulvskov, Hemicelluloses, Annu. Rev. Plant Biol. 61 (2010) 263–289.

[7] R. Fauré, C.M. Courtin, J.A. Delcour, C. Dumon, C.B. Faulds, G.B. Fincher, S. Fort, S.C. Fry, S. Halila, M.A. Kabel, L. Pouvreau, B. Quemener, A. Rivet, L. Saulnier, H.A. Schols, H. Driguez, M.J. O'Donohue, A brief and informationally rich naming system for oligosaccharide motifs of heteroxylans found in plant cell walls, Aust. J. Chem. 62 (2009) 533–537.

[8] W.F. Broekaert, C.M. Courtin, K. Verbeke, T. Van de Wiele, W. Verstraete, J.A. Delcour, Prebiotic and other health-related effects of cereal-derived arabinoxylans, arabinoxylan-oligosaccharides, and xylooligosaccharides, Crit. Rev. Food Sci. 51 (2011) 178–194.

[9] I. Figueroa-González, G. Quijano, G. Ramírez, A. Cruz-Guerrero, Probiotics and prebiotics-perspectives and challenges, J. Sci. Food Agric. 91 (2011) 1341–1348.

[10] R. Crittenden, S. Karppinen, S. Ojanen, M. Tenkanen, R. Fagerström, J. Mättö, M. Saarela, T. Mattila-Sandholm, K. Poutanen, In vitro fermentation of cereal dietary fibre carbohydrates by probiotic and intestinal bacteria, J. Sci. Food Agric. 82 (2002) 781–789.

[11] P. Falck, S. Precha-Atsawanan, C. Grey, P. Immerzeel, H. Stålbrand, P. Adlercreutz, E. Nordberg Karlsson, Xylooligosaccharides from hardwood and cereal xylans produced by a thermostable xylanase as carbon sources for *Lactobacillus brevis* and *Bifidobacterium adolescentis*, J. Agric. Food Chem. 61 (2013) 7333–7340.

[12] P. Immerzeel, P. Falck, M. Galbe, P. Adlercreutz, E. Nordberg Karlsson, H. Stålbrand, Extraction of water-soluble xylan from wheat bran and utilization of enzymatically produced xylooligosaccharides by *Lactobacillus*, *Bifidobacterium* and *Weissella* spp, LWT - Food Sci. Technol. 56 (2014) 321–327.

[13] V. Lombard, H. Golaconda Ramulu, E. Drula, P.M. Coutinho, B. Henrissat, The carbohydrate-active enzymes database (CAZy) in 2013, Nucleic Acids Res. 42 (2014) D490–D495.

[14] B. Henrissat, A classification of glycosyl hydrolases based on amino acid sequence similarities, Biochem. J. 280 (1991) 309–316.

[15] J.A. Linares-Pastén, A. Aronsson, E. Nordberg Karlsson, Structural considerations on the use of endo-xylanases for the production of prebiotic xylooligosaccharides from biomass, Curr. Protein Pept. Sci. 17 (2016) 1–20.

[16] P. Biely, M. Vršanská, M. Tenkanen, D. Kluepfel, Endo-β-1,4-xylanase families: differences in catalytic properties, J. Biotechnol. 57 (1997) 151–166.

[17] Ľ. Urbániková, M. Vršanská, K.B.R. Mørkeberg Krogh, T. Hoff, P. Biely, Structural basis for substrate recognition by *Erwinia chrysanthemi* GH30 glucuronoxylanase, FEBS J. 278 (2011) 2105–2116.

[18] F.J. St John, J.C. Hurlbert, J.D. Rice, J.F. Preston, E. Pozharski, Ligand bound structures of a glycosyl hydrolase family 30 glucuronoxylan xylanohydrolase, J. Mol. Biol. 407 (2011) 92–109.

[19] G.J. Davies, K.S. Wilson, B. Henrissat, Nomenclature for sugar-binding subsites in glycosyl hydrolases, Biochem. J. 321 (1997) 557–559.

[20] A. Pollet, J.A. Delcour, C.M. Courtin, Structural determinants of the substrate specificities of xylanases from different glycoside hydrolase families, Crit. Rev. Biotechnol. 30 (2010) 176–191.

[21] G. Pell, L. Szabo, S.J. Charnock, H. Xie, T.M. Gloster, G.J. Davies, H.J. Gilbert, Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases, J. Biol. Chem. 279 (2004) 11777–11788.

[22] L. Lo Leggio, J. Jenkins, G.W. Harris, R.W. Pickersgill, X-ray crystallographic study of xylopentaose binding to *Pseudomonas fluorescens* xylanase A, Proteins: Struct. Funct. Bioinf. 41 (2000) 362–373.

[23] G.W. Harris, J.A. Jenkins, I. Connerton, N. Cummings, L. Lo Leggio, M. Scott, G.P. Hazlewood, J.I. Laurie, H.J. Gilbert, R.W. Pickersgill, Structure of the catalytic core of the family F xylanase from *Pseudomonas fluorescens* and identification of the xylopentaose-binding sites, Structure 2 (1994) 1107–1116.

[24] G. Zolotnitsky, U. Cogan, N. Adir, V. Solomon, G. Shoham, Y. Shoham, Mapping glycoside hydrolase substrate subsites by isothermal titration calorimetry, P. Natl. Acad. Sci. USA 101 (2004) 11275–11280.

[25] G. Pell, E.J. Taylor, T.M. Gloster, J.P. Turkenburg, C.M.G.A. Fontes, L.M.A. Ferreira, T. Nagy, S.J. Clark, G.J. Davies, H.J. Gilbert, The mechanisms by which family 10 glycoside hydrolases bind decorated substrates, J. Biol. Chem. 279 (2004) 9597–9605.

[26] Z. Fujimoto, S. Kaneko, A. Kuno, H. Kobayashi, I. Kusakabe, H. Mizuno, Crystal structures of decorated xylooligosaccharides bound to a family 10 xylanase from *Streptomyces olivaceoviridis* E-86, J. Biol. Chem. 279 (2004) 9606–9614.

[27] M. Vardakou, J. Flint, P. Christakopoulos, R.J. Lewis, H.J. Gilbert, J.W. Murray, A. Family, 10 *Thermoascus aurantiacus* xylanase utilizes arabinose decorations of xylan as significant substrate specificity determinants, J. Mol. Biol. 352 (2005) 1060–1067.

[28] H. Xie, J. Flint, M. Vardakou, J.H. Lakey, R.J. Lewis, H.J. Gilbert, C. Dumon, Probing the structural basis for the difference in thermostability displayed by family 10 xylanases, J. Mol. Biol. 360 (2006) 157–167.

[29] G.A. Alfredsson, J.K. Kristjansson, S. Hjörleifsdottir, K.O. Stetter, *Rhodothermus marinus*, gen. nov., sp. nov., a thermophilic, halophilic bacterium from submarine hot springs in Iceland, Microbiology 134 (1988) 299–306.

[30] E. Nordberg Karlsson, E. Bartonek-Roxå, O. Holst, Cloning and sequence of a thermostable multidomain xylanase from the bacterium *Rhodothermus marinus*, BBA-Gen Struct. Exp. 1353 (1997) 118–124.

[31] M. Abou-Hachem, F. Olsson, M.P. Williamson, S. Linse, S.J. Crennell, G.O. Hreggvidsson, J.K. Kristjansson, O. Holst, E. Nordberg Karlsson, The modular organisation and stability of a thermostable family 10 xylanase, Biocatal. Biotransform. 21 (2003) 253–260.

[32] E. Nordberg Karlsson, L. Dahlberg, N. Torto, L. Gorton, O. Holst, Enzymatic specificity and hydrolysis pattern of the catalytic domain of the xylanase Xyn1 from *Rhodothermus marinus*, J. Biotechnol. 60 (1998) 23–35.

[33] M. Abou-Hachem, F. Olsson, E. Nordberg Karlsson, Probing the stability of the modular family 10 xylanase from *Rhodothermus marinus*, Extremophiles 7 (2003) 483–491.

[34] P. Falck, A. Aronsson, C. Grey, H. Stålbrand, E. Nordberg Karlsson, P. Adlercreutz, Production of arabinoxylan-oligosaccharide mixtures of varying composition from rye bran by a combination of process conditions and type of xylanase, Bioresour. Technol. 174 (2014) 118–125.

[35] D.M. Salas-Veizaga, R. Villagomez, J.A. Linares-Pastén, C. Carrasco, M.T. Álvarez, P. Adlercreutz, E. Nordberg Karlsson, Extraction of glucuronoarabinoxylan from quinoa stalks (*Chenopodium quinoa* Willd.) and evaluation of xylooligosaccharides produced by GH10 and GH11 xylanases, J. Agric. Food Chem. 65 (2017) 8663–8673.

[36] P.J. Simpson, S.J. Jamieson, M. Abou-Hachem, E. Nordberg Karlsson, H.J. Gilbert,

O. Holst, M.P. Williamson, The solution structure of the CBM4-2 carbohydrate binding module from a thermostable *Rhodothermus marinus* xylanase, Biochemistry 41 (2002) 5712–5719.

[37] E. Krieger, G. Vriend, YASARA View—molecular graphics for all devices—from smartphones to workstations, Bioinformatics 30 (2014) 2981–2982.

[38] E. Krieger, K. Joo, J. Lee, J. Lee, S. Raman, J. Thompson, M. Tyka, D. Baker, K. Karplus, Improving physical realism, stereochemistry, and side-chain accuracy in homology modeling: four approaches that performed well in CASP8, Proteins: Struct. Funct. Bioinf. 77 (2009) 114–122.

[39] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, B.L.A.S.T. Gapped, PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.

[40] B.E. Suzek, Y. Wang, H. Huang, P.B. McGarvey, C.H. Wu, the UniProt Consortium, UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches, Bioinformatics 31 (2015) 926–932.

[41] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, L.N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242.

[42] R.W.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272.

[43] R.W.W. Hooft, C. Sander, M. Scharf, G. Vriend, The PDBFINDER database: a summary of PDB, DSSP and HSSP information with added value, Comput. Appl. Biosci. 12 (1996) 525–529.

[44] D.T. Jones, Protein secondary structure prediction based on position-specific scoring matrices, J. Mol. Biol. 292 (1999) 195–202.

[45] R.D. King, M.J.E. Sternberg, Identification and application of the concepts important for accurate and reliable protein secondary structure prediction, Protein Sci. 5 (1996) 2298–2310.

[46] I.N. Shindyalov, P.E. Bourne, Protein structure alignment by incremental combinatorial extension (CE) of the optimal path, Protein Eng. 11 (1998) 739–747.

[47] Y. Ye, A. Godzik, Flexible structure alignment by chaining aligned fragment pairs allowing twists, Bioinformatics 19 (2003) ii246–ii255.

[48] J. Jung, B. Lee, Protein structure alignment using environmental profiles, Protein Eng. 13 (2000) 535–543.

[49] J. Qiu, R. Elber, SSALN: an alignment algorithm using structure-dependent substitution matrices and gap penalties learned from structurally aligned protein pairs, Proteins: Struct. Funct. Bioinf. 62 (2006) 881–891.

[50] U. Mückstein, I.L. Hofacker, P.F. Stadler, Stochastic pairwise alignments, Bioinformatics 18 (2002) S153–S160.

[51] A.A. Canutescu, R.L. Dunbrack, Cyclic coordinate descent: a robotics algorithm for protein loop closure, Protein Sci. 12 (2003) 963–972.

[52] A.A. Canutescu, A.A. Shelenkov, R.L. Dunbrack, A graph-theory algorithm for rapid protein side-chain prediction, Protein Sci. 12 (2003) 2001–2014.

[53] A. Klamt, Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena, J. Phys. Chem. 99 (1995) 2224–2235.

[54] E. Krieger, R.L. Dunbrack, R.W. Hooft, B. Krieger, Assignment of protonation states in proteins and ligands: combining pK_a prediction with hydrogen bonding network optimization, Computational Drug Discovery and Design, Springer, 2012, pp. 405–421.

[55] E. Krieger, T. Darden, S.B. Nabuurs, A. Finkelstein, G. Vriend, Making optimal use of empirical energy functions: force-field parameterization in crystal space, Proteins: Struct. Funct. Bioinf. 57 (2004) 678–683.

[56] U. Essmann, L. Perera, M.L. Berkowitz, T. Darden, H. Lee, L.G. Pedersen, A smooth particle mesh Ewald method, J. Chem. Phys. 103 (1995) 8577–8593.

[57] E. Krieger, J.E. Nielsen, C.A.E.M. Spronk, G. Vriend, Fast empirical pK_a prediction by Ewald summation, J. Mol. Graphics Modell. 25 (2006) 481–486.

[58] J. Bowie, R. Luthy, D. Eisenberg, A method to identify protein sequences that fold into a known three-dimensional structure, Science 253 (1991) 164–170.

[59] R. Lüthy, J.U. Bowie, D. Eisenberg, Assessment of protein models with three-dimensional profiles, Nature 356 (1992) 83–85.

[60] Č. Venclovas, A. Zemla, K. Fidelis, J. Moult, Some measures of comparative performance in the three CASPs, Proteins: Struct. Funct. Bioinf. 37 (1999) 231–237.

[61] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.

[62] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.

[63] M. Wiederstein, M.J. Sippl, ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins, Nucleic Acids Res. 35 (2007) W407–W410.

[64] M.J. Sippl, Recognition of errors in three-dimensional structures of proteins, Proteins Struct. Funct. Genet. 17 (1993) 355–362.

[65] C. Colovos, T.O. Yeates, Verification of protein structures: patterns of nonbonded atomic interactions, Protein Sci. 2 (1993) 1511–1519.

[66] F. Melo, E. Feytmans, Novel knowledge-based mean force potential at atomic level, J. Mol. Biol. 267 (1997) 207–222.

[67] F. Melo, E. Feytmans, Assessing protein structures with a non-local atomic interaction energy1, J. Mol. Biol. 277 (1998) 1141–1152.

[68] F. Melo, D. Devos, E. Depiereux, E. Feytmans, ANOLEA: a www server to assess protein structures, ISMB 5 (1997) 187–190.

[69] M.J. Abraham, D. van der Spoel, E. Lindahl, B. Hess, the GROMACS development team, GROMACS User Manual version 2016, www.gromacs.org, (2016).

[70] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: a message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1995) 43–56.

[71] Y. Duan, C. Wu, S. Chowdhury, M.C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations, J. Comput. Chem. 24 (2003) 1999–2012.

[72] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, J. Chem. Phys. 126 (2007) 014101.

[73] S. Nosé, M.L. Klein, Constant pressure molecular dynamics for molecular systems, Mol. Phys. 50 (1983) 1055–1076.

[74] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2010) 455–461.

[75] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.

[76] E. Krieger, G. Koraimann, G. Vriend, Increasing the precision of comparative models with YASARA NOVA—a self-parameterizing force field, Proteins: Struct. Funct. Bioinf. 47 (2002) 393–402.

[77] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand–protein interaction diagrams for drug discovery, J. Comp. Inf. Modell. 51 (2011) 2778–2786.

[78] T. Yui, K. Imada, N. Shibuya, K. Ogawa, Conformation of an arabinoxylan isolated from the rice endosperm cell wall by X-ray diffraction and a conformational analysis, Biosci. Biotechnol. Biochem. 59 (1995) 965–968.

[79] J.A. Linares-Pastén, P. Falck, K. Albasri, S. Kjellström, P. Adlercreutz, D.T. Logan, E. Nordberg Karlsson, Three-dimensional structures and functional studies of two GH43 arabinofuranosidases from *Weissella* sp. strain 142 and *Lactobacillus brevis*, FEBS J. 284 (2017) 2019–2036.

[80] M.J. Bailey, P. Biely, K. Poutanen, Interlaboratory testing of methods for assay of xylanase activity, J. Biotechnol. 23 (1992) 257–270.

[81] M.H.J. Koch, J. Bordas, X-ray diffraction and scattering on disordered systems using synchrotron radiation, Nucl. Inst. Methods 208 (1983) 461–469.

[82] C.J. Boulin, R. Kempf, A. Gabriel, M.H.J. Koch, Data acquisition systems for linear and area X-ray detectors using delay line readout, Nucl. Instrum. Meth. A 269 (1988) 312–320.

[83] A. Gabriel, F. Dauvergne, The localisation method used at EMBL, Nucl. Inst. Methods 201 (1982) 223–224.

[84] P.V. Konarev, V.V. Volkov, A.V. Sokolova, M.H.J. Koch, D.I. Svergun, PRIMUS: a Windows PC-based system for small-angle scattering data analysis, J. Appl. Crystallogr. 36 (2003) 1277–1282.

[85] L.A. Feigin, D.I. Svergun, G.W. Taylor, General principles of small-angle diffraction, in: G.W. Taylor (Ed.), Structure Analysis by Small-angle X-Ray and Neutron Scattering, Springer, Boston, 1987, pp. 25–55.

[86] D.I. Svergun, A direct indirect method of small-angle scattering data treatment, J. Appl. Crystallogr. 26 (1993) 258–267.

[87] D.I. Svergun, Determination of the regularization parameter in indirect-transform methods using perceptual criteria, J. Appl. Crystallogr. 25 (1992) 495–503.

[88] A. Guinier, La diffraction des rayons X aux tres petits angles: applications a l'etude de phenomenes ultramicroscopiques, Ann. Phys. (1939) 161–237.

[89] D.I. Svergun, M.V. Petoukhov, M.H.J. Koch, Determination of domain structure of proteins from X-ray solution scattering, Biophys. J. 80 (2001) 2946–2953.

[90] V.V. Volkov, D.I. Svergun, Uniqueness of ab initio shape determination in small-angle scattering, J. Appl. Crystallogr. 36 (2003) 860–864.

[91] M.B. Kozin, D.I. Svergun, Automated matching of high- and low-resolution structural models, J. Appl. Crystallogr. 34 (2001) 33–41.

[92] M.V. Petoukhov, N.A.J. Eady, K.A. Brown, D.I. Svergun, Addition of missing loops and domains to protein models by X-ray solution scattering, Biophys. J. 83 (2002) 3113–3125.

[93] D.I. Svergun, C. Barberato, M.H.J. Koch, CRYSOL - a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates, J. Appl. Crystallogr. 28 (1995) 768–773.

[94] M.V. Petoukhov, D.I. Svergun, Global rigid body modeling of macromolecular complexes against small-angle scattering data, Biophys. J. 89 (2005) 1237–1250.

[95] C.R. Santos, Z.B. Hoffmam, V.P. de Matos Martins, L.M. Zanphorlin, L.H. de Paula Assis, R.V. Honorato, P.S. Lopes de Oliveira, R. Ruller, M.T. Murakami, Molecular mechanisms associated with Xylan degradation by *Xanthomonas* plant pathogens, J. Biol. Chem. 289 (2014) 32186–32200.

[96] M. Abou-Hachem, E. Nordberg Karlsson, P.J. Simpson, S. Linse, P. Sellers, M.P. Williamson, S.J. Jamieson, H.J. Gilbert, D.N. Bolam, O. Holst, Calcium binding and thermostability of carbohydrate binding module CBM4-2 of Xyn10A from *Rhodothermus marinus*, Biochemistry 41 (2002) 5720–5729.

[97] K. Mazeau, C. Moine, P. Krausz, V. Gloaguen, Conformational analysis of xylan chains, Carbohydr. Res. 340 (2005) 2752–2760.