Chapter 4

Structural Basis for the Substrate Specificity of a *Bacillus* 1,3-1,4-β-Glucanase

4.1 Crystallization and Crystal Structure Analysis

For crystallization set-ups, H(A16-M)$^{E105Q/E109Q}$, concentrated to 10 mg/mL (0.42 mM) and supplemented with 2 mM $CaCl_2$, was pre-incubated with the natural substrate β-glucan hexasaccharide Glcβ4Glcβ4Glcβ3Glcβ4Glcβ4Glc in a molar protein-to-carbohydrate ratio of 1:10. Crystals were readily obtained within 2-3 days with the method of hanging-drop vapor-diffusion by mixing the protein-carbohydrate solution with an equal volume of crystallization buffer containing 100 mM MES pH 6.5, 10 mM $ZnSO_4 \bullet 7H_2O$, and 25% (v/v) PEG mono-methyl ether 550. The crystals had $P2_12_12_1$ symmetry and diffracted beyond 1.6 Å resolution. Initial protein phases were obtained by using the native-like hybrid enzyme H(A16-M) as a model. The structure was improved by several rounds of phase refinement with a step-wise increase of the resolution range, additionally considering ions and water molecules which were located by a $|F_o|$-$|F_c|$ difference density peak search. The carbohydrate ligand was not accounted for in the early stage of refinement to avoid model bias. However, the difference electron density unambiguously revealed the presence of a β-glucan tetrasaccharide moiety consisting of four covalently linked glucosyl units in the substrate binding cleft. Model refinement considering reflections in the range of 34.5 – 1.64 Å finally converged to an *R*-value of 16.3% (R_{free} = 19.7%). Analysis of the polypeptide backbone conformation showed no residues in the generously allowed and disallowed regions of the Ramachandran plot, whereas more than 90% of all non-Pro/Gly residues were in the most favored regions. Data collection and refinement statistics are listed in [44].

The asymmetric unit contains four protein-carbohydrate complexes (Figure 13). The four protein molecules are very similar, showing identical secondary structure and only minor conformational heterogeneities with an RMSD for all C^{α} atoms of 0.142 Å. H(A16-M) is an all β-protein with two major seven-stranded antiparallel β-sheets arranged atop each other which are bent, forming a cleft in which substrate binding and hydrolysis takes place. The

substrate binding cleft may accommodate six glucosyl residues. Four calcium (Ca^{2+}) ions are each bound with pentagonal-bipyramidal geometry to the protein backbone oxygen atoms of Pro9, Gly75 and Asp207, the carboxylate oxygen $O^{\delta 1}$ from Asp207, and three water molecules. In addition, two zinc (Zn^{2+}) ions are located on a two-fold symmetry axis, coordinating in tetrahedral geometry to His145 $N^{\delta 1}$ and Asp161 $O^{\delta 2}$ from one protein molecule and to the equivalent amino acid residues from a symmetry-related molecule. A total of 1596 firmly bound water molecules (i.e. 399 water molecules per protein-carbohydrate complex, on average) were identified in the asymmetric unit. The high quality of the diffraction data and the presence of four crystallographically independent complexes allowed unambiguous determination of carbohydrate conformation, linkage-type and enzyme-ligand interactions. A detailed description of crystal analysis and structure solution is given in [44].

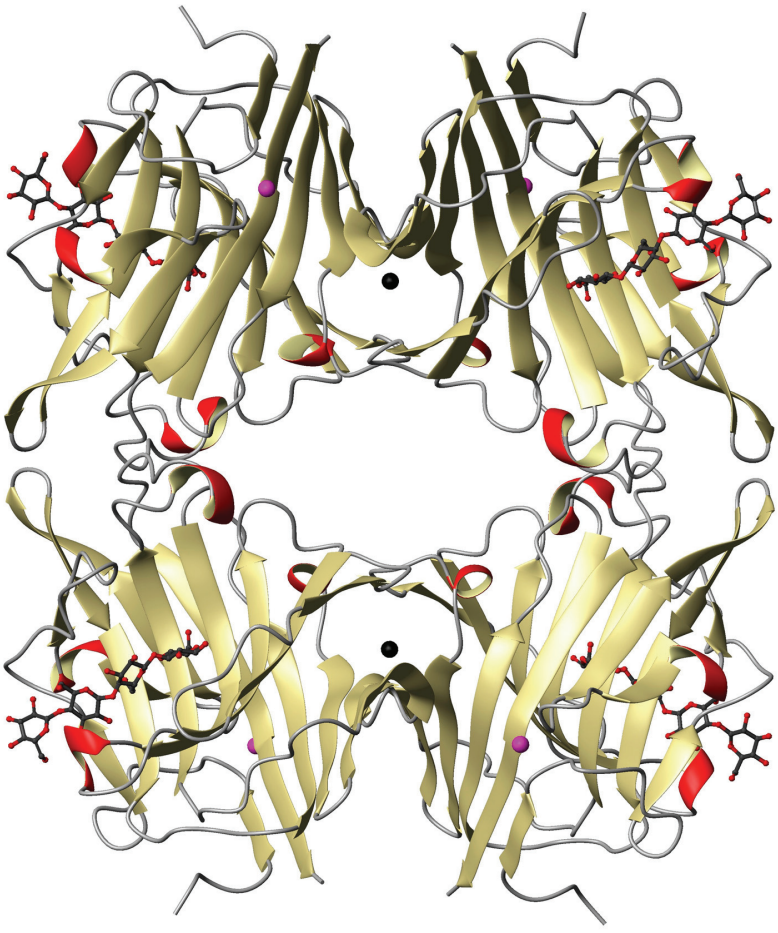

Figure 13: Crystal packing of the H(A16-M)$^{E105Q/E109Q}$-tetrasaccharide complex.
Molecular assembly in the asymmetric unit comprising four non-covalent complexes between the β-glucanase variant H(A16-M)$^{E105Q/E109Q}$ and the tetrasaccharide ligand (Glcβ4Glcβ4Glcβ3Glc). In addition, the asymmetric unit contains four calcium (Ca^{2+}) ions (colored magenta), each coordinated to the protein's calcium binding site, and two zinc (Zn^{2+}) ions (colored black) which mediate crystal contacts between two symmetry related protein-carbohydrate complexes. The protein backbone is depicted as ribbon diagram with β-strands colored light yellow and α-helices colored red. The carbohydrate moiety is shown in ball-and-stick representation with carbon and oxygen atoms colored grey and red, respectively.

4.2 The H(A16-M)$^{E105Q/E109Q}$-Tetrasaccharide Complex

4.2.1 Carbohydrate Binding

Crystal structure analysis of the enzyme-carbohydrate complex revealed a tetrasaccharide moiety bound in subsites –IV to –I with a β-1,3 glycosidic linkage between the glucosyl residues located in subsites –II and –I. The tetrasaccharide ligand Glcβ4Glcβ4Glcβ3Glc (3-O-β-cellotriosyl-β-D-glucopyranose) represents the natural reaction product of 1,3-1,4-β-glucanases. Thus, although the *hexa*saccharide substrate has been used for co-crystallization with the catalytically impaired enzyme variant H(A16-M)$^{E105Q/E109Q}$, a *tetra*saccharide complex was found in the crystal, representing a non-covalent enzyme-product (E·P) complex rather than the expected enzyme-substrate (E·S) complex. This result was well surprising as previous biochemical studies have shown that mutation of the catalytic glutamate residues to the isosteric glutamines (E105Q, E109Q) results in complete loss of enzymatic activity. The tetrasaccharide ligand is firmly bound in relaxed conformation, covering the glucosyl-binding subsites –IV to –I, whereas subsites +I and +II are occupied by a network of well-ordered, discrete water molecules (Figures 13 and 14). This effectively rules out the possibility that the two missing glucosyl units expected to be bound in subsites +I/II might be actually present but not visible in the electron density due to their high flexibility. It is therefore concluded that the hexasaccharide substrate used for crystallization has been catalytically processed by trace impurities of wild-type enzyme which might have been present in the crystallization set-ups as a result of (partial) deamidation of Gln105 and Gln109 in H(A16-M)$^{E105Q/E109Q}$.

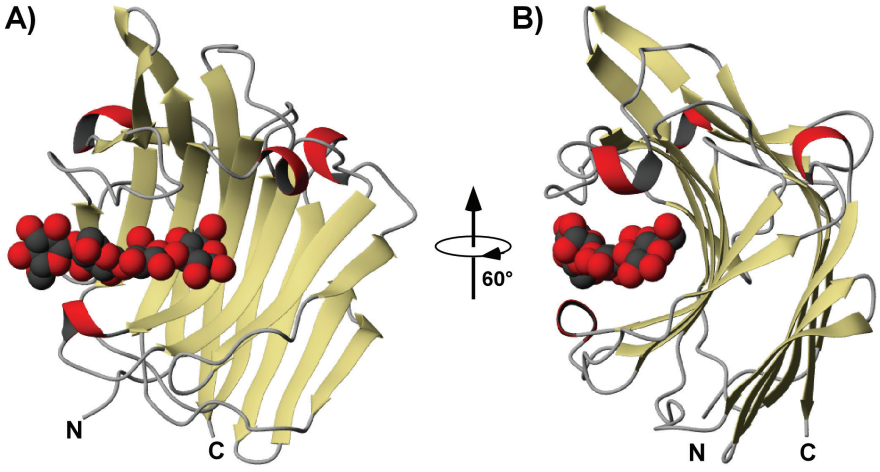

Figure 14: Structure of H(A16-M)$^{E105Q/E109Q}$ in complex with β-glucan tetrasaccharide.
(a) Side-view of the non-covalent protein-carbohydrate complex with the non-reducing end of the tetrasaccharide ligand pointing to the left. **(b)** View along the substrate binding cleft from the reducing end. The protein is depicted as ribbon diagram, β-strands and α-helices colored light yellow and red, respectively. The carbohydrate moiety is shown as space-filling model with carbon atoms colored grey and oxygen atoms colored red.

4.2.2 Molecular Basis of Substrate Specificity

1,3-1,4-β-glucanases are highly specific enzymes that catalyze the hydrolysis of β-1,4 glycosidic bonds in polysaccharides with mixed β-1,3 and β-1,4 linkage, as opposed to all β-1,3 linked laminarin and all β-1,4 linked cellulose, thereby discriminating between different linkage-types and OH-substitution patterns. Binding of a laminaribiosyl unit to subsites –II/–I is a major determinant of substrate recognition, and enzyme-carbohydrate interactions may therefore elucidate the molecular basis of substrate specificity. Notably, in a β-1,3 linked laminaribiosyl unit, as in Glc-II-(β-1,3)-Glc-I of the E·P complex, the 6-CH_2-OH side chains are on the same side of the glucopyranosyl rings (Figure 15), whereas the β-1,4 linkage in a cellobiosyl unit causes the orientation of the 6-CH_2-OH side chains to alternate with respect to the sugar rings. The E·P complex reveals indeed an intricate network of hydrogen bonds which mediate specific interactions with the carbohydrate hydroxyl groups in subsites –IV to –I. A maximum of hydrogen bonds is observed in the catalytic subsite –I, including interactions between the O^1 and O^2 hydroxyls with the nucleophile equivalent Gln105, and between the O^6 hydroxyl and the acid/base catalyst equivalent Gln109. Beyond numerous hydrogen bonds, two direct hydrophobic stacking interactions involving Tyr94 and Phe92 in subsites –II and –I, respectively, contribute further to carbohydrate binding (Figure 15). The combination with kinetic and mutagenesis data provided further insights into the molecular basis of highly specific substrate recognition and enzymatic hydrolysis.

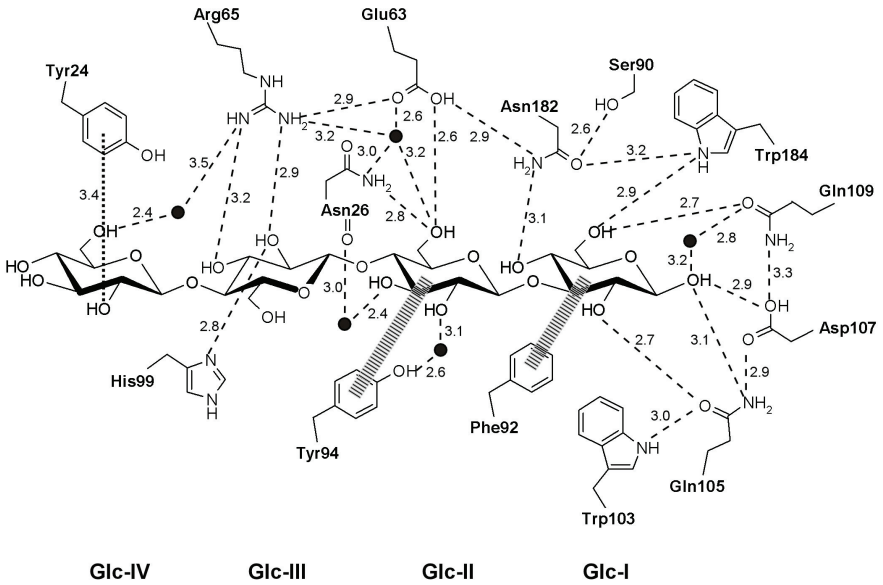

Figure 15: Enzyme-carbohydrate interactions in the glucosyl-binding subsites –I to –IV.
Schematic depiction E·P interactions. Amino acid side chains are shown with their functional groups. H-bonds (3.5 Å upper-limit cut-off) are shown with broken lines; numbers denote donor-acceptor distances in Å. Water molecules are shown as black balls. Hydrophobic stacking interactions between aromatic amino acid side chains and the sugar rings are also indicated (|||||). Glc-I to Glc-IV mark glucosyl residues bound in subsites –I to –IV.

4.3 Discussion

4.3.1 Hydrophobic Interactions and Substrate Sliding

An emerging common feature of saccharide binding proteins, may they be carbohydrate binding modules (CBMs) or processing enzymes (e.g. glycosyl hydrolases), is the presence of a number of aromatic residues that are involved in hydrophobic stacking interactions with the polysaccharide substrate. Although carbohydrates are highly hydrophilic due to the many polar hydroxyl groups they may be well involved in hydrophobic interactions by the rather nonpolar, almost flat faces of the glucopyranose sugar rings. In fact, hydrophobic interactions are considered the main driving force for unspecific saccharide binding, whereas substrate specificity in terms of anomeric configuration, conformation and linkage-type is based on the positioning of hydroxyl groups and an associated sugar-type specific hydrogen bonding pattern. A particularly interesting feature of carbohydrate binding is observed in some linkage-type specific glycosyl hydrolases, in which the binding energy profile of the hydrophobic interactions does not match exactly with that of the hydrogen bonding interactions, the former being slightly off-positioned with respect to the latter. At first glance this might be counter-intuitive, since substrate recognition and initial binding becomes negatively affected by the hydrophobic interactions. However, as outlined for the case of cellobiohydrolase I (CBH-I) [68], the energy separation of protein-carbohydrate interactions in different subsites is modulated and effectively smoothened, thus allowing effective substrate sliding within the binding cleft [69]. Although a multiple attack mechanism is not clearly shown for retaining *endo*-glucanohydrolases, carbohydrate-sliding might be an important feature of product release: upon cleavage of the scissile glycosidic bond, the reaction product bound in subsites –I to –IV is free to adopt ground state conformation, concomitant with a reduction in enzyme affinity and sliding within the binding cleft.

4.3.2 pK_a Modulation of the General Acid/Base Catalyst

The double-displacement reaction mechanism of retaining β-glucanases requires the pK_a of the acid/base catalyst to cycle along the reaction coordinate: the very same residue acts as general acid in the first step (glycosylation) whereas it is the general base in the second step (deglycosylation). pK_a cycling is supported by additional hydrogen bond interactions between the acid/base catalyst and nearby auxiliary amino acid residues. As such, Asp107 rotates away from the nucleophile (Glu105) upon glycosylation to form a new hydrogen bond with the acid/base catalyst Glu109, thereby assisting its pK_a drop by stabilizing the

deprotonated carboxylate group. After deglycosylation, Asp107 returns to its initial position, leaving the acid/base catalyst to interact with Gln119 and a network of four water molecules in the area of subsite +I. Thus, pK_a cycling of the general acid/base (Glu109) along the reaction coordinate is mediated by hydrogen bond interactions with Asp107 which form upon removal of the negative charge of the nucleophile when it becomes covalently bound in the glycosyl-enzyme intermediate. As to whether the auxiliary amino acid residues Asp107 and Trp103 play a similar role in deglycosylation by means of stabilizing the leaving carboxylate is unknown, since glycosylation is the rate-limiting step even for activated substrates.

In the E·P structure, both Trp184 and Gln109 interact with the O^6 hydroxyl of Glc-I, whereas Gln119 forms weak water-mediated hydrogen bonds to the endocyclic O^5 and the β-configured O^1, the latter being equivalent to the glycosidic oxygen in the substrate. As discussed in the publication describing the results of this structural study [44], the aforementioned hydrogen bond network observed in the E·P complex might suit for proton transfer to the glycosidic oxygen in the Michaelis complex and stabilization of the transition state. This opens the intriguing possibility of carbohydrate assisted pK_a modulation of the general acid/base catalyst, which implicates a ***functional induced fit mechanism*** for catalysis.

4.3.3 Substrate Distortion in the Michaelis Complex

The reaction mechanism outlined in Chapter 1.2.1 ("Carbohydrate Depolymerization by Retaining β-Glucanases") implies an oxocarbenium-ion-like transition state, characterized by a partial positive charge on the anomeric carbon atom C^1. In addition, the conformation of the sugar ring in subsite –I (Glc-I) is expected to become distorted in the E·S complex upon binding. Such ring distortion has been observed in E·S complexes of various glycoside hydrolases with different substrate specificity. In these cases the sugar ring in the catalytic subsite mostly adopts a $^1S_3$ skew-boat conformation that is assumed to be of particular advantage for the stereochemistry and stereoelectronics of the (rate-limiting) glycosylation reaction.

Recently, enzyme-substrate interactions of H(A16-M) have been modeled using a combined quantum mechanics and molecular dynamics approach (QM/MM protocol) with the aim of calculating the sugar conformation in the catalytic subsite –I [70]. However, the advanced QM/MM modeling algorithm failed to reproduce the expected substrate distortion in subsite –I when simulating the binding event *in silico*. This might be due to the artificial substrate molecule used in these calculations with a methylumbelliferyl aglycon to be bound in subsite +I, and the difficulties related to the assessment of the energetic contributions

associated with hydrophobic interactions, H-bonds, and electrostatic interactions in the E·S complex. The calculation of *binding energies* suggested a 4 kcal/mol stabilization of the $^{1}S_{3}$ skew-boat conformation compared to the relaxed $^{4}C_{1}$ chair conformation. Furthermore, as a result of the modeling studies, interconversion between $^{1}S_{3}$ and $^{4}C_{1}$ is effectively blocked by a huge energy barrier of 80 kcal/mol. Applying the QM/MM protocol to the reaction product suggested a $^{4}C_{1}$ chair conformation of Glc-I in the E·P complex, well consistent with the crystallographic results of this thesis work. Thus, the substrate distortion predicted by the Michaelis complex model would be induced primarily by E·S interactions in subsite +I (i.e. beyond the cleavage point). However, since in the case of the methylumbelliferyl aglycon the stacking interactions but not the hydrogen bonding pattern in subsite +I are properly accounted for, the results of this modeling study may deviate from the situation in solution and should be therefore used with some caution. This raises the question as to the "real" situation in solution. Assuming distortion ($^{1}S_{3}$ skew-boat conformation) of Glc-I in the Michaelis complex, initial substrate binding may occur with the sugar ring becoming positioned in the catalytic subsite –I either in the relaxed $^{4}C_{1}$ conformation or in the distorted $^{1}S_{3}$ conformation. Since the latter is energetically disfavored, the equilibrium is shifted (far) to the relaxed $^{4}C_{1}$ conformation. By this the fraction of substrate molecules in the preformed $^{3}S_{1}$ conformation would be far too small to account for the observed catalytic performance of glycosyl hydrolases. However, forcing the Glc-I sugar ring from the relaxed $^{4}C_{1}$ chair conformation into the distorted, unfavorable $^{1}S_{3}$ conformation may be accomplished by the binding event itself, with most of the required energy coming from the favorable interactions with the enzyme. Therefore it is most likely that the carbohydrate substrate is initially recognized and bound in its preferred $^{4}C_{1}$ conformation followed by a conformational change along the reaction coordinate to adopt the $^{1}S_{3}$ conformation in subsite –I upon productive binding in the Michaelis complex.